ASPAI
CORFU • 2022

Advances in Signal Processing and Artificial Intelligence

Proceedings of the 4th International Conference on Advances in Signal Processing and Artificial Intelligence (ASPAI' 2022)

Edited by Sergey Y. Yurish

IFSA

Advances in Signal Processing and Artificial Intelligence:

Proceedings of the 4th International Conference on Advances in Signal Processing and Artificial Intelligence

19-21 October 2022
Corfu, Greece

Edited by Sergey Y. Yurish

IFSA Publishing

Sergey Y. Yurish, *Editor*
Advances in Signal Processing and Artificial Intelligence
ASPAI' 2022 Conference Proceedings

E-mail (for orders and customer service enquires): ifsa.books@sensorsportal.com

Visit our Home Page on https://sensorsportal.com/ifsa_publishing.html

ASPAI Conference Website: https://aspai-conference.com/

ISSN: 2938-5350
ISBN: 978-84-09-45050-3
BN-20221014-XX
BIC: UYQ

Contents

Foreword

On behalf of the ASPAI' 2022 Organizing Committee, I introduce with pleasure these proceedings devoted to contributions from the 4th International Conference on Advances in Signal Processing and Artificial Intelligence (ASPAI' 2022).

Advances in artificial intelligence (AI) and signal processing are driving the growth of the artificial intelligence market as improved appropriate technologies is critical to offer enhanced drones, self-driving cars, robotics, etc. Today, more and more sensor manufacturers are using machine learning to sensors and signal data for analyses. The machine learning for sensors and signal data is becoming easier than ever: hardware is becoming smaller and sensors are getting cheaper, making Internet of things devices widely available for a variety of applications ranging from predictive maintenance to user behavior monitoring. Whether we are using sounds, vibrations, images, electrical signals or accelerometer or other kinds of sensor data, we can build now richer analytics by teaching a machine to detect and classify events happening in real-time, at the edge, using an inexpensive microcontroller for processing - even with noisy, high variation data.

The global artificial intelligence market size was valued at USD 93.5 billion in 2021 and is projected to expand at a compound annual growth rate (CAGR) of 38.1% from 2022 to 2030. Artificial intelligences currently transforming the manufacturing industry. Virtual reality, automation, Internet of Things (IoT), and robotics are some important features of AI that are benefitting the manufacturing industry. AI has been one of the fastest-growing technologies in recent years. The market growth is mainly driven by factors such as the increasing adoption of cloud-based applications and services, growing big data, and increasing demand for intelligent virtual assistants. The major restraint for such market is the limited number of AI technology experts.

The Series of ASPAI Conferences have been launched to fill-in this gap and to provide a forum for open discussion and development of emerging artificial intelligence and appropriate signal processing technologies focused on real-word implementations by offering Hardware, Software, Services, Technology (Machine Learning, Natural Language Processing, Context-Aware Computing, Computer Vision and Signal Processing). The goal of the conference is to provide an interactive environment for establishing collaboration, exchanging ideas, and facilitating discussion between researchers, manufacturers and users. The first ASPAI conference has taken place in Barcelona, Spain in 2019, the second and the third – in the virtual format due to the COVID-19 pandemic – in Berlin, Germany in 2020.

The conference is organized by the International Frequency Sensor Association (IFSA) - one of the major professional, non-profit association serving for sensor industry and academy more than 20 years, in technical cooperation with media partners – IOS Press (journal *'Integrated Computer-Aided Engineering'*) and World Scientific (*International Journal of Neural Systems*). The conference program provides an opportunity for researchers interested in signal processing and artificial intelligence to discuss their latest results and exchange ideas on the new trends.

I hope that these proceedings will give readers an excellent overview of important and diversity topics discussed at the conference.

We thank all authors for submitting their latest work, thus contributing to the excellent technical contents of the Conference. Especially, we would like to thank the individuals and organizations that worked together diligently to make this Conference a success, and to the members of the International Program Committee for the thorough and careful review of the papers. It is important to point out that the great majority of the efforts in organizing the technical program of the Conference came from volunteers.

Prof., Dr. Sergey Y. Yurish
ASPAI' 2022 Conference Chairman

(0001)

Image Processing for Digital Twin Technology

J. D. Hobby
Nokia Bell Labs, Murray Hill, New Jersey, USA

Summary: Consider the problem of constructing a digital twin by taking a two-dimensional CAD drawing such as a building floor plan, and augmenting it with new 3-D data. The new 3-D data is a point cloud in some unknown coordinate system. It can come from Laser range finding or from comparing common features in multiple images from a video stream, but it might also be useful to match 2-D data against a CAD drawing.

An algorithm that has worked well on real world data is first to analyze the point cloud to decide which way is up, then apply polyline-based matching. The vertical alignment algorithm involves optimization of a "how well clustered" metric. The polyline-based matching is applied after projecting the point set onto a floor plane, and analyzing vertical histograms to decide which parts of the projected image most resemble walls. The vertical histograms are also used to correct for residual tilt due to inaccuracy in the "which way is up?" result.

Keywords: Digital twin, 3-D point cloud, Polyline fitting, Optimization, Vertical histogram.

1. Introduction

We are given up to a few hundred thousand 3D points in an unknown coordinate system. The points are typically on floors, walls and ceilings of some building, and we have a CAD floor plan for the building. Although CAD data may be available in vectorized formats, it is safer to assume that we have a computer-generated image. For instance, I have seen CAD-generated floor plans that have been updated via editing tools that operate on PNG files.

The 3-D points are presumably tagged with RGB values and/or other information in order to facilitate the construction of a digital twin. However, this paper focuses exclusively on the problem of registering the point cloud to the CAD floor plan image. Section 2 explains how to decide on the direction in the point cloud coordinate system that should be treated as vertical. This is basically an optimization problem that expects the point cloud to have tight clusters when projected onto the proposed floor plane.

Once we know the "up" direction, we must match wall-like structures in the point cloud to the CAD data. Some of those points are likely to correspond to furniture and/or clutter, so Section 3 analyzes vertical histograms in order to find floors, ceilings and walls.

Since Section 3 assumes there is CAD data to match against, it is a somewhat easier problem than recognizing structures in 3D point clouds (See [4, 6].) On the other hand, these authors tend to assume that the "up" direction is already known.

2. Which Way is Up?

The point cloud is given in an arbitrary coordinate system, so we need to align it with gravity. Consider what happens to the 3D point cloud if we choose the proper "up" vector, **U**, and project all the points on to a perpendicular plane. Points on walls will cluster near a few straight lines as shown in Fig. 1.

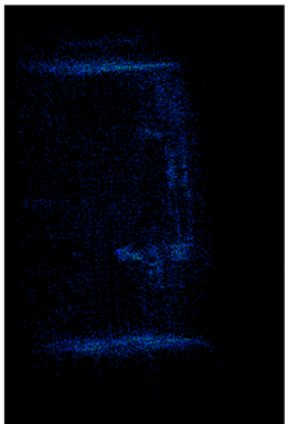


Fig. 1. A projected image based on a 3D point cloud that was derived from comparing common features in images from a video stream.

To optimize the **U** vector, we need a function that measures how tightly the points are clustered in an image like Fig. 1. We want a function $R_6(\mathbf{U})$ that estimates the distance from a 2D projection of a point from the cloud to its 6th nearest neighbor. (The parameter 6 is not very important, but it makes the function a little more well behaved.) For efficiency, we snap all the projected points to a grid, where the grid spacing is a few times smaller than what we expect for good values of $R_6(\mathbf{U})$. This small grid was 3 cm for our tests. The heavily populated portions of the projected point set can be analyzed by providing a 3 cm grid with a hit count for each grid cell. Since overall dimensions can be as large as 100 m, sparser portions of the projected 2D world can use a coarser bucketing scheme for the 6th nearest neighbor calculation.

2.1. Using $R_6(\mathbf{U})$ to Find Which Way is Up

Function $R_6(\mathbf{U})$ is hard to optimize because it does not vary continuously as the **U** vector changes, and you cannot change **U** very much before neighboring $R_6(\mathbf{U})$

values start to lose their correlation. Furthermore, optmizing the function is not sufficient, because a 3D point cloud for a typical building will give low $R_6(\mathbf{U})$ values when **U** is off by 90 degrees.

We begin by evaluating $R_6(\mathbf{U})$ for a complete constellation of **U** values. Consider the octahedron $|x| + |y| + |z| = 1$, and subdivide each edge into some number of parts n (suggested value $n = 10$). The **U** values for one face of the octahedron are

$$\left(\sin\alpha_i \cos\beta_{ij}, \sin\alpha_i \sin\beta_{ij}, \cos\alpha_i\right), \quad (1)$$

where $0 \le i \le n$, $j \le n$-i and $\alpha_i = \pi/2 - i\pi/2n$ and $\beta_{ij} = j\pi/2(n-i)$. Since negating **U** has no effect on $R_6(\mathbf{U})$, we need only four other faces of the octahedron, and they can be obtained via simple transformations of (1).

Find the best $R_6(\mathbf{U})$ value from the octahedral constellation and use Nelder-Mead optimization to look for better **U** vectors nearby [5]. This method is appropriate for nondifferentiable functions in low-dimensional spaces (2 degrees of freedom due to scale indolence in the **U** vector).

Now we have an optimal **U** vector, $\boldsymbol{U}_0$, but it might be 90 degrees away from the desired "up" vector. Scan the octahedral constellation for the best $R_6(\mathbf{U})$ among all **U** vectors in the constellation that are roughly orthogonal to $\boldsymbol{U}_0$, apply Nelder-Mead and call the result $\mathbf{U}_1$. Then find a **U** vector perpendicular to both $\boldsymbol{U}_0$ and $\boldsymbol{U}_1$, apply Nelder Mead optimization and call the result $\boldsymbol{U}_2$.

Now we have three suggested ``up'' vectors and we need a new way of choosing among them. A pretty good assumption is that most of the points in the cloud should be within 2 or 3 meters of the floor, and the room horizontal dimensions can be significantly larger Hence, for each **U**, we find the 5th and 95th percentiles of the **U** components of the 3D cloud points. The true up vector is the $\boldsymbol{U}_k$ that minimizes the separation between the 5th and 95th percentile planes.

3. Find Walls and Match with CAD Data

We need to find wall-like structures in the 3D point cloud after rotating it based on the U vector from Section 2. The CAD data is basically a collection of wall polylines, and it should be available in this form. However, it may well be necessary to extract polylines from a digital image. Reference [2] explains how to do this in a computer-generated image, and [3] explains how to cope with noisy images.

The rotated point cloud can be divided into buckets based on (x, y) coordinates, and we can collect a cumulative distribution function (CDF) for each bucket that shows how the z coordinates are distributed. Fig. 2 shows a typical result from this process. The raw CDF's at the bottom of the figure have been grouped into 7 clusters. The cluster representatives appear on a row above, and each cluster representative is surrounded by a colored square. These colors appear in the top part of the diagram with a bold colored dot for each (x, y) bucket. The green dots happen to mean "there is a floor here but no wall".

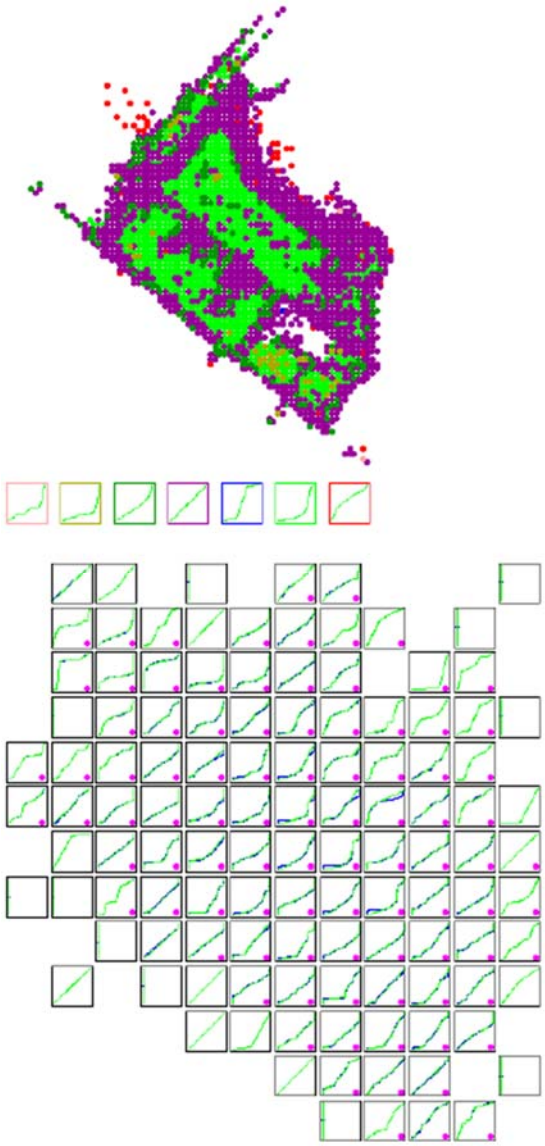


Fig. 2. Vertical CDF's from a 3D point cloud that has been rotated based on the "up" direction from Section 2. The colors in the top half of the diagram illustrate a clustering process. The lower half is a low-res picture with CDF diagrams in place of pixels.

3.1. Clustering CDFs for Floor and Wall Detection

The clustering strategy is to convert each CDF into a simplified form, then define a distance metric based on the simplified form and apply single linkage clustering [7]. The simplified form consists of cut values $c_0 \le c_1 \le \cdots \le c_m$ and fractions $f_0, f_1, \cdots f_{m-1}$ where $m \le M$ tells what fraction of the points lie between c_i and c_{i+1}. (The suggested value is $M = 8$).

Suppose we have a sorted sequence of z values, and we want to choose the parameters of the simplified form so as to fit this as well as possible. The cost to be minimized is the total squared error from a sequence

of linear regression problems. There are M such problems, and the i-th problem involves the z values that fall between $c_i$and c_{i+1}. If the number of z values is k, we can use dynamic programming to fill in an m by k matrix, where each entry in a row $0 < j \leq m$ gives an optimal total cost so far and an index that tells which entry to use in row $j-1$. This takes time $O(Mk^2)$ to find the simplified form for the CDF of k z values.

Notice that the simplified from of a CDF can be thought of as a piecewise-linear approximation: the i^{th} vertex is $(c_i, f_0 + \cdots + f_{i-1})$. The distance metric for clustering such functions is first to rescale them so that the last vertex is (1,1), and then to compute the symmetric difference; i.e., the integral between 0 and 1 of $|F(x) - G(x)|$. To get the actual clusters, use this distance metrix in the SLINK algorithm [7].

3.2. Correct for Tilt and Find Floor and Ceiling

We now have rotated the point cloud so z should be vertical, and we have buckets based on (x, y) with a simplified CDF that describes the z coordinates in each bucket. Furthermore, the simplified CDF's have been clustered. For instance, the green dots at the top of Fig. 2 gives the (x, y) positions of buckets in one cluster. Each member of such a cluster has a slightly different set of $c_i$values and $f_j$values, but we can get a canonical version by simply averaging them. For each member of the cluster, we can also find a z shift amount that makes its $c_0 + c_1 + \cdots + c_m$ value agree with the canonical version. In other words, there are buckets based on (x, y) and each bucket has a desired z shift.

Use linear regression to fit a plane $z = ax + by + c$ to these z shift amounts. This leads to a revised **U** vector, but it is not necessary to rerotate the 3D point cloud. It suffices to apply a z shift to each simplified CDF; i.e., add some Δz to each c_i.

The floor plane is based on the assumption that one of the 0.5-meter x, y buckets will have a simplified CDF that is heavily dominated by floor points. Find the 90th percentile z value for each such bucket, and use the lowest such result to define the floor plane.

The so-called ceiling plane is really a z value near the top of buckets that appear to be devoted to wall points. Among all buckets whose 20th percentile z value if above the floor plane, find their 90th percentile z values, and compute a weighted average. The weight is the number of points in the bucket times the amount by which the 20th percentile z value exceeds the floor plane.

3.3. Wall Segments and Line Statistics

The CAD data is basically a line drawing of a building floor plan, so it is easy to get wall line segments. As explained above, it may be necessary to use [2, 3] if the CAD data is given as a PNG image.

In order to find 2D transforms that appear to align wall-like features from the point cloud with the CAD-based wall segments, we need a 2D collection of points that appear to belong to walls. Fig. 3, shows a weighted collection of wall-like points with connections based on 2 nearest neighbors.

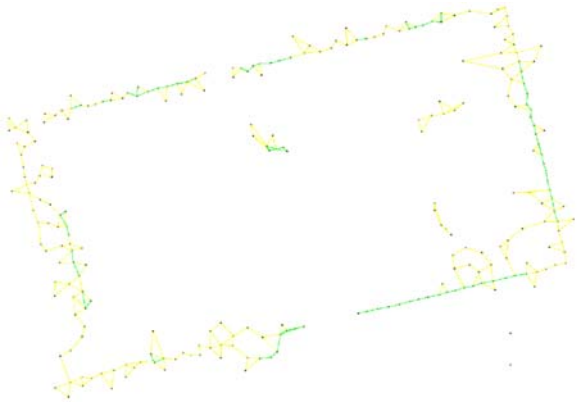


Fig. 3. Wall-like points from the coarsened point cloud, where green points are weighted higher than 100. Connections are based on the two nearest neighbors.

The main idea is that each point p in the cloud (after rotating so z is vertical) has weight $(z_c - z_p)(z_p - z_f)$, where z_f and z_c are the floor and ceiling heights. Nearest neighbors within such a weighted point set provide a way of computing, for any line $ax + by = c$ a quantitative metric for line-like features in the point cloud that are within 60 cm of the line. It is easy to compute a similar metric for the CAD floorplan data.

Experience has shown that it is compute-intensive and unnecessary to consider the entire point cloud when looking for a 2D transform that matches it to the CAD data. Hence, we use a scheme that resembles K-means clustering to reduce the point cloud to a coarser one; e.g., from 260770 points to 6317.

Start with a hexagonal grid of seed points spaced to that neighbors are 40 cm apart.

Associate each point from the fine cloud with the nearest seed point. Use a k-D tree [1].

Readjust each seed point to be the centroid of its set of fine cloud points.

Repeat Steps 2 and 3 until the seed points stop changing.

Now each point in the coarse cloud has a weight equal to its number of fine cloud points. The 2D transform candidates can be based on statistics for what types of lines are present in the CAD dataset and the point cloud dataset. Line length versus direction in 1-degree increments gives an array of 180 numbers. Use a greedy algorithm to find $C = 10$ non-overlapping intervals of length at most 18: Find the heaviest interval of length 18, then find the heaviest legal nonoverlapping interval, etc.

For each of the C line direction categories, scan all the lines assigned to that category and find the average of the orientation vectors. This makes it possible to define a set of possible line positions for each category:

parallel lines of the chosen orientation that are 60 cm apart. Thus, if we scan the line segments and assign them to the nearest possible position, we have a total line length for each of the C categories and each possible line position. These are the line statistics. We have them for each of the two datasets, and they are the basis for 2D transformations that match the point cloud dataset to the CAD dataset.

3.4. Transform Point Cloud to Match CAD Data

We need a 2D transform with rotation, translation and (within a modest range) rescaling. There are C^2 ways to rotate so one of the line direction categories from the point cloud dataset aligns with a direction category from the CAD dataset. For each one, we can consider a range of values for the scale factor σ, and a range of values for the bias b. Legal 2D transforms are described by a formula $P_{cad} = \sigma P_{cl} + b$, where P_{cad} and P_{cl} are line position parameters. Quantize the σ values into 63 possiblities, and quantize the b values at 20 cm spacing. These quantized σ and b values can be the indices into an array that gives a score for each possible 2D transform. To fill in the array, we need 3 nested loops: the 63 scale factors and all lines of the appropriate category from both data sets.

We can collect a set of candidates 2D transforms by looking for large values in the array and expecting to find a few neighboring array entries that should be thought of as a unit. For each σ value, look for triples of consecutive b values where the sum of those three array entries is a relative maximum: shifting the triple by one notch either way reduces the sum. Once such a relative maximum is found, look at neighboring σ values and consider shifted versions of the triple of b values at the neighboring σ value. It suffices to shift by at most one array position. After finding a transform candidate in this fashion, the relevant array entries should be zeroed out.

A shortcoming of the above process for candidate transforms is that the line statistics are based infinitely long lines and we never really specified which part of the point cloud dataset line matches which part of the CAD line. We can look at the individual contributions to the line statistics in order to get a rough idea, but this still results in large uncertainties in the transform candidates. The solution is to remember that we had C^2 pairs of line categories, and we have candidate transforms for each pairing.

Each of the *C* line categories in the point cloud data corresponds to a specific line direction, and the same is true for the CAD data. Hence, for any pair of categories (i, j), the candidate transforms all have the same rotation. Furthermore, it is likely that some other pair (i', j') will share that rotation angle. This is because buildings have a lot of 90-degree angles. Hence, we can get better candidate transforms:

For each pair of line direction categories (i, j) look for pairs (i', j') that give almost the same rotation angle.

Look for transforms from the (i', j') list that agree is scale factor with an (i, j) transform and have compatible shift amounts.

The compatibility test in Step 2 involves the transform uncertainties. For each compatible pair, output a new transform with reduced uncertainty.

3.5. Select Improved High-scoring Transforms

For each candidate transform, we need to test the transformed point cloud with weights $(z_c - z_p)(z_p - z_f)$, where z_f and z_c are the floor and ceiling heights. The score is the weighted sum of D_p^2, where $D_p$is the distance to the nearest wall in the CAD data, dividing by the total weight and taking the square root gives RMS matching error.

Efficiently scoring a large number of candidates transforms requires precomputed distances. We can do this at the pixel resolution of the CAD floorplan image file (or a 5 cm grid if we are lucky enough to have polyline CAD data). It helps to apply bilinear interpolation to these precomputed distances. Bilinear interpolation makes the total weighted squared error a differentiable function of the transform matrix, so we can use gradient steps to locally optimize the 2D transform parameters before scoring. We do 3 gradient steps with an extra function evaluation for each gradient step in order to decide how far to go in the gradient direction.

4. Results and Conclusions

Two data sets were available for testing, and the larger one differed only slightly from data used when developing the algorithm. Fig. 4 shows how the algorithm's top-choice transformation causes the 3D point cloud to fit the CAD floorplan data. The CAD data was given as a.png image, and it describes one floor of a large building that is 129 m by 165 m. The 3D point cloud had 266770 points and covered a 30 m by 37 m rectangle with a few outliers. The coarsened point cloud for this dataset had 6317 points.

The other dataset covers a rectangular room whose outside dimensions are 12.6 m by 7.0 m. Its point cloud has 7269 points with 268 points in the coarsened point cloud. The "CAD image" for this dataset was manually created and omitted semi-permanent structures in the room's interior. Given an overlay image like Fig. 4 for this room, it is hard for even a human observer to see if the transformation is correct or 180 degrees off.

Sections 2 and 3 of this paper were implemented as separate C++ programs, running single-threaded on a Linux server 6 years old. Processing of the large dataset took 120 sec for Section 2 and 1070 sec for Section 3. For the small dataset, the times were 5.49 sec and 1.86 sec.

There were no problems with wrong "up" vectors from Section 2, except that the Section 3 algorithm had to consider mirror image transforms due to the "up" vector possibly being negated. The Section 3 program

was designed to report reasonable alternatives to its preferred answer. For the large dataset, it reports 5 alternatives with RMS matching error ranging from 14.2 cm to 15.2 cm. The top-choice answer scores 9.7 cm, and Fig. 4 illustrates its correctness. For the small dataset, the program reports a top choice and 2 alternatives with RMS match errors 5.5 cm, 6.1 cm and 6.3 cm. The third choice is correct.

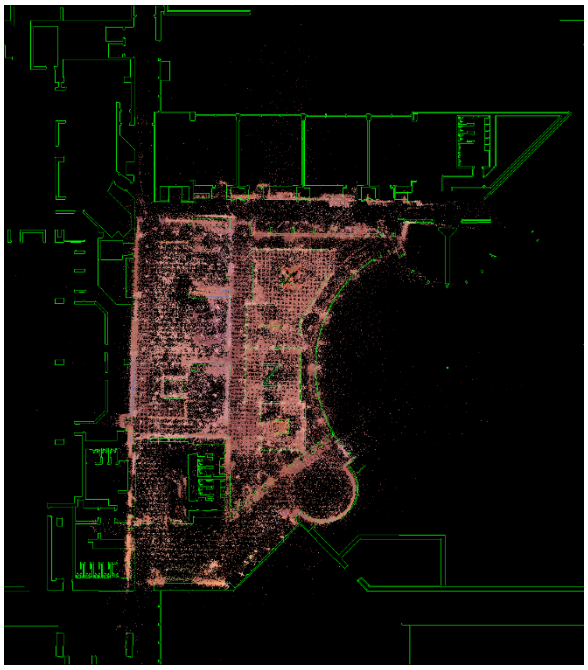


Fig. 4. The transformed point cloud in pink overlaid on CAD data in green. This gives RMS matching error of 9.7 cm. The 2nd best transform gives 14.2 cm.

We can make the following conclusions: First, the "up" vector from Section 2 seems to be reliable, except that it may be necessary to modify the final step of choosing the alternative that minimize the separation of the 5th percentile height and the 95th percentile height. Secondly, a web GUI was developed so a human can choose between various alternative matching transforms from Section 3. Thirdly, run time can grow nonlinearly with the problem size, and it is hard to characterize in a meaningful way. For instance, the large dataset has a CAD image that covers a far bigger region than the point cloud. Searching this large area increases the run time. Fourthly, registered floorplans and cloud maps can find various applications in robotics, surveillance, extended reality, and digital twin scenarios

References

[1]. J. L. Bentley, K-d trees for semi dynamic point sets, in *Proceedings of the 6th Annual Symposium on Computational Geometry (SoCG'90)*, 1990, pp. 1878-197.

[2]. J. D. Hobby, Smoothing digitized contours, in Theoretical Foundations of Computer Graphics and CAD (R. A. Earnshaw, Ed.), *Springer Verlag*, 1988, pp. 777-793.

[3]. J. D. Hobby, Polygonal approximations that minimize the number of Inflections, in *Proceedings of the Fourth Annual Symposium on Discrete Algorithms (SODA'93),* Jan. 1993, pp. 93-102.

[4]. K.-S. Kim, J.-S. Park, Cuboid-shaped space recognition from noisy point cloud in indoor AR workspace, in *IEEE Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct'21)*, 2021, pp. 512-513.

[5] J. A. Nelder, R. Mead, A simple method for function minimization, *Computer Journal,* Vol. 7, 1965, pp. 308-313.

[6]. B. Noujoks, H.-J. Wuensche, An orientation corrected bounding box fit based on the convex hull under real time constraints, *IEEE Intelligent Vehicle Systems*, June 2018, pp. 415-420.

[7]. R. Sibson, SLINK: An optimally efficient algorithm for the single-link clustering method, *Computer Journal,* Vol. 16, 1973, pp. 30-34.

(0362)

Monitoring Real-time Personal Activity and Recognition Based on Fusion and Complex Feature Extraction of Multimodal Sensors

D. Teran-Pineda and E. Dominguez
[33]University of Málaga, Dept. of Computer Science, 29071, Malaga, Spain
E-mails: {dfteran, enriqued}@uma.es

Summary: Doctors widely use gait analysis to detect anomalies and conclude possible treatments for the patient. Despite the extensive applications, human activity classification and recognition remain inaccurate. To improve the detection and accuracy of human activity recognition, we create a novel model classification and system acquisition data with a recursive algorithm, reducing the complexities of extracting features. As a result, we created a dataset of fifteen participants wearing three inertial sensors and were selected by a specialist. Our model extracts a power spectre; due to the high dimensionality of features, the sliding windows techniques are used to determine frequency bandwidth automatically, where the QRS algorithm selects the first dominant spectrum amplitude. A feature selection algorithm and PCA has been used to improve feature extraction. We have considered widely used classifiers, which we evaluated in the HuGaDB dataset and validated in our dataset, achieving an accuracy rate of 100 % in both datasets.

Keywords: Multimodal sensor, Motion classification, Computational intelligence, Complex feature extraction, Activity recognition, QRS algorithm.

1. Introduction

Human gait is one of the most common and natural activities that involves the combined effort of the brain, nerves, and muscles. It is necessary to understand the movements of humans (kinematics). The forces (kinetics) that are applied in human joints [1], in addition, physiotherapists, orthopedists, and neurologists examine the human tendency to analyse and assess the status, treatment, and rehabilitation of patients. Various extrinsic, internal, psychological, and pathological physical elements influence normal gait in regular human activities, making identifying normal gait parameters a tough endeavour [2].

There has been a rising desire to improve community health and safety. In remote health monitoring and caregiving systems, reliable and timely recognition of daily human activities throughout the day is necessary [3]. For example, only cases of injuries in older individuals [1] indicate that unintentional or untreated falls are the top cause of death (56 % in 2014) in the United States [4], including 33 % of Australians, 19.3 % in Hong Kong, 21.6 % in Bridgetown (Barbados), 29 % in Havana (Cuba), and 33 % in Mexico City (Mexico) and 34 % in Santiago (Chile). When traumas and pathologies in older persons are not adequately treated, it is vital to have a reliable remote health monitoring system that recognises motion patterns and human activities, among other uses.

A large number of methodologies have been published about (1) acquiring body signals using sensors, (2) determining features, and (3) they have applied classification methods with Artificial or Computational Intelligence. Body-worn inertial sensors (multimodal sensor) have gained wide research attention and encouraged the development of novel Human Activity Recognition applications, these applications include e-health rehabilitation, security surveillance, emergency services, wellness aid, smart homes and biofeedback systems, gait analysis, motion symmetry studies, or monitoring human activities [5], each of these applications require continuous monitoring. Nevertheless, accurate and efficient activity recognition requires selecting the most relevant features or removing redundant information.

A small number of research using frequency measures [6] and entropy have produced encouraging results when extracting more complicated features. [7] employed specific frequency power components as characteristics to distinguish normal and ataxic gait using a perceptive neuron system for data collecting with a maximum accuracy of 98.59 per cent, although it required several expensive sensors. There have also been studies that reduce sensors to reduce computing time; for example, in [8], the author directly used features from raw acceleration data as input to the classifier and did not extract features as a pre-processing step to save training time.

The main focus of this paper is to improve the detection and accuracy of human activity recognition (HAR); we create a novel model classification and system acquisition data with a recursive algorithm named DAS System, reducing the complexities of extracting features. As a result, we created a dataset of fifteen participants wearing three inertial sensors and were selected by a specialist. Our model extracts a power spectre; due to the high dimensionality of features, the sliding windows techniques are used to determine frequency bandwidth automatically, where the QRS algorithm selects the first dominant spectrum amplitude. For performance evaluation, we applied the proposed model to our Database, we also applied in a public benchmark dataset named Human Gait Database (HuGaDB) and compare it with another research.

2. Methodology

2.1. Signal Acquisition

The DAS system acquires information from two IMU sensors and a force sensor (FSR). The IMU sensor has i2c communication, and the FSR send ADC information. All data are acquired for the microcontroller and are digitally processed to be transmitted via wireless communication from a PC. The PC receives all the data information. To solve the measurement error, a discrete estimation algorithm is developed; it separates the noise based on the state space and does not require a cut-off frequency; this algorithm is called Kalman Filter (KF).

2.2. Signal Feature

In this work, we intend to use the raw accelerometer data as experiments in our proposed methodology; in addition, the raw features can still have nuances unintentionally filtered by the parameter extraction process. The system selects the activity. Also, we designed: 1) an algorithm to get relevant features, 2) an algorithm to feature extraction, 3) an algorithm to extract QRS complex from the spectrum, and 4) an algorithm to reduce the dimensionality of features using the sliding windows techniques.

3. Results

The measured samples are used to produce features, which are then fed into several classification models, including decision trees (DT), naïve Bayes (NB), support vector machine (SVM), k-nearest neighbor (k-NN), kernel approximation (kA), and neural networks (NN) (NN). The data is randomly split into five folds in the 5-folds cross-validation (CV) process. The average accuracy, sensitivity, and specificity of all five folds are used to assess performance. Table 1 represents the comparison result between the proposed method and other states of the art. research.

Table 1. Comparison of recognition accuracy.

	Proposed Method		Other Method	
Method	**DAS (%)**	**HuGaDB (%)**	**IM-WSHA (%)**	**HuGaDB (%)**
DT	98.40	100		
NV	78.80	100		
SVM	69.70	100	72.84	85.68
kA	90.20	100		
NN	85.20	100	80.13	91.76

4. Conclusions

The DAS acquisition system can be used in both legs; for this experiment, it was used in the dominant leg of each patient, and the database was obtained without any problem. In addition, the system automatically selects the acceleration data to be able to use our method. During the experimentation, we used the HuGaDB database because other investigations also used it; this helped to compare the performance with different methods and correspond with other designed systems that use inertial sensors. In our method, you can also modify the window size or sensor fusion to reduce feature extraction, reducing the amount of data and time in machine learning training. It is not recommended to decrease by less than 70 % because the results are not excellent.

References

[1]. D. Teran, E. Dominguez, Human gait model based on a machine learning and filtering noisy signals with recursive algorithm, in *Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (BIBM'20)*, 2020, pp. 1142-1145.

[2]. Y. Song, J. Zhang, L. Cao, M. Sangeux, On discovering the correlated relationship between static and dynamic data in clinical gait analysis, *Machine Learning and Knowledge Discovery in Databases*, Vol. 3, 2013, pp. 563-578.

[3]. A. A. Badawi, A. Al-Kabbany, H. A. Shaban, Sensor type, axis, and position-based fusion and feature selection for multimodal human daily activity recognition in wearable body sensor networks, *Journal of Healthcare Engineering*, Vol. 2020, 2020, 7914649.

[4]. Centers for Disease Control and Prevention, Web-based Injury Statistics Query and Reporting System, www.cdc.gov

[5]. R. Voicu, C. Dobre, L. Bajenaru, R. Ciobanu, Human physical activity recognition using smartphone sensors, *Sensors*, Vol. 18, 2019, 458.

[6]. M. Z. Uddin, M. M. Hassan, Activity recognition for cognitive assistance using body sensors data and deep convolutional neural network, *IEEE Sensor Journal*, Vol. 19, Issue 19, 2019, pp. 8413-8419.

[7]. D. Phan, N. Nhan, P. Pathirana, M. Horne, L. Power, D. Szmulewicz, Quantitative assessment of ataxic gait using inertial sensing at different walking speeds, in *Proceedings of the 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC'19)*, 2019, pp. 4600-4603.

[8]. A. Nait Aicha, G. Englebienne, K. van Schooten, M. Pijnappels, B. Kröse, Deep learning to predict falls in older adults based on daily-life trunk accelerometry, *Sensors*, Vol. 18, Issue 5, 2018, 1654.

(0458)

Unsupervised Text Clusterisation to Characterize Adverse Drug Reactions from Hospitalization Reports

Xuchun Zhang [1,2], Milou-Daniel Drici [1,3] and Michel Riveill [1,2]
[1] Université Côte d'Azur, France
[2] CNRS, INRIA, France
[3] CHU Nice, France
E-mails: {xuchun.zhang, michel.riveill}@inria.fr, drici.md@chu-nice.fr

Summary: The detection of Adverse Drug Reactions (ADRs) in clinical records plays a pivotal role in pharmacovigilance (PhV). Achieving near-ideal practice relies on well-trained health professionals, who are trained to identify, assess, and report to health authorities ADRs occurring after drug marketing approval, including those that are infrequent. However, the number of experts trained in this practice is low and despite reporting ADRs being mandatory for healthcare professionals, pharmacovigilance still suffers from a significant under-reporting, accounting for only 5-10 % of all ADRs. Yet, drug safety is crucial for assessing the benefit/risk ratio of a given drug. It is therefore important to circumvent under-reporting and to be able to collect ADRs automatically from medical reports. The most natural approach would be to train a model in a supervised manner, which requires annotation of a large volume of data, but this is unfortunately not possible. We therefore propose here an unsupervised approach to distinguish between ADRs-related and non-related reports. From a more formal point of view, we address this problem as a clustering task aiming at distinguishing medical reports containing the description of an ADR from those without.

Keywords: Text clustering, Unsupervised learning, Adverse drug reactions.

1. Introduction

Pharmacovigilance (PhV), by its definition from World Health Organisation (WHO), is "the science and activities relating to the detection, assessment, understanding, and prevention of adverse effects or any other medicine/vaccine related problem" [1], which concerns drug regulatory to ensure that the authorities of medical products are well studied on safety issues in everyday practice. Whereas rigorous testing must be done during the drug development program before its marketing approval, the issue of safety is not absolute. One of the reasons is that the clinical trials involve a relatively small number of quite selected participants comparing to the large potential number of patients who will use the drug in real life. Another reason is that these trials are conducted within a limited time frame, which precludes the characterization of certain chronic adverse reactions that may occur over a longer period.

Managing Adverse Drug Reactions (ADRs) is one of the most important post-marketing PhV practice, since serious ADRs are thought to be responsible for 5-10 % of hospitalisations. Pharmacovigilance aims at detecting and monitoring ADRs in real life settings, and more frequently nowadays, from hospital clinical reports or Electronic Health Records (EHRs), owing to the rich information about patient health and the structured textual content that were written by professionals working in the domain. After review, confirmation, and causality assessment by trained pharmacologists, this detection will be recorded in the National Agency for the Safety of Medicines and health products (ANSM)[1]. Then, data from the national database will be fed into the VigiBase database at the Uppsala Monitoring Centre (UMC). VigiBase is the unique global database of WHO (World Health Organization) reporting potential side effects of medicinal products. It is the largest database of its kind in the world, with over 30 million reports of suspected adverse effects of medicines, submitted, since 1968, by member countries of the WHO Programme for International Drug Monitoring (PIDM). It is continuously updated with incoming reports.

Achieving a more ideal practice relies heavily on well-trained health professionals, who are more likely to have sufficient experience to identify, assess and report important ADRs [2]. Despite being mandatory for health care probationers to report ADRs when suspected, notifications of ADRs amount to a mere 5-10 % of all ADRs. However, the efficiency to detecting ADRs is limited due to the lack of well-trained professionals, the underreporting and the enormous number of clinical reports at disposition.

Deep learning has boosted the development of Natural Language Processing (NLP) and showed that NLP can be a solution to practice efficiently and accurately in biological analysis, and it is getting more and more attentions from the researchers. Many shared tasks/workshops [3-5] are conducted in exploitation of possibilities of ADR detection by modern deep learning NLP techniques, which provides us with an

[1] https://ansm.sante.fr/

overview of how powerful these techniques are on the annotated corpus. Despite the good performances, the state-of-art supervised NLP techniques could achieve in ADRs detection from annotated corpus, we cannot ignore that one need a large number of annotated data to train a supervised model but getting such number of annotations is extremely expensive. On the other hand, the rapid increasing amount of EHRs without annotations are remain unexploited. To bridge this gap, we present in this paper a new unsupervised approach to help finding potential EHRs with ADRs descriptions.

2. Related Work

Because of the rarity of annotated EHRs related to adverse events and the limited public access to clinical records, given patient privacy and confidentiality. Since the first approach which try to characterize the likelihood of a candidate drug-symptom relation to be categorized as a true ADR [6], this domain has got more and more attention from NLP research community. The 2018 National NLP Clinical Challenges shared task (n2c2) [7] provided 505 discharge summaries for 3 different tasks: concept extraction, relation classification, and end-to-end systems construction, where among the best performance systems, Wei et al. [8] applied a joint-learning-based BiLSTM-CRF for both NER and RI tasks, where they conducted rule-based postprocessing to fix the obvious errors and improve the prediction. Christopoulou et al. [9] proposed a weighted BiLSTM combining a walk-based model to reasoning intra-sentence relations and a Transformer-based network to memorising inter-sentence relations. IBM Research team explored a combination of piecewise neural networks [10] and an attention-based BiLSTM. More recently, El-allay [11] proposed a joint model with transformer and Weighted Graph Convolutional Network (WGCN) to capture ADR relations and proved its state of-the-art performance on n2c2 dataset With similar tracks as 2018 n2c2 shared task, the MADE (Medications and Adverse Drug Events from Electronic Health Records) 1.0 challenge [4] provides real de-identified EHRs and corresponding annotations for medications, symptoms and ADRs. For MADE data, Chapman et al. [12] developed a two-stage approach by first identifying the named entities based on conditional random field (CRF), and then assigning the relevant relation type between entities based on random forest (RF) and achieve the highest score for Relation Identification (RI) task. Dandala et al. [13] adopted a combined bidirectional long short-term memory (BiLSTM) with CRF for named entities recognition (NER) and applied attention-based BiLSTM network together with medical domain ontology information from unified medical language system (UMLS) to RI task, which is the highest performing system in joint Relation Identification (NER-RI) task.

In recent years, the NLP community has demonstrated the great power of supervised machine learning techniques for ADR extraction. However, the unsupervised approaches still remain uncertain and under-exploited. Pérez et al. [14] first tried analysing vector representation for ADRs from EHRs written in Spanish by linking word2vec embeddings of drug-symptom entities pair in semantical space, which shows the potential of expressing correlation between ADR and non-ADR. More recently, Bampa et al. [15] explored encoding the document type without considering too much the textual content and by clustering aggregation [16] techniques to grasp information about the phenotype of patient/document, which provided decent cluster structure for ADR analysis.

3. Method

3.1. Preprocessing

We assume that for any ADR, both the drug and the adverse effect are described within the same block of textual content. We defined henceforth "block" as the basic unit of textual content to analyse, which can be either whole document, paragraph, phrase, sentence, etc.

Then we can define the problem as: Let $B = \{\beta_1, \beta_2, \ldots \beta_N\}$ with N blocks of literature, each block β_i contains textual contents together with annotations for drug and for symptom entities (In a text, it is simple to locate drugs by consulting domain ontologies that explain the molecules and trade names of those who has marketing authorisation, and symptoms have also their universally codified medical definitions). Take the block "He was better controlled on Velcade, but developed significant peripheral neuropathy" as an example, where we see the drug "Velcade" and the symptom "peripheral neuropathy" in the text.

We want to separate the blocks with the description of ADR (noted as positive block β^+ from those who don't (noted as negative block β^-). As a result, the blocks that do not include any drug or symptom will be of little interest to us. We made the hypothesis that the ADR relation lies in the contextual content between drug and symptom entities, based on which, we want to reduce the influence of drug and symptom entities and increase the model's emphasis on the context. To preprocess the drug/symptom entities, taking sentence "He was better controlled on Velcade, but developed significant peripheral neuropathy" as example, we presented four strategies:

- **Keep the entities**: "He was better controlled on Velcade, but developed significant peripheral neuropathy";
- **Replace drug entities by word 'drug' and symptom entities by word 'symptom'**: "He was better controlled on drug, but developed significant symptom";

- **Masking both drug and symptoms entities**: "He was better controlled on [MASK], but developed significant [MASK]";
- **Remove the drug/symptom entities**: "He was better controlled on, but developed significant".

We took finally the "removing drug/symptom entities" strategy to preprocess text with entities information as its best performance among the models. In this section, we describe our unsupervised ADR-related records detect system. Fig. 1 shows the overall structure of our model as well as its main components. By the definition of the ADR, it is obvious that its occurrence will always relate to a drug-symptom entities pair, and the contextual contents around the target drug and target symptom indicates its existence. Since most clinical records were generated by hospital health care practitioners, the documents bear a well-organised structure with many medical terms like medication, chemical names, symptoms, medical observations, and diagnoses etc... We assume that the source mentions for drug and symptom related entities is given, and we need to find ADR-related records. Our system takes the clinical records as inputs and process the records and apply a filter algorithm to choose the potential blocks for further purpose. Then, the blocks will be tokenized and fed to the model for unsupervised learning.

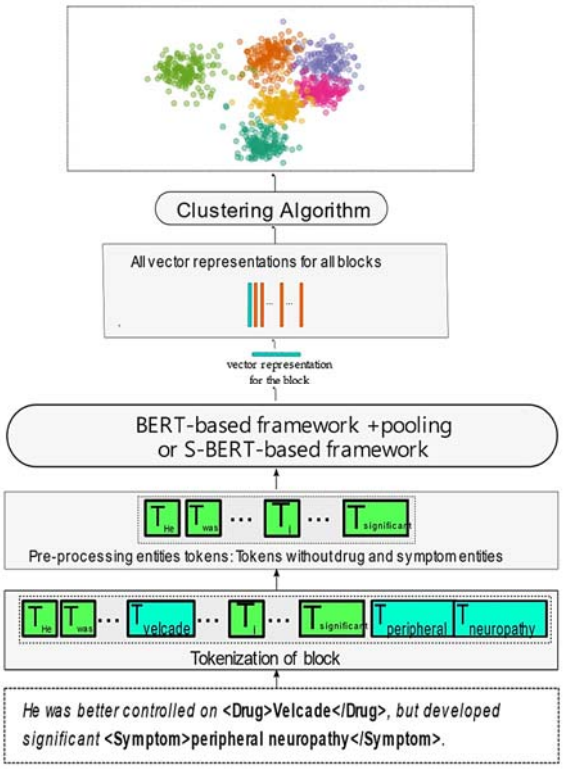


Fig. 1. The structure of our model. For each block, only the contextual tokens around drug/symptom named entities are selected for BERT-related embedding. For BERT-like models we applied a pooling strategy between tokens from the last layer to obtain a single vector representation for the block as the Sentence-BERT like models do, and then the output embedding vectors will go through the clustering algorithm to get the cluster assignment for each one of them.

3.2. Unsupervised BERT Based ADR Block Detection

The pre-trained language models, including BERT (Bidirectional Encoder Representation from Transformer) [17], a two-stage Transformer-based [18] natural language representation framework proposed by Google Brain in 2018, shows in recent years its great potential in extraction of features from textual content, which push significantly the state-of-the-art performance in many aspects in NLP domains. We here utilised the pre-trained BERT models without fine-tuning it since the latter requires a huge corpus to support.

BERT-based transformation model split each word in input text into word-piece tokens and takes the tokenized words sequence as its own input to encode each input text into vectors of the same size in the same semantic space, which means that the basic BERT-based models embed each word-piece token but not the whole sequence. As to infer a single representation for one block, we chose to applied pooling to the embedded tokens. Asides from the original BERT model, we also tried Sentence-Bert (SBERT) [19] that take BERT as basic component and considered training it with a siamese and triplet networks, in order to catch representation not for words but for the whole sequence and thus it can map directly a sentence like input to the vector space with common similarity measures like cosine-similarity. In our case, we used the pre-trained sentence encoder part from Sentence-Bert to encode the block content into one single vector representation.

We make the hypothesis that, the description of ADR-related information lies between the medication and symptom entities. We therefore chose the textual content for each block by removing all drug or symptom associated entities as the input for all language models. This input is processed by the tokenizer of the model and then the model itself to represent each input in their own way. As we mentioned above, for the output of BERT-based models, since they encode every word-piece token of a block into vectors with same size, we need to apply an extra average pooling to it to obtain one vector representation for one block as the SBERT model does. Once getting all vector representations from the language model, we applied KMedoids cluster algorithm to create clusters of similar blocks. KMedoids clustering is similar to the popular KMeans clustering algorithm, both aim to reduce the distance between points labelled as belonging to a cluster and a point designated as the cluster's centre, we choose the former due to its robustness to noise and outliers than the latter and also its flexibility with arbitrary dissimilarity measures. The ideal practice is to obtain a cluster with only positive blocks and another with only negative ones. The structure of the model is shown in Fig. 1.

4. Experimental Results

4.1. Dataset

We chose here block in sentence-level, which leads to a relatively small span of text been chosen comparing to a whole length of documentation, which also means the tokenized sequence nearly exceeds the length limit of BERT-based models. We also note that we didn't take the irrelevant examples as input for our models, such as blocks with no entities or blocks with only one type of entity. We used here two datasets: MADE [4] dataset and CSIRO Adverse Drug Event Corpus (CADEC) [20] dataset. The former is the data used in MADE 1.0 challenge [4], whose corpora collected from 21 randomly selected cancer patients at the University of Massachusetts Memorial Medical Center, with the annotations of drugs, symptoms and ADEs. We choose this dataset considering the nature of source being real life clinical notes and its high quality of annotations.

For MADE data, we sampled two sets of blocks that contains both drug and symptom (since in unsupervised system, we have no idea in advance that whether the symptom is adverse effect of the drug or the cause of taking the drug or even an irrelevant symptom) by the help of entity type information for drugs and symptoms in EHR and we can extract two datasets as following:

- **MADE multi-d-s:** All examples that contains only drugs or only symptoms were removed and thus we got a dataset where each block has at least one drug and one symptom. This dataset contains long blocks from EHR corpus with a nearly balanced distribution with 571 negative blocks and 651 positive blocks.
- **MADE 1d1s:** For the dataset above, we extract those who has exactly one drug and one symptom, which called "1d1s" as "perfect situation", a nearly balanced (with a number of 416 negative blocks and 301 positive ones) dataset with short blocks from well-written EHR corpus.

The CADEC Dataset has a rich annotated corpus of medical forum posts "Ask a Patient", which is dedicated to ADE-related consumer reviews on medications with length of several sentences. These posts are mostly written in colloquial language and often deviate from the formal rules of English grammar and punctuation. The annotations contain entities such as drugs, ADEs, symptoms and diseases related to their respective concepts in MedDRA. We performed the same pre-selection as we did for MADE data, we kept all blocks that has at least one drug and one symptom

4.2. Experimental Settings

We chose a fully supervised yet simple approach, a Bag of Words + Logistic Regression Classifier as the baseline of upper bound and a Bag of Words + completely random classifier as lower bound.

As evaluation metrics, we chose then precision, recall and F1-Score as categories of metric to evaluate the result produced. The fraction of documents retrieved that are relevant to the ADE, is known as precision, which can be given by the formula $\text{Precision} = \frac{\text{TP}}{\text{TP+FP}}$, where TP and FP represent the number of real positive examples and real negative examples among all that have been retrieved as positive examples. Recall is the number of correct results divided by the number of expected results, whose formula is $\text{Recall} = \frac{\text{TP}}{\text{TP+FN}}$, where FN indicates the number of retrieved negative examples that are really positive ones. F1-score is the harmonic mean of recall and precision, with the formula as $\text{F}_1 = 2 \times precision \times recall/(precision + recall)$. We computed the average score of a 5-fold cross validation as the final score for each term.

As we mentioned in Section 3.2, we mainly used three models to encode information from text: 1) The original BERT model "Bert-base-cased". 2) The BioBERT [21] model, who uses the same structure as the BERT model pre-trained and fine-tuned on biomedical corpora instead of employing general domain text corpora, to create a BERT model that specialises in describing features in biomedical literature. We introduced BioBERT here to verify if domain-specific knowledge has great impact on represented latent ADE information in the textual content. We used in our experiments the model "biobert-base-cased-v1.1". And 3) the Sentence-BERT (S-Bert) [19] "sentence-transformers/bert-base-nli-mean-tokens". To get fully representation for whole block for the first two models 1) and 2), as well as obtaining the corresponding block representation in high quality, we applied average pooling to the output from short block in **MADE 1d1s** dataset and extract the "cls" token for long block in **MADE multi-d-s** and **CADEC** dataset. We choose cosine similarity as the metric and use KMedoids as clustering algorithm due to its flexibility with this measure, and set number of clusters as 2, to agree with the ADE-related/non-ADE blocks.

4.3. Experimental Results

The experimental results are shown in the Table 1, from which we can see that for the MADE 1d1s dataset, compared to the lower bound, the representation provided by basic BERT model itself is not enough to capture the essential information about ADR. However, BioBERT wins BERT for its biomedical domain specified dictionary which helps it to represent better the medical text, with a highest recall value among the unsupervised methods, which means it is more reliable when we focus on retrieving more examples from the real positive ones. On the other hand, comparing to BioBERT, the S-Bert embedding + clustering strategy could achieve a higher F1 score (0.669 vs 0.657 for BioBERT) with a higher precision (0.733 vs 0.651 for BioBERT) but lower

recall value. This has showed us that the Table 2: Comparison with supervised baseline and our unsupervised approach, we report the average Precision, Recall and F1 scores of 5-folds cross validation.

As for the MADE multi-d-s data, the augmented number of textual data with longer length boost the performance for supervised baseline. Introducing more textual content means more access to potential information, but also leads to more irrelevant content being considered. As we can see from the same table, the S-BERT still stay strong in processing the sentences and thus achieve the best performance in unsupervised methods, but we can also see the drop of precision comparing to the MADE 1d1s dataset. The performance of BERT improves a little thanks to its general domain dictionary gathering more information from longer texts. Moving from short block to long block does introduce more resources which can be helpful in representing ADRs, but also makes BioBERT with average pooling representation difficult to tell the ADR information from the text, taking here the traditional "cls" token representation from last hidden layer showed us the best performance among the unsupervised methods. For CADEC dataset whose corpus contains more informal structures and spells and extremely unbalanced example distribution, the results seem less stunning as for the MADE data, but we can still observe that the strength of BioBERT in capturing features to represent an ADR correlated semantic content.

Table 1. Comparison with supervised baseline and our unsupervised approach, we report the average Precision, Recall and F1 scores of 5-folds cross validation. The results for unsupervised approaches (*) are always followed by a KMedoids clustering.

Category	Exps	MADE 1d1s			MADE multi-d-s			CADEC		
		Prec	Recall	F1	Prec	Recall	F1	Prec	Recall	F1
Supervised Classifier	BOW+LR	0.702	0.797	0.746	0.809	0.847	0.828	0.939	0.993	0.965
Unsupervised Clustering	BERT*	0.549	0.463	0.502	0.591	0.634	0.612	0.950	0.520	0.672
	BioBERT*	0.651	0.663	0.657	0.653	0.673	0.663	0.938	0.570	0.709
	S-BERT*	0.733	0.615	0.669	0.666	0.593	0.627	0.958	0.492	0.650
Supervised Classifier	BOW+Random	0.514	0.529	0.522	0.509	0.440	0.472	0.946	0.509	0.662

We have also explored taking not the context around entities but only masking the entities as input for BERT models and performed the same pipeline as we did before, and it turns out that fully removing entities remains better with respect to all datasets, which lead us to the point that the content around entities did infers the information. Even more, we trained also LR classifier with the three BERT embeddings whose results grand us confidence that this kind of representation did grasp important information in distinguish ADR and non-ADR relations. Overall, the representation provided by BERT is a helpful representation as features for ADR-related block classification, during which the whole progress is fully unsupervised, which proves potential for more future explorations.

5. Conclusion

Unsupervised learning can be a powerful resource in post-marketing pharmacovigilance, as it can exploit the big amount of data produced by daily trials of a larger populations and avoiding simultaneously the big cost of annotating data. We proposed a model to make use of modern text features extraction technique with BERT based models and explored the possibility of clustering ADR-related representations together in semantic space. The results indicate that with only contextual tokens as input, the model representation, especially those who obtained from domain-specific pretrained model like BioBERT, can be helpful in classifying ADR-related textual blocks with non-ADR blocks, especially for corpus like EHRs.

References

[1]. Institute of Medicine (US) Committee on Quality of Health Care in America, To Err is Human: Building a Safer Health System (L. T. Kohn, J. M. Corrigan, M. S. Donaldson, Eds.), *National Academies Press*, Washington (DC), USA, 2000.

[2]. The Importance of Pharmacovigilance, Safety Monitoring of Medicinal Products, *W. H. Organization*, 2002.

[3]. A. Jagannatha, F. Liu, W. Liu, H. Yu, Overview of the first natural language processing challenge for extracting medication, indication, and adverse drug events from electronic health record notes (MADE 1.0), *Drug Safety*, Vol. 42, 2019, pp. 99-111.

[4]. A. Jagannatha, F. Liu, W. Liu, H. Yu, Overview of the first natural language processing challenge for extracting medication, indication, and adverse drug events from electronic health record notes (MADE 1.0), *Drug Safety*, Vol. 42, 2019, pp. 99-111.

[5]. D. Weissenbacher, A. Sarker, A. Magge, A. Daughton, K. O'Connor, M. J. Paul, G. Gonzalez-Hernandez, Overview of the Fourth Social Media Mining for Health (SMM4H) shared tasks at ACL 2019, in *Proceedings of the Fourth Social Media Mining for Health Applications Workshop & Shared Task (SMM4H'19)*, 2019.

[6]. N. Kang, B. Singh, C. Bui, Z. Afzal, E. M. van Mulligen, J. A. Kors, Knowledge-based extraction of adverse drug events from biomedical text, *BMC Bioinformatics*, Vol. 15, March 2014, 64.

[7]. S. Henry, K. Buchan, M. Filannino, A. Stubbs, O. Uzuner, 2018 n2c2 shared task on adverse drug events and medication extraction in electronic health records, *Journal of the American Medical Informatics Association*, Vol. 27, October 2019, pp. 3-12.

[8]. Q. Wei, Z. Ji, Z. Li, J. Du, J. Wang, J. Xu, Y. Xiang, F. Tiryaki, S. Wu, Y. Zhang, C. Tao, H. Xu, A study of deep learning approaches for medication and adverse drug event extraction from clinical text, *Journal of the American Medical Informatics Association*, Vol. 27, May 2019, pp. 13-21.

[9]. F. Christopoulou, T. T. Tran, S. K. Sahu, M. Miwa, S. Ananiadou, Adverse drug events and medication relation extraction in electronic health records with ensemble deep learning methods, *Journal of the American Medical Informatics Association*, Vol. 27, January 2020, pp. 39-46.

[10]. D. Zeng, K. Liu, Y. Chen, J. Zhao, Distant supervision for relation extraction via piecewise convolutional neural networks, in *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP'15)*, Lisbon, 2015, pp. 1753-1762.

[11]. E.-d. El-Allaly, M. Sarrouti, N. En-Nahnahi, S. Ouatik El Alaoui, An attentive joint model with transformer-based weighted graph convolutional network for extracting adverse drug event relation, *Journal of Biomedical Informatics*, Vol. 125, January 2022, 103968.

[12]. A. B. Chapman, K. S. Peterson, P. R. Alba, S. L. DuVall, O. V. Patterson, Detecting adverse drug events with rapidly trained classification models, *Drug Safety*, Vol. 42, January 2019, pp. 147-156.

[13]. B. Dandala, V. Joopudi, M. Devarakonda, Adverse drug events detection in clinical notes by jointly modeling entities and relations using neural networks, *Drug Safety*, Vol. 42, 2019, pp. 135-146.

[14]. A. Perez, A. Casillas, K. Gojenola, Fully unsupervised low-dimensional representation of adverse drug reaction events through distributional semantics, in *Proceedings of the Fifth Workshop on Building and Evaluating Resources for Biomedical Text Mining (BioTxtM'16)*, Osaka, 2016, pp. 50-59.

[15]. M. Bampa, P. Papapetrou, J. Hollmen, A clustering framework for patient phenotyping with application to adverse drug events, in *Proceedings of the IEEE 33nd International Symposium on Computer-Based Medical Systems (CBMS'20)*, 2020, pp. 177-182.

[16]. A. Gionis, H. Mannila, P. Tsaparas, Clustering aggregation, *ACM Transactions on Knowledge Discovery from Data*, Vol. 1, March 2007.

[17]. J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: pre-training of deep bidirectional transformers for language understanding, *arXiv Preprint*, May 2019, arXiv:1810.04805.

[18]. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need, in Advances in Neural Information Processing Systems (I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, Eds.), Vol. 30, *Curran Associates Inc.*, 2017.

[19]. N. Reimers, I. Gurevych, Sentence-BERT: Sentence embeddings using siamese BERT-Networks, in *Proceedings of the Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP'19)*, Hong Kong, China, 2019, pp. 3982-3992.

[20]. S. Karimi, A. Metke-Jimenez, M. Kemp, C. Wang, CADEC: A corpus of adverse drug event annotations, *Journal of Biomedical Informatics*, Vol. 55, 2015, pp. 73-81.

[21]. J. Lee, W. Yoon, S. Kim, D. Kim, S. Kim, C. H. So, J. Kang, BioBERT: A pre-trained biomedical language representation model for biomedical text mining, *Bioinformatics*, Vol. 36, 2020, pp. 1234-1240.

(0482)

Robust Sensor Spike Detection Method Based on Dynamic Time Warping

F. Deuschle [1,2], B. Cornelis[1] and K. Gryllias [2,3]
[1] Siemens Digital Industries Software, 68 Interleuvenlaan, B-3001 Leuven, Belgium
[2] KU Leuven, Department of Mechanical Engineering,
Celestijnenlaan 300, B-3001, Heverlee, Belgium
[3] Dynamics of Mechanical and Mechatronics Systems,
Flanders Make, Celestijnenlaan 300, B-3001, Heverlee, Belgium
E-mail: federico.deuschle@siemens.com

Summary: In the frame of automated multi-sensor validation for automotive proving ground testing, this paper addresses the problem of detecting a specific type of signal acquisition anomaly known as "spikes". This issue can be seen in time series data as an unexpected, sharp peak with no correlation to previous samples. In addition, proving ground testing involves procedures where several types of impulsive events are encountered (i.e., sudden maneuver shifts, passing over cobblestones or sudden impacts due to potholes in the road). This leads to physical "peaks" in the measurement data which must not be confused with spike sensor faults. Thus, this research proposes a novel technique based on Dynamic Time Warping to classify spikes as either physical peaks (non-anomalous patterns) or spike sensor faults (anomalous patterns) to reduce false positive rates of an unsupervised spike anomaly detection approach for time-variant signals.

Keywords: Machine learning, Dynamic time warping, Spikes, Sensors, Data-driven, Proving ground testing, Automotive.

1. Introduction

The automotive industry demands more reliable tests to validate their digital design models. Thus, several studies are being conducted to make tests more effective and more efficient. Artificial Intelligence or Machine Learning techniques can enhance the accuracy of a given test, or to be more precise, to automatically detect any issue in the acquired signals while the test is conducted.

In automotive proving ground testing, the vehicle is traditionally heavily instrumented with different types of physical sensors that are needed to characterize the dynamical behavior of a newly designed variant of a certain vehicle. This activity is prone to have sensor issues while the test is conducted. Most common hardware problems that can be found are misconnection issues due to defective cables or sensors, ADC stuck values, EMI due to other sources functioning close to the sensors, ground loop noise, etc. All these types of hardware problems can be seen in the raw measured time series as spikes, drifts, noise, or stuck-at-constant anomalies [1].

Every type of sensor can exhibit anomalies while a test is conducted (i.e., microphones, strain gages, accelerometers, tachometer, wheel force transducers, etc.). However, especially for the case of accelerometers, they may also be sensitive to physical impulses or shock events (non-anomalous patterns) which can be confused with actual spike sensor faults (anomalous patterns). The distinction between the two cases can only be made by an experienced test engineer.

Additionally, the interest of the industry into faster testing campaigns does not match well with the idea of having to pre-train an anomaly detection model in order to get accurate results. Thus, this research assumes that no training data is available to develop an automated procedure for spike detection, which is able to distinguish physical peaks from actual spike sensor anomalies through Dynamic Time Warping.

2. Methodology

2.1. Dynamic Time Warping Algorithm

In this section, a brief introduction about the Dynamic Time Warping (DTW) algorithm will be given. DTW, which was introduced in continuous speech recognition applications in [2], allows to compare the shape of waveforms with different amplitudes or time scaling, in contrast to the traditional Euclidean distance. DTW finds the optimal alignment between two time series and captures flexible similarities by aligning the coordinates inside both sequences. Fig. 1 shows graphically how the most similar elements are linked by a grey solid line. The cost of the optimal alignment can be recursively computed by the following expression:

$$D(i,j) = \zeta(x_i, y_j) + \min\begin{Bmatrix} D(i-1,j-1) \\ D(i,j-1) \\ D(i-1,j) \end{Bmatrix} \quad (1)$$
$$i = 1,\ldots,N;\ j = 1,\ldots,M$$

where $x = (x_1, \ldots, x_N)$ and $y = (y_1, \ldots, y_M)$ are two time series and where ζ is the distance between 2 elements. These terms form *D*, a N-by-M matrix that will allow to establish the "least-costly" warping path. Intuitively, the distance function has a small value when the sequences are similar and a large one if they are different. Thus, DTW finds the optimal path that runs through the low-cost "valleys" in the cost matrix.

For this research, all DTW related calculations were executed with DTAIDistance Python package [3].

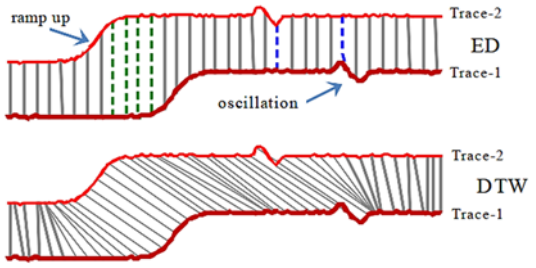


Fig. 1. Comparison between Euclidean Distance (ED) and Dynamic Time Warping (DTW). Grey lines indicate linked elements between time series [4].

As is stated in the previous section, it is important for an automated sensor validation method to be able to distinguish between physical peaks (PP) and artificial spikes (AS). We assume that the difference is observable in the waveform shapes as follows:

- *Physical peaks (PP)* show a slow oscillating decay after the maximum value is reached. The decay time and oscillation frequency are determined by the mechanical system properties (mass, damping, stiffness).
- *Artificial spikes (AS)* can be considered as outliers with no dynamic response (e.g., no correlation with previous and subsequent samples).

The acceleration magnitudes obtained in time domain signals after an impulsive force acted on a mechanical structure present a time energy decay proportional to the structural damping. For a Single-Degree-Of-Freedom (SDOF) linear system, the dynamic behavior can be expressed as:

$$mX'' + cX' + kX = 0, \qquad (2)$$

where m, c and k are the mass, damping and stiffness of the structure and X'', X' and X are the acceleration, velocity and displacement. The response of more complex real-world structures corresponds to a sum of several independent frequencies and damping ratios which leads to more complex waveform shapes [5]. Nevertheless, for the case of artificial spikes, such dynamic behavior cannot be observed. Fig. 2 shows signals from real world test data. For a timeframe of 10 ms, the difference between a PP and an AS can clearly be observed.

Based on this knowledge, template signals can be constructed which represent each signal type (cf. Fig. 3). By comparing extracted spikes with these template signals through DTW and calculating the distances, the algorithm can determine whether an extracted spike is a PP or an AS.

2.2. Pre-processing

For this case we consider a sensor time series that will be checked as an isolated signal, referred to as $x(t)$. The algorithm starts by differentiating the signal to enhance the presence of the most relevant changes in the gradient $x'(t)$. Then, Z-scores are calculated to normalize the time series, this means to subtract and divide each sample with the overall standard deviation and mean values [6]. We will define this output variable as $z(t)$. After this first processing, the signal is divided into equally sized window slices. For this case, the window size is 2048 samples and for each window slice an exponentially weighted standard deviation (EWSD) calculation is performed with a forgetting factor α of 0.3. The EWSD is calculated in the same way as the exponentially weighted moving average (EWMA) [7], i.e.:

$$\mathrm{EWSD}_k = \alpha * \hat{\sigma}_k(z) + (1-\alpha) * \mathrm{EWSD}_{k-1}, \qquad (3)$$

where k is the window slice index, α is a forgetting factor between 0 and 1, EWSD_{k-1} corresponds to the EWSD from the previous window slice and $\hat{\sigma}_k(z)$ corresponds to the standard deviation of $z(t)$ in window slice k. The results for a real-world accelerometer signal can be observed in Fig. 4. It is noticeable that the slices with peaks show higher values of standard deviation compared to neighboring peak less slices, while there is also a global increasing trend due to the nonstationary operational conditions of the measurement (i.e., a vehicle performing a run-up maneuver, cf. Section 3).

2.3. Peak Extraction

Following the previously described pre-processing procedure, an EWSD value per window slice is obtained. In order to identify potential peaks, a dynamic threshold is then set as EWSD*Factor. The value of Factor will tune the sensitivity of the algorithm to find more or less peaks to be further analyzed by the DTW method. Each peak that surpasses the threshold and the successive 125 samples are extracted and stored for further analysis. 125 samples correspond to a timeframe close to 2.5 ms for the sampling frequency used in these measurements (i.e., 51200 Hz).

This sample size of 125 is considered to represent well enough the expected response of a structure to any PP that could be confused with an AS. Moreover, this value is also still suitable for lower sampling frequencies because of the type of phenomena that we are studying. For sampling frequencies close to ~2048 Hz, 125 samples correspond to ~60 ms, a timeframe suitable to represent a PP that could be confused with an artificial spike. Fig. 5 shows an extraction of 2 peaks in which a slowly decaying behavior can be observed for peak 1, while peak 2 exhibits a more sudden energy loss (i.e., it is most likely that peak 2 corresponds to an AS). As can also be noted in the figure, the extracted peaks are scaled to unit amplitude to reduce the influence of amplitude while comparing with the reference templates (cf. next section).

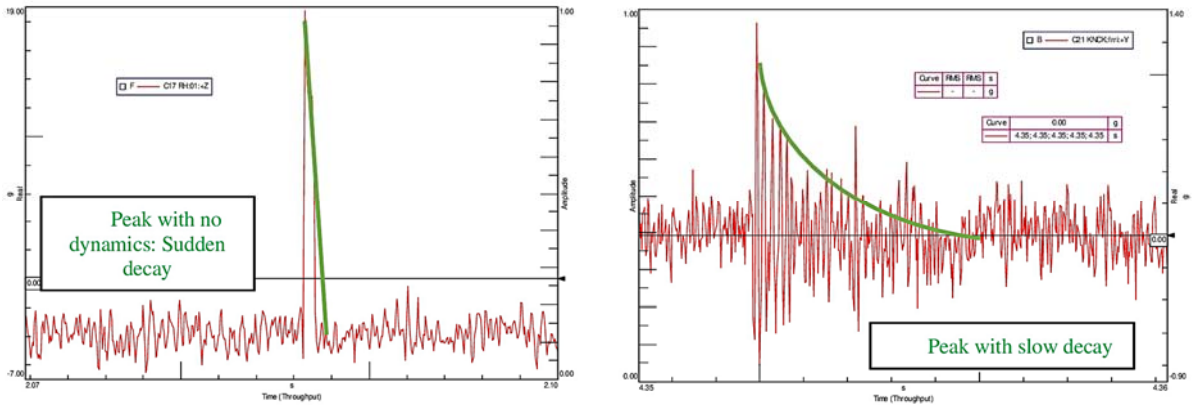


Fig. 2. Examples of artificial spikes (left) and physical peaks (right).

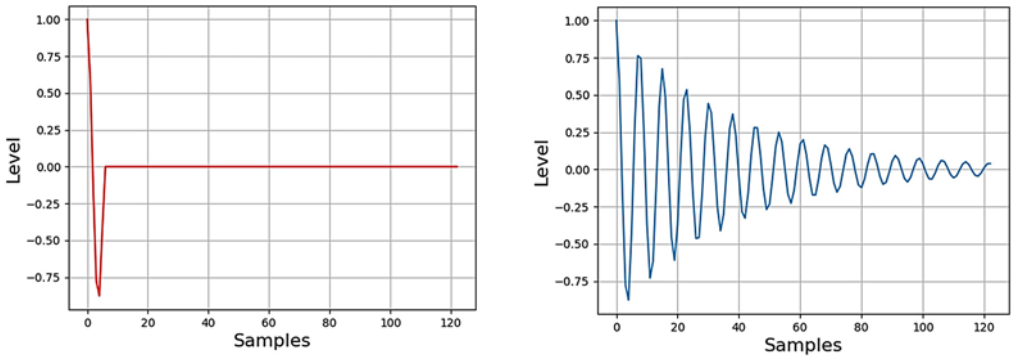


Fig. 3. Reference templates of artificial spike (left) and physical peak (right).

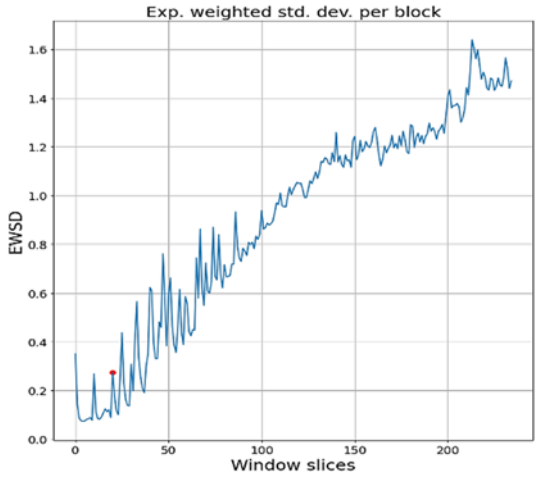


Fig. 4. Exponentially weighted standard deviation (EWSD) per window slice.

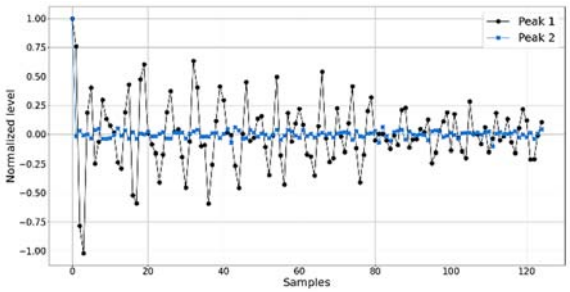


Fig. 5. Extracted peaks from a time series.

2.4. Dynamic Time Warping Block

Once the peaks have been extracted from the pre-processed time series, each one of them is compared to the reference templates (cf. Fig. 3). The reference templates are simulated time series that represent two models of the expected behavior of a PP and an AS. Since the peaks were all extracted from their maximum value and then scaled with this same value, it means that they all start at amplitude equal to 1. Nevertheless, since the frequencies and damping ratios will vary from each specimen under test, the use of Dynamic Time Warping (DTW) is essential to perform waveform shapes studies (cf. Section 2.1). The overall slow decay in amplitudes can be captured by the algorithm without a significant impact of frequency variations that can be found in different structures under test. Once the similarity metric is calculated for each template, we will have two resulting metrics for each peak. The lowest between both will define if the peak is considered as anomalous or not (e.g., if the lowest similarity metric corresponds to the comparison against the artificial spike template, then the peak will be considered as anomalous). The subtraction of these two metrics leads to the creation of the ΔDTW for each peak:

$$\Delta DTW = DTW_{artificial} - DTW_{physical}, \quad (4)$$

$$\begin{cases} If\ \Delta DTW > 0 \Rightarrow Not\ Anomaly \\ If\ \Delta DTW < 0 \Rightarrow Anomaly \end{cases} \quad (5)$$

One important aspect to remark is that since traditional implementations of DTW have a computational complexity of $O(N^2)$ [8], where N is the quantity of samples to be fed into the DTW block, it is important to keep N small. The most typical applications of DTW are the ones that search for specific patterns in extremely long time series. Depending on the length of the array, the computation times can become excessively large for an online application. Rakthanmanon *et al.* [9] have worked on a significant improvement of the computational times by discarding samples that do not comply with the expected values at the initial and ending part of the subsequence under test. Still, our research tackles the computational complexity issue in a different manner by only applying DTW to the suspicious peaks and successive 125 samples, which were obtained in the peak extraction block (cf. Section 2.3). This procedure leads to short computation times (as $N = 126$), which allow for online use while a test is conducted.

2.5. Sequence Scheme

Fig. 7 depicts a summary of the processes described in the previous sections as four independent blocks that process and classify extracted peaks into physical peaks or artificial spikes.

3. Real-world Data

Real measurements performed on an electric vehicle in a run-up condition were used to validate the methodology. The evaluated sensors are a 3D accelerometer mounted on a wheel hub and a 1D accelerometer mounted on an interior seat. These sensors were selected because they were sensitive to physical peaks caused by a traction control unit. They were acquired with a sampling frequency of 51200 Hz and a total length of 9.28 seconds. In addition, 24 artificial spike anomalies were induced to evaluate the accuracy of the method. The maximum amplitude of the anomalies corresponds to the maximum value of the time series. Fig. 6 depicts a wheel hub accelerometer signal with a combination of physical peaks (marked within a green area) and artificial spikes induced over the entire time series with a fixed distance between them, such that it is possible to study the performance of the algorithm for the same anomaly at different background noise levels. Annex I show plots of the other signals under test.

4. Results

We have compared the performance of the DTW-based spike detector with two other low complexity outlier detection techniques. These are: i) An adaptive threshold technique based on standard deviation per block that we will refer to as "statistic-based spike detector". This method corresponds to the first 3 processing blocks of the scheme in Fig. 7; ii) The Grubbs test, also known as extreme studentized deviate test, a technique to detect outliers in univariate datasets [10].

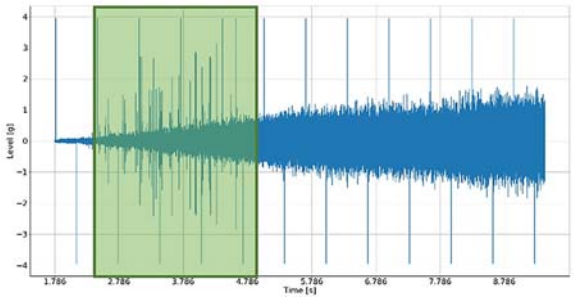


Fig. 6. Wheel hub accelerometer (Y axis) with physical peaks and induced artificial spike anomalies.

The DTW-based Spike detection algorithm was set with the following parameters:

Table 1. DTW-based Spike detector: parameters.

Parameters	
Threshold factor:	5
Spacing between peaks:	65 Samples
Rolling window size:	2048 Samples
Forgetting factor:	0.3

There are other valid approaches, such as Principal Component Analysis (PCA) or methods based on the Autoregressive Integrated Moving Average (ARIMA). However, the last one was not considered for this analysis due to reported high false positive rates for short length anomalies [11], and the attempts of using PCA led to similar disadvantages.

Fig. 8 presents the F1-Score results obtained for each of the spike detection strategies. Annex II depicts the tables with more detailed results (including precision and recall values). The DTW-based spike detector obtains a superior F1-score compared to the other 2 approaches. However, for the Wheel hub X axis sensor, the difference is less pronounced. Upon further investigation, the performance drop is due to the peak extraction block which did not extract sufficient candidate peaks, and not due to wrong classifications by the DTW block.

Considering the more detailed results reported in Annex II, the DTW-based spike detector shows high precision numbers (precision close to 1) for all evaluated sensors, while the recall numbers are between 0.68 and 1. This implies that not every true spike anomaly was detected, while on the other hand a low quantity of false alarms were raised (i.e., a physical peak classified as an artificial spike anomaly). This is considered acceptable from the user perspective, as a small number of missed detections are admissible in this application. The statistic-based spike

detector on the other hand has a high number of false positives for the wheel hub accelerometer (i.e., physical peaks erroneously classified as artificial spikes) such that the obtained precision values are low. This demonstrates that the DTW block is indeed crucial to avoid raising false alarms (cf. Fig. 9). Finally, the Grubbs test based detector results are highly dependent on the sensor channel which is considered. Further investigations are required to understand why the performance is better for some channels over others.

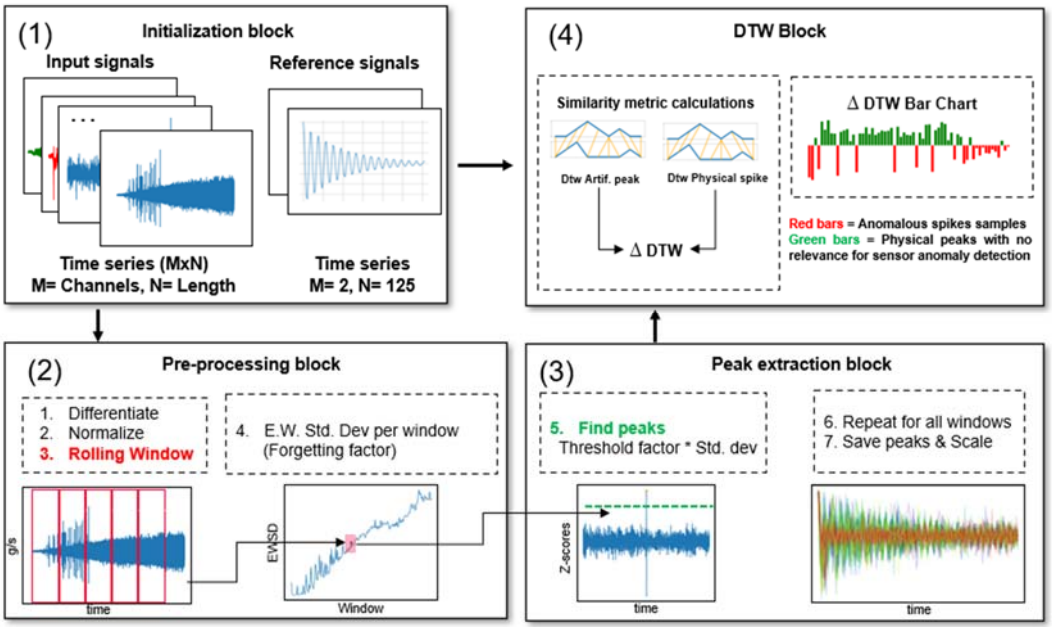


Fig. 7. DTW-based spike detector: processing blocks.

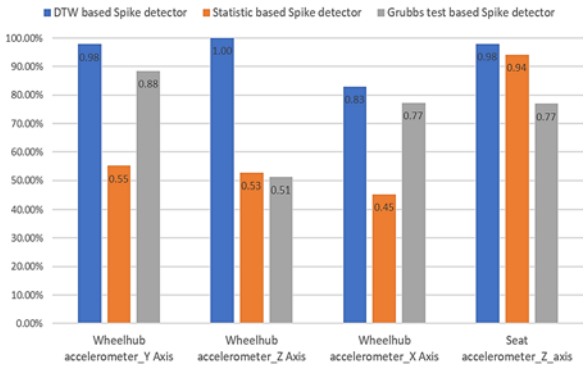


Fig. 8. F1-Score results for each spike detection technique.

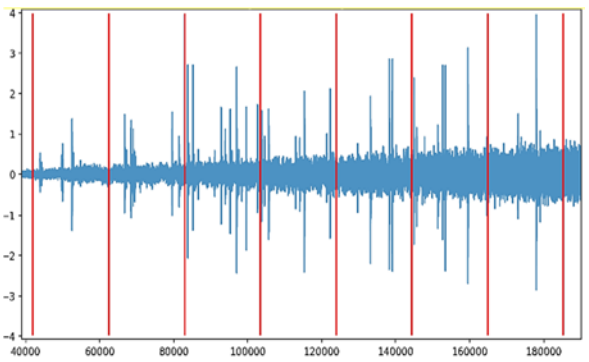


Fig. 9. Zoom in on green area (cf. Fig. 6): artificial spikes are rejected (red lines) while physical peaks do not raise false alarms.

The 3 methods were also compared in terms of computational complexity. Table 2 presents the results obtained for the time consumed to process each of the channels. It can be observed that the DTW and statistical approach have computation times compatible with an online usage, as the processing time is significantly shorter than the duration of the measurement (i.e., 9.28 seconds). However, for the Grubbs test based detector, the processing latency can be considered as incompatible with an online/inline anomaly detection tool.

Table 2. Comparison computation times.

	Wheelhub accelerometer_ Y Axis	Wheelhub accelerometer_Z Axis	Wheelhub accelerometer_X Axis	Seat accelerometer_Z_axis	Unit
DTW	0.04	0.03	0.03	0.03	**seconds**
Stat	0.03	0.03	0.03	0.03	**seconds**
Grubbs test	14.58	13.01	13.56	14.06	**seconds**

5. Conclusions

This paper presented a robust spike detection technique based on DTW with low computational cost and without needing training data, which is suitable for applications in automotive proving ground testing. The method consists of 4 processing blocks which pre-process the sensor signal and extract candidate

peaks, which are then compared to reference template signals through DTW in order to classify them as either physical peaks (non-anomalous) or artificial spikes (anomalous). The DTW-based spike detector was compared to 2 other low complexity outlier detection techniques and validated on real measurements performed on an electric vehicle in a run-up maneuver. It was demonstrated that the DTW-based spike detector obtains a superior performance in terms of the F1-score, while also a low false alarm rate is achieved which is important from the user perspective.

6. Future Work

As further steps in this research, new real-world datasets will be tested with real anomalies, to study in detail the performances of this technique for long duration tests, lower sampling frequencies and more challenging driving maneuvers (i.e., durability test campaigns, ride & handling tests, etc.).

Acknowledgements

We gratefully acknowledge the European Commission for its support of the Marie Sklodowska Curie program through the H2020 ETN MOIRA project (GA 955681).

References

[1]. K. Ni, N. Ramanathan, M. N. H. Chehade, L. Balzano, S. Nair, S. Zahedi, E. Kohler, G. Pottie, M. Hansen, M. Srivastava, Sensor network data fault types, *ACM Transactions on Sensor Networks*, Vol. 5, Issue 3, 2009, pp. 1-29.

[2]. F. Petitjean, A. Ketterlin, P. Gançarski, A global averaging method for dynamic time warping with applications to clustering, *Pattern Recognition*, Vol. 44, 2011, pp. 678-693.

[3]. W. Meert, K. Hendrickx, T. Van Craenendonck, P. Robberechts, DTAIDistance, Version 2, https://zenodo.org/record/3981067

[4]. Y. Chen, B. Hu, E. Keogh, G. E. Batista, DTW-D: time series semi-supervised learning from a single example, in *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'13)*, Chicago, USA, 2013, pp. 383-391.

[5]. C. Harris, Shock and Vibration Handbook, *McGraw-Hill*, 2002.

[6]. F. Deuschle, B. Cornelis, J. Lanslots, K. Gryllias, Overload detection in MEMS microphones-based acoustic arrays, in *Proceedings of the International Conference on Noise and Vibration Engineering*, Leuven, Belgium, 2022.

[7]. T. Fehlmann, E. Kranich, Exponentially Weighted Moving Average (EWMA) prediction in the software development process, in *Proceedings of the Joint Conference of the International Workshop on Software Measurement and the International Conference on Software Process and Product Measurement*, Rotterdam, Netherlands, 2014, pp. 263-270.

[8]. A. Mueen, E. Keogh, Extracting optimal performance from dynamic time warping, in *Proceedings of the 22th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'16)*, San Francisco, USA, 2016, pp. 2129-2130.

[9]. T. Rakthanmanon, B. Campana, M. Abdullah, G. Batista, B. Westover, Q. Zhu, J. Zakaria, E. Keogh, Searching and mining trillions of time series subsequences under dynamic time warping, in *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'12)*, Beijing, China, 2012, pp. 262-270.

[10]. H. Woodrow, Time Series Anomaly Detection, https://github.com/HamishWoodrow/anomaly_detection

[11]. A. B. Sharma, L. Golubchik, R. Govindan, Sensor faults: Detection methods and prevalence in real-world datasets, *ACM Transactions on Sensor Networks*, Vol. 6, Issue 23, 2010, pp. 1-39.

Annex I

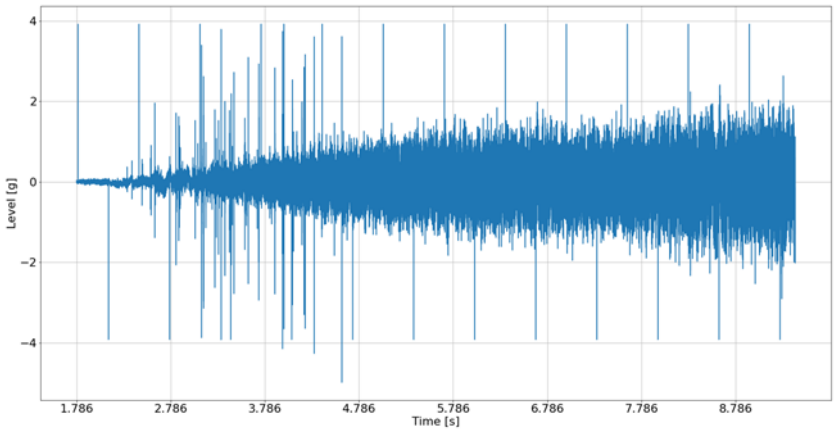


Fig. 10. Wheel hub accelerometer (Z axis) with physical peaks and induced artificial spike anomalies.

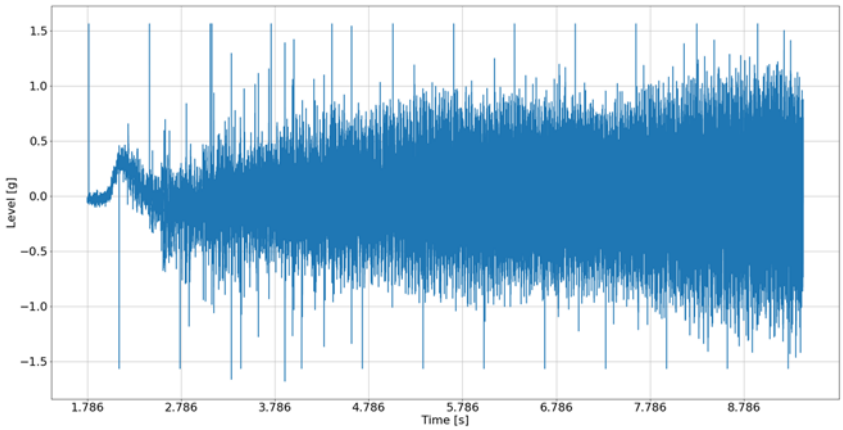


Fig. 11. Wheel hub accelerometer (X axis) with physical peaks and induced artificial spike anomalies.

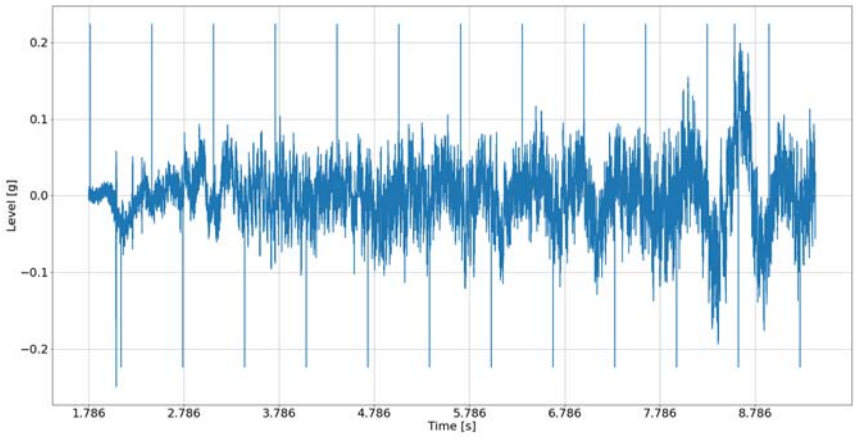


Fig. 12. Seat accelerometer (Z axis) with physical peaks and induced artificial spike anomalies.

Annex II

Table 3. DTW based Spike detector detailed results.

DTW based Spike detector				
	Wheelhub accelerometer_Y Axis	Wheelhub accelerometer_Z Axis	Wheelhub accelerometer_X Axis	Seat accelerometer_Z_axis
Anomalies induced	24	24	24	24
Peaks found	63	67	51	27
TP	23	24	17	24
FP	0	0	0	1
FN	1	0	7	0
TN	39	43	33	2
Accuracy	0.98	1.00	0.88	0.96
Precision	1.00	1.00	1.00	0.96
Recall	0.96	1.00	0.71	1.00
F1-Score	0.98	1.00	0.83	0.98
F2-Score	0.97	1.00	0.75	0.99

Table 4. Statistical based Spike detector detailed results.

Statistic based Spike detector				
	Wheelhub accelerometer_Y Axis	Wheelhub accelerometer_Z Axis	Wheelhub accelerometer_X Axis	Seat accelerometer_Z_axis
Anomalies induced	24	24	24	24
Peaks found	63	67	51	27
TP	24	24	17	24
FP	39	43	34	3
FN	0	0	7	0
TN	0	0	0	0
Accuracy	0.38	0.36	0.29	0.89
Precision	0.38	0.36	0.33	0.89
Recall	1.00	1.00	0.71	1.00
F1-Score	0.55	0.53	0.45	0.94
F2-Score	0.75	0.74	0.58	0.98

Table 5. Grubbs test based Spike detector detailed results.

Grubbs test based Spike detector				
	Wheelhub accelerometer_Y Axis	Wheelhub accelerometer_Z Axis	Wheelhub accelerometer_X Axis	Seat accelerometer_Z_axis
Anomalies induced	24	24	24	24
Peaks found	19	15	20	15
TP	19	10	17	15
FP	0	5	3	0
FN	5	14	7	9
TN	0	0	0	0
Accuracy	0.79	0.34	0.63	0.63
Precision	1.00	0.67	0.85	1.00
Recall	0.79	0.42	0.71	0.63
F1-Score	0.88	0.51	0.77	0.77
F2-Score	0.83	0.45	0.73	0.68

(0508)

Convergence of Gradient Descent Algorithm Based Smoothing Double Regularization for Pi-Sigma Neural Networks

Khidir Shaib Mohamed [1,2]
[1] Department of Mathematics, College of Sciences and Arts in Uglat Asugour, Qassim University, Buraydah, Kingdom of Saudi Arabia
[2] Department of Mathematics and Computer, College of Science, Dalanj University, Dalang, South Kordofan, Sudan
E-mails: K.Idris@qu.edu.sa, khshm7@yahoo.com

Summary: This paper proposed a gradient descent algorithm with smoothing double regularization for pi-sigma neural networks and the analysis of its convergence. The term double regularization is known as a combination of L1 and L2 regularization it involves the absolute value at the origin, and the gradient methods cannot be used directly since is non-differentiable. To get rid of this obstacle, we try to use a differentiable and continuous function to replace that absolute value and end up with the smoothing double regularization. Numerical results show that the smoothing double regularization improved better generalization performance and accelerate the learning process. The weak and strong convergence results are also presented to support our proposes.

Keywords: Convergence, Gradient descent algorithm, Pi-sigma neural network, Smoothing double regularization.

1. Introduction

High-order neural networks (HONNs) are networks that utilize higher combinations of inputs. HONNs have been successfully applied to a variety of real-world tasks, including prediction, classification, signal processing, image recognition, and particularly in covid-19, business, and for trading the EUR/USD exchange rate [1-3], respectively. In a research paper published by Ghosh and Shin [4], they introduced the pi-sigma neural network, a type of high-order neural network, which has shown to be effective in predicting patterns and approximating functions. Since the weights between hidden units and outputs are fixed at 1, the model has multiplication neurons in output unit layers. A neural network of this type uses the product of sums, has a regular structure, is much faster at learning, and can be incrementally enlarged to achieve a desired level of complexity. Data is essential for building learning models and can be used in the modeling pipeline in two ways. In batch learning, the data is at rest, while in online learning, the data flows into the learning algorithm in streams [5], and this paper flow the first approach. The pi-sigma neural network (PSNN) joined a gradient descent algorithm is capable to dispose of the nonlinear problems with much more powerful mapping and higher computational speed [6, 7]. The terms of PSNNs are accepted to address several difficult different problems: such as time series forecasting [8], for decade progress [9], based on sine cosine optimization algorithm [10], forecasting [11], and improved spotted hyena optimizer with space transformational search [12].

In [13], Zou and Hastie introduced a novel double regularization method by extending the L1 regularization by adding L2 regularization for the variable selection method in order to overcome the limitation of the L1 regularization. This term has been used in several studies for example as used for portfolios with risk minimization [14], high-dimensional longitudinal data [15], and discriminate analysis classifiers with automatic parameter selection [16]. The double regularization algorithm is widely used in practice and is implemented in many libraries for machine learning. The modified error function joins double regularization, which takes the following form:

$$\boldsymbol{E(W)} = \check{\boldsymbol{E}}(\boldsymbol{W}) + \boldsymbol{\lambda_1}\|\boldsymbol{W}\|^2 + \boldsymbol{\lambda_2}\|\boldsymbol{W}\|_1, \quad (1)$$

where $\check{\boldsymbol{E}}(\boldsymbol{W})$ is the stander error function depending on the weights $\boldsymbol{W}$, $\|\cdot\|^2$ and $\|\cdot\|_1$ denotes 2-norm and 1-norm, respectively, λ_1, $\lambda_2$are regularization parameters, when $\boldsymbol{\lambda_1} = \boldsymbol{0}, \boldsymbol{\lambda_2} = \boldsymbol{\lambda}$, it is the $\boldsymbol{GL1}$ method, while $\boldsymbol{\lambda_2} = \boldsymbol{0}, \boldsymbol{\lambda_1} = \boldsymbol{\lambda}$ is the $\boldsymbol{GL2}$ method, $\|\boldsymbol{W}\|^2 = \sum_{i=1}^{d} \boldsymbol{w_i^2}$ and $\|\mathbf{W}\|_1 = \sum_{\mathbf{i}=1}^{\mathbf{d}} |\mathbf{w_i}|$, respectively, $\|\cdot\|$ is the normal Euclidean norm, and $|\cdot|$ is the normal absolute value function. In generally many regularization types take the form of the q-norm of the weights $\mathbf{W}$ of the network as

$$\|\boldsymbol{W}\|_{\boldsymbol{q}} = (\sum_{\boldsymbol{m}=1}^{\boldsymbol{M}} |\boldsymbol{w_m}|^{\boldsymbol{q}})^{\frac{1}{\boldsymbol{q}}}, \quad (2)$$

where $q = 0, 0.5, 1$ and 2 are L0 regularization, L1/2 regularization, L1 regularization, and L2 regularization, respectively.

Neural network training using the gradient method has been commonly applied, but sometimes, the training process is poor and overfitting. The idea of adding a regularization term into the error has become common to improve the generalization performance of the network and to make the network weights become

smaller during the training. Various types of regularization terms have been used in many applications of neural network areas. L2 and L1 are the most common terms used which include 2-norm and 1-norm, respectively. The L2 regularization also called weight decay may be the most popular regularization technique used [17-19]. While the L1 regularization has become widespread for improving the sparse solution [20, 21]. Sparse optimization involving the L0 regularization in objective function has a wide application in many fields. AIC and BIC [22], well-known model selection criteria, are special cases of L0 regularization. In [23], a novel L1/2 regularization has been providing more sparsity than the L0 and L1 regularizations. However, since the L0, L1/2, and L1- regularization norms of weights are non-differentiable; we cannot incorporate it directly as a regularization term into the objective function. To get rid of this dilemma, many researchers have begun to use a smoothing technical that removes the inability of a function to be a string, for example (cf. [24, 25]). We try to investigate the proposal of this study which is related to these ideas.

In Section 2, we show the gradient descent algorithm with double regularization ($\boldsymbol{GDr}$), and the gradient descent algorithm with smoothing double regularization ($\boldsymbol{GSDr}$). We selected convergence results are given in Section 3. We present numerical results in Section 4. The proofs of the convergence results in Section 5 with their conclusion summary of this work. Finally, conclusion summary of this work in Section 6.

2. Gradient Descent Algorithm Based on Smoothing Double Regularization (GSDr)

PSNN consists of an input unit, a single hidden layer of linear summation units and product units in the output layer are $\boldsymbol{p, n}$ and $\mathbf{1}$ respectively. The term pi-sigma (ΠΣ) use products of sums of input components. PSNNs have only one unit of adjustable weights, the weights of the output unit are commonly fixed at $\mathbf{1}$ (see Fig. 1). The weight vector connecting the input unit and the k-th summing unit, introduced by $\boldsymbol{w_j} = (\boldsymbol{w_{j1}, w_{j2}, \dots, w_{jp}})\boldsymbol{T} \in \mathbb{R}^{\boldsymbol{P}}$, and given $\boldsymbol{w} = (\boldsymbol{w_1^T, w_2^T, .., w_n^T}) \in \mathbb{R}^{\boldsymbol{nP}}, (\mathbf{1} \le \boldsymbol{j} \le \boldsymbol{n})$. Let $\boldsymbol{g}: \mathbb{R} \to \mathbb{R}$ be a given activation function. For any given input x and weight w, the output of the PSNN is

$$y = g[\prod_{j=1}^{n}(\sum_{i=1}^{p}(w_{ji}x_i))] = \\ = g[\prod_{j=1}^{n}(w_j \cdot x)], \tag{3}$$

where $\boldsymbol{w_j} \cdot \boldsymbol{x}$ represents the inner product of $\boldsymbol{w_j}$ and $\boldsymbol{x}$. The topological structure of PSNN algorithm is shown below:

Let $\{\boldsymbol{x^l}, \boldsymbol{O^l}\}_{\boldsymbol{l=1}}^{\boldsymbol{L}} \subset \mathbb{R}^{\boldsymbol{P}} \times \mathbb{R}$ be the set of training samples with $\boldsymbol{O^l}$ is the desired objective output for the input $\boldsymbol{x^l}$. For any fixed weight w, the output error double regularization ($\boldsymbol{GDr}$) is defined as

$$E(w) = \frac{1}{2}\sum_{l=1}^{L}[O^l - g(\prod_{j=1}^{n}(w_j \cdot x^l))]^2 + \\ +\lambda_1 \sum_{j=1}^{n}\|w_j\|^2 + \lambda_2 \sum_{j=1}^{n}|w_j|, \tag{4}$$

where $\|\cdot\|$ is the normal Euclidean norm, and $|\cdot|$ is the absolute value function. Since the L1- norm of weights includes the absolute value which means is non-differentiable; we cannot incorporate it directly as a regularization term into the objective function. To get rid of this problem, we try to use a differential and continuous function $f(t)$ instead of an absolute value as in the above formula. Specifically, the following piecewise polynomial function is used in [24, 25] as follows:

$$f(t) = \begin{cases} |t| \; if \; |t| \ge \omega \\ -\frac{1}{8\omega^3}t^4 + \frac{3}{4\omega}t^2 + \frac{3\omega}{8}, if \; |t| < \omega, \end{cases} \tag{5}$$

where $\omega > 0$ is a small positive constant. This method provides the following advantages:

- $f(t) \to \left[\frac{3\omega}{8}, +\infty\right)$;
- $f'(t) \to [-1, 1]$;
- $f''(t) \to \left[0, \frac{3}{2\omega}\right]$.

The new error function with smoothing double regularization ($\boldsymbol{GSDr}$) is as follows:

$$E(w) = \frac{1}{2}\sum_{l=1}^{L}[O^l - g(\prod_{j=1}^{n}(w_j \cdot x^l))]^2 + \\ +\lambda_1 \sum_{j=1}^{n} w_j^2 + \lambda_2 \sum_{j=1}^{n} f(w_j) \tag{6}$$

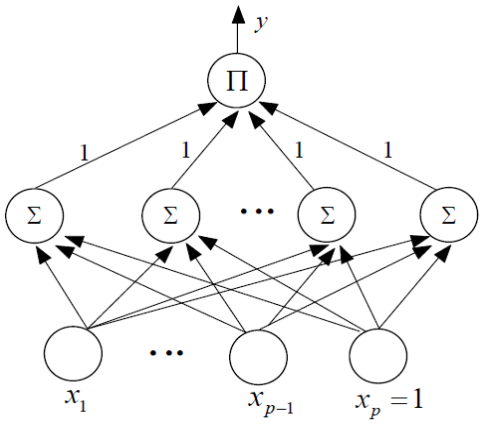


Fig. 1. Structure of pi-sigma neural network.

The purpose of above equation is to obtain w^* such that

$$E(w) = \min E(w^*) \tag{7}$$

The gradient descent algorithm is often used to solve this type of problem. Starting from an arbitrary initial value w^0, the updates the weights w^m iteratively by

$$w_j^{m+1} - w_j^m = \Delta w_j^m = -\eta E_{w_j}(w^m), \quad (8)$$

where $E_{w_j}(w^m) = \frac{\partial E(w^m)}{\partial w_j^m}$ and

$\Delta w_j^m = -\eta[\sum_{l=1}^{L} \delta_l(\prod_{j=1}^{n}(w_j^m \cdot x^l)) \prod_{\substack{k=1 \\ k\neq j}}^{n}(w_k^m \cdot x^l)\, x^l +$
$+2\lambda_1 w_j^m + \lambda_2 f'(w_j^m)], j = 1,2,\dots,J, m = 0,1,2,\cdots,$

where $\eta > 0$ is learning rate and $\delta_l(t) = -[O^l - g'(t)g'(t)]$.

3. Main Results

The following propositions will later be applied to prove the convergence theorem.

Proposition 1.

$|\delta_j(t)|, |\delta_j'(t)|, |g(t)|, |g'(t)|, |g''(t)| \le C_0.$

Proposition 2.

$$max\{|x^l|, |w_j^m \cdot x^l|\} \le C_0.$$

Proposition 3.

The learning rate η and the regularization parameters λ_1, λ_2 are chosen to satisfy $0 < \eta < \frac{1}{\lambda_1 + \lambda_2 M + C}$, where M = 3\2ω and C = C_2+C_3.

Proposition 4.

There exists a compact set φ such that $w^m \in \varphi$ and $\varphi_0 = \{w \in \varphi : E_w(w) = 0\}$ is the contains finite points.

Theorem 1.

Let the error function $E(w)$ be defined by (6) and the weight $\{w^m\}_{m=0}^{\infty}$ is generated by the iteration algorithm (8) for a random initial value. If propositions 1 to 3 are correct, then there are the following estimates:

I. $E(w^{m+1}) \le E(w^m), m = 0,1,2,\cdots.$
II. There exists $E^* \ge 0$ such that
$$\lim_{m\to\infty} E(w^m) = E^*.$$
III. $\lim_{m\to\infty} \left\| E_{w_j}(w^m) \right\| = 0, j = 1,2,\dots,n.$

Furthermore, if proposition 4 is also valid, then we have the following strong convergence: there exists a point $w^* \in \Phi_0$ such that $\lim_{m\to\infty} w^m = w^*$.

4. Experimental Results

PSNNs have a higher-order neural network structure and preferable anticipation performance, making them a different kind of neural network when considered computationally. In this study, the training of our proposed algorithm (GSDr) is performed by differential numerical experimentations of parity problem, tow function approximation problems, and sonar benchmark problem.

4.1. Example 1 (Parity Problem)

The parity problem is known as a normative classification problem. In this problem, we consider 4 input layers, 5 summation layers, and 1 output layer. Selected the initial coefficients carefully in order to correspond to the exercise process to meet the loan from this process as: $\eta = 0.03$ to be the learning rate, $\lambda_1 = 0.0001$ and $\lambda_2 = 0.0003$ are regularization parameters, the weight size is randomly chosen between -0.5 and 0.5, and 1800 repetitions are allowed.

In particular, as predicted by theorem 1, we can see that the proposed learning algorithm GSDr compared with GDr, GL2, GL1 and GSDr has the best learning accuracy. In Table 1, the ANE means the Average Numbers of neurons Eliminated in brief by pruning over 10 trials for the error and the norm of gradients for each learning methods. There is a monotonic decrease in the error function ofGSDr, and its norm of gradient moves to zero more quickly than it does forGDr, GL2, and GL1 see Figs. 2 and 3. This result supports the proposal of this study.

Table 1. Numerical results for Example 1.

Learning methods	Average training error	Norm of gradient	ANE	Training time (s)
GL1	3.1433e-05	2.5336e-04	4.3	5.331998
GL2	2.1846e-05	1.8055e-04	4.1	5.356528
GDr	1.9049e-05	1.5900e-04	3.9	5.360701
GSDr	1.0383e-05	2.8449e-05	4.8	5.308263

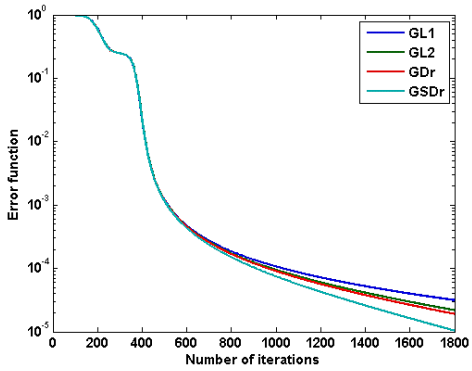


Fig. 2. Learning errors of different methods in example 1.

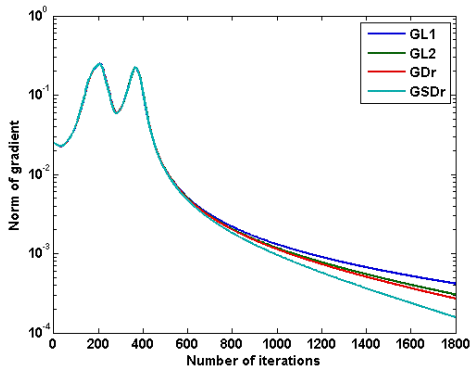


Fig. 3. Norm of gradient of different methods in example 1.

4.2. Example 2 (Approximation Function)

In order to verify the effectiveness of the learning, the algorithm approximates the function: $cos(x)$, $x \in [-1, 1]$ is used and the numbers of the training samples are101 data. While the initial weights are randomly assigned values between $[-0.2, 0.2]$. In this example, the network it is considered with the structure 2 input samples, 3 hidden layer neurons, and 1 output, the learning rate η is 0.05 and the regularization parameters λ_1 and λ_2 is 0.001 and 0.003, respectively. Training periods are determined by updated weights ending in 10,000.

Compared the performance of the GDr, GL2, GL1 and GSDr methods are shown in Figs. 4 to 7. We see that error of GSDr is decreases monotonically and it is the norm of gradient trends to zero, as depicted by the convergence Theorem 1. It is clear from Table 2 that GSDr has much stronger prediction capability and achieves better ANE over 10 trials than the common GL2, GL1, and GDr.

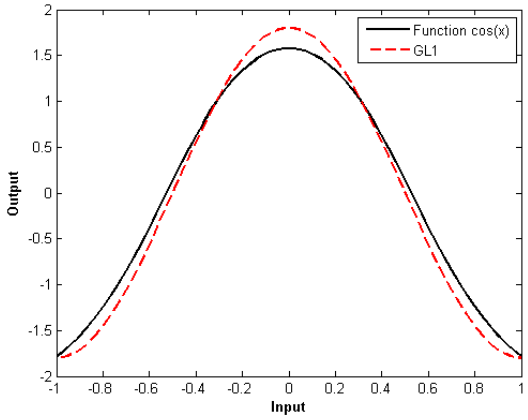


Fig. 4. Function approximation by $\boldsymbol{GL1}$ in Example 2.

4.3. Example 3 (Gabor Approximation Function)

We use the 2D Gabor function to illustrate the possibility of GSDr. As shown in Fig. 8, the two-dimensional Gabor function has the following form:

$$h(a,b) = \frac{1}{2\pi(0.5)} exp\left(\frac{a^2+b^2}{2(0.5)^2}\right) cos[2\pi(a+b)]$$

Then, 325 entry points were selected for training and testing to display the ability of the GSDr. For training, 36 entry points were randomly selected from 6×6 enablers evenly spaced at -0.5×0.5 and -0.5×0.5 out of these 325 points. The remaining 289 points were randomly selected from 17×17 equally spaced at $-0.5 \leq a \leq 0.5$ and $-0.5 \leq b \leq 0.5$ for testing from these 325 points.

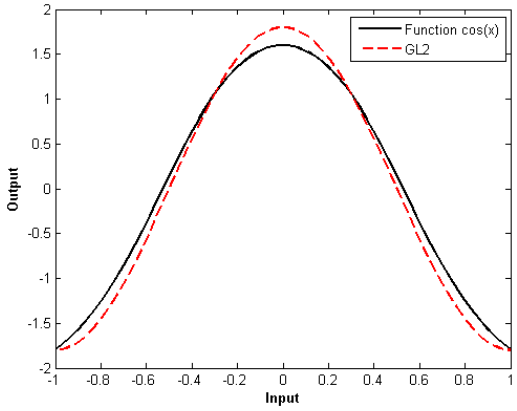


Fig. 5. Function approximation by $\boldsymbol{GL2}$ in Example 2.

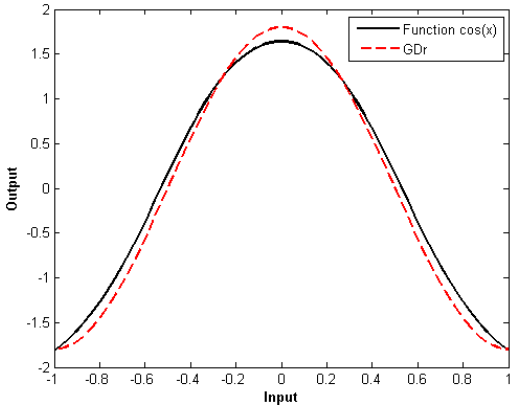


Fig. 6. Function approximation by $\boldsymbol{GDr}$ in Example 2.

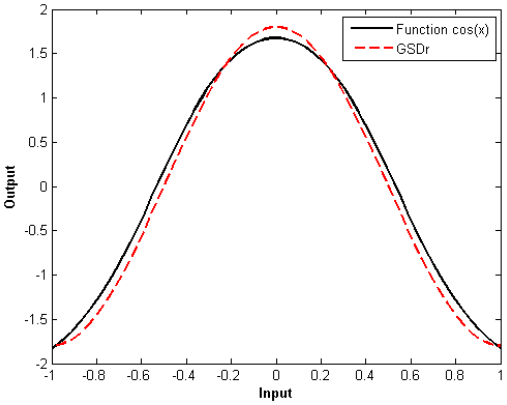


Fig. 7. Function approximation by $\boldsymbol{GSDr}$ in Example 2.

Table 2. Numerical results for Example 2.

Learning methods	Average training error	Average test error	ANE	Training time (s)
$GL1$	1.6036e-05	0.0048	5.9	0.555564
$GL2$	5.4861e-05	0.0044	6.1	0.566383
GDr	4.1465e-04	0.0037	6.0	0.560159
$GSDr$	8.8803e-04	0.0035	7.3	0.542739

In Table 3, the results are compared for GDr, GL2, GL1 and GSDr algorithms, where the Average Numbers of neurons Eliminated was taken for 10 different iterations. From the results obtained our method evaluates a good performance and accelerated learning planning capabilities compared to others. From Figs. 9 to 12, we observe that our proposed algorithm generalizes stellar and achieves a smooth insertion among the training process than the GDr, GL2, and GL1.

Table 3. Numerical results for Example 3.

Learning methods	Average training error	Average test error	ANE	Training time (s)
GL1	0.0407	0.0483	2.0	5.428512
GL2	0.0409	0.0480	2.2	5.493166
GDr	0.0406	0.0481	2.5	5.483647
GSDr	0.0101	0.0111	3.3	5.415376

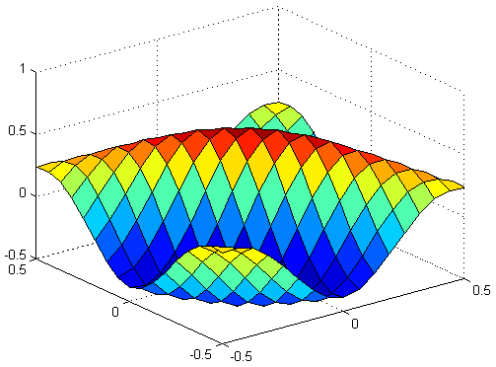


Fig. 8. Gabor function in example 3.

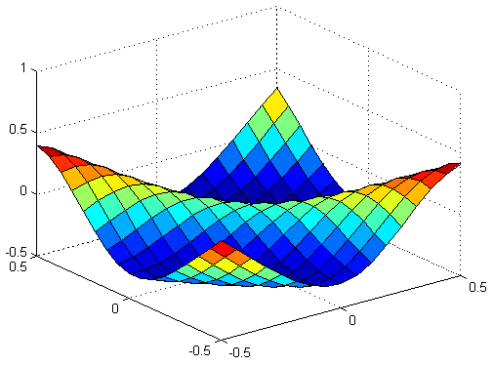


Fig. 9. Gabor approximation by $\boldsymbol{GL1}$ in example 3.

4.4. Example 4 (Sonar Classification Problem)

Sonar is a linearly inseparable classification problem. The target of this data has been used to classify reflected sonar signals into metal chambers and shakes). Analyzing the performance of 208 input vectors, each of which has 60 interval components, which are used to test the presented training algorithms. Informatics indexes for preparation and testing are randomly selected from 208 information vectors, and the neurons of the organization contain 60 information units, 6 summation units and 1 production unit. This test is completed by choosing $\eta = 0.6$ is the learning rate, $\lambda_1 = 0.0001$ and $\lambda_2 = 0.0005$ regularization parameters, and update weights ending with 5000 are the maximum number of training periods. From Table 4, we found that our proposed method gives better accuracy (%), as the result in Table 4 was taken out of the ten best results obtained under the same conditions for four algorithms. The result of this comparison is generally acceptable to support theoretical funding.

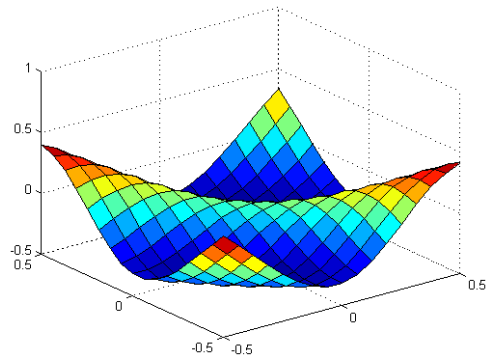


Fig. 10. Gabor approximation by $\boldsymbol{GL2}$ in example 3.

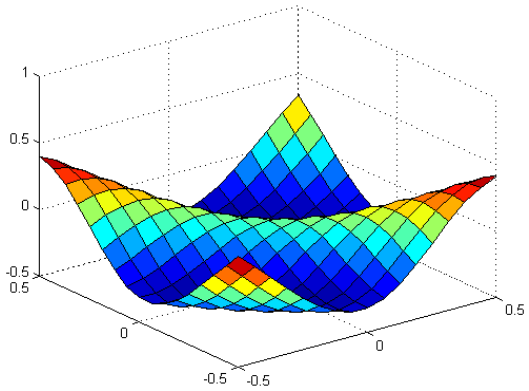


Fig. 11. Gabor approximation by $\boldsymbol{GDr}$ in example 3.

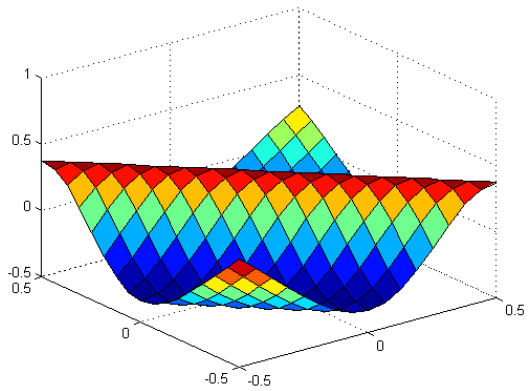


Fig. 12. Gabor approximation by $\boldsymbol{GSDr}$ in example 3.

Table 4. Numerical results for Example 4.

Learning methods	Accuracy training (%)	Accuracy testing (%)	Training time (s)
GL1	95.80	92.22	15.358937
GL2	96.40	93.27	15.398114
GDr	94.72	90.98	15.849914
GSDr	98.41	97.05	15.277646

5. Proofs

The following tow lemmas are a crucial tool for our analysis. For sake of convenience, we introduce the notations:

$$\Pi^{m+1,l} = \prod_{j=1}^{n}(w_j^{m+1}\cdot x^l) \tag{9}$$

Lemma 1. Let the propositions 1 and 2 are satisfied, and that $\{\mathrm{w}^{\mathrm{m}}\}_{\mathrm{m}=0}^{\infty}$ is grated by (8). We have

$$(\Pi^{m+1,l}-\Pi^{m,l})^2 \le C_1\sum_{j=1}^{n}\left\|\Delta w_j^m\right\|^2, \tag{10}$$

and

$$\begin{aligned}&\sum_{l=1}^{L}\delta_l(\Pi^{m,l})\,[\Pi^{m+1,l}-\Pi^{m,l}]\le\\ &\le\left[-\frac{1}{\eta}+C_2\right]\sum_{j=1}^{n}\left\|\Delta w_j^m\right\|^2\end{aligned} \tag{11}$$

Proof. We write Following the Taylor expansion first and second orders. We write

$$\Pi^{m+1,l}-\Pi^{m,l} = \sum_{j=1}^{n}\Big(\prod_{\substack{k=1\\k\neq j}}^{n}t_k\Big)\left(\Delta w_j^m\cdot x^l\right), \tag{12}$$

and

$$\begin{aligned}\Pi^{m+1,l}-\Pi^{m,l} &= \sum_{j=1}^{n}\prod_{k\neq i}^{n}(w_k^m\cdot x^l)\left(\Delta w_j^m\cdot x^l\right)+\\ &+\frac{1}{2}\sum_{\substack{j=1\\j\neq s}}^{n}\Big[\prod_{\substack{k=1\\k\neq j\\k\neq s}}^{n}\left(t_k^{j,s}\right)\Big]\left(\Delta w_j^m\cdot x^l\right)\left(\Delta w_s^m\cdot x^l\right)=\\ &=\sum_{j=1}^{n}\prod_{k\neq i}^{n}(w_k^m\cdot x^l)\left(\Delta w_j^m\cdot x^l\right)+\sigma,\end{aligned} \tag{13}$$

where t_k and $t_k^{j,s}$ are in between $\left(w_j^{m+1}\cdot x^l\right)$ and $\left(w_j^m\cdot x^l\right)$. By using proposition 2, easy to see that

$$|t_k|\le 2C_0, \left|t_k^{i,s}\right|\le 2C_0, k = 1,2,\dots,n \tag{14}$$

Thus, can be easily to obtained (10), from (12) and (14). Through (14), we write

$$\begin{aligned}|\sigma| &= \frac{1}{2}\sum_{\substack{j=1\\j\neq s}}^{n}\Big[\prod_{\substack{k=1\\k\neq j\\k\neq s}}^{n}\left(t_k^{j,s}\right)\Big]\left(\Delta w_j^m\cdot x^l\right)\left(\Delta w_s^m\cdot x^l\right)\le\\ &\le\frac{1}{4}(2C_0)^{n-2}\max_{1\le l\le L}\|x^l\|^2\sum_{\substack{j=1\\j\neq s}}^{n}\sum_{s=1}^{n}\left(\left\|\Delta w_j^m\right\|^2+\left\|\Delta w_s^m\right\|^2\right)\le\\ &\le C_2'\sum_{j=1}^{n}\left\|\Delta w_j^m\right\|^2,\end{aligned} \tag{15}$$

where $C_2' = \frac{1}{4}(2\mathrm{C}_0)^{\mathrm{n}-2}\max_{1\le l\le L}\left\|\mathrm{x}^1\right\|^2$.

From collecting (8), (13), (15) and proposition 1. We have

$$\begin{aligned}&\sum_{l=1}^{L}\delta_l\,(\Pi^{m,l})(\Pi^{m+1,l}-\Pi^{m,l})=\\ &=\sum_{l=1}^{L}\delta_l(\Pi^{m,l})\,[\sum_{j=1}^{n}\left(\Delta w_j^m\cdot x^l\right)\prod_{k\neq j}^{n}(w_k^m\cdot x^l)+\sigma]\le\\ &\le\sum_{j=1}^{n}[\sum_{l=1}^{L}\delta_l(\Pi^{m,l})\prod_{\substack{k=1\\k\neq j}}^{n}(w_k^m\cdot x^l)\,x^l]\Delta w_j^m+\\ &+LC_0C_2'\sum_{j=1}^{n}\left\|\Delta w_j^m\right\|^2=\left[-\frac{1}{\eta}+C_2\right]\sum_{j=1}^{n}\left\|\Delta w_j^m\right\|^2\end{aligned} \tag{16}$$

where $C_2 = LC_0C_2'$ This completes the proof.

Lemma 2. Let $\mathbb{Q}\colon\mathbb{R}^{\mathrm{s}}\to\mathbb{R}$is continuous and differentiable on a compact set $\upsilon\subset\mathbb{R}^s$ and that $\Omega = \left\{h\in\upsilon\middle|\frac{\partial F(h)}{\partial h}=0\right\}$ has only finite number of points. If a sequence $\{h^z\}_{z=1}^{\infty}\in\upsilon$ satisfies

$$\lim_{z\to\infty}\|h^{z+1}-h^z\| = 0, \lim_{z\to\infty}\left\|\frac{\partial F(h^z)}{\partial h}\right\| = 0$$

Then there exists a point $h^*\in\Omega$ such that

$$\lim_{m\to\infty}h^z = h^*$$

Proof Lemma 2. is almost the same as Lemma 1 in [26] and the detail of the proof is omitted.

There are four parts to the proof of Theorem 1: statements (I), (II), (II), and (IV).

Proof to (*I*) of Theorem 1. Combining (4), (10) (11), and the Taylor expansion, we get

$$\begin{aligned}&E(w^{m+1})-E(w^m)=\\ &=\frac{1}{2}[\sum_{l=1}^{L}(o^l-\Pi^{m+1,l})^2-\sum_{l=1}^{L}(o^l-\Pi^{m,l})^2]+\\ &+\lambda_1\sum_{j=1}^{n}\left[\left(w_j^{m+1}\right)^2-\left(w_j^m\right)^2\right]+\\ &+\lambda_2\sum_{j=1}^{n}\left[f\left(w_j^{m+1}\right)+f\left(w_j^m\right)\right]=\\ &=\sum_{l=1}^{L}\delta_l\,(\Pi^{m,l})(\Pi^{m+1,l}-\Pi^{m,l})+\\ &+\frac{1}{2}\sum_{l=1}^{L}\delta_l'\,(t_l')(\Pi^{m+1,l}-\Pi^{m,l})^2+\\ &+\lambda_1\sum_{j=1}^{n}\left[2w_j^m+\Delta w_j^m\right]\Delta w_j^m+\\ &+\lambda_2\sum_{j=1}^{n}\left[f'\left(w_j^m\right)+f''(t_{j,m})\Delta w_j^m\right]\Delta w_j^m=\\ &=\sum_{j=1}^{n}[\sum_{l=1}^{L}\delta_l(\Pi^{m,l})\prod_{\substack{k=1\\k\neq j}}^{n}(w_k^m\cdot x^l)\,x^l]\Delta w_j^m+\\ &+\left[2\lambda_1w_j^m+\lambda_2f'(w_j^m)\right]\sum_{j=1}^{n}\Delta w_j^m+\\ &+\frac{1}{2}LC_0C_1\sum_{j=1}^{n}\left\|\Delta w_j^m\right\|^2+C_2\sum_{j=1}^{n}\left\|\Delta w_j^m\right\|^2+\\ &+\left[\lambda_1+\lambda_2f''(t_{j,m})\right]\sum_{j=1}^{n}\left\|\Delta w_j^m\right\|^2=\\ &=\left[-\frac{1}{\eta}+\lambda_1+\lambda_2f''(t_{j,m})\right]\sum_{j=1}^{n}\left\|\Delta w_j^m\right\|^2+\\ &+\left[\frac{1}{2}LC_0C_1+LC_0C_2'\right]\sum_{j=1}^{n}\left\|\Delta w_j^m\right\|^2\le\\ &\le-\left[\frac{1}{\eta}-\lambda_1-\lambda_2f''(t_{j,m})\right]\sum_{j=1}^{n}\left\|\Delta w_j^m\right\|^2+\\ &+[C_2+C_3]\sum_{j=1}^{n}\left\|\Delta w_j^m\right\|^2,\end{aligned} \tag{17}$$

where $C_3 = \frac{1}{2}LC_0C_1\,t_l'$ is in between $\prod_{j=1}^{n}\left(w_j^{m+1}\cdot x^l\right)$ and $\prod_{j=1}^{n}\left(w_j^m\cdot x^l\right)$ and $t_{j,m}$ is in between w_j^{m+1} and w_j^m. With (17), proposition 3, and the Lagrangian Mean Value Theorem for $f(t)$ and Let $M = f''\left(t_{j,m}\right) = \frac{3}{2\omega}$, then we have

$$\begin{aligned} & E(w^{m+1}) - E(w^m) = \\ & = [-\frac{1}{\eta} + \lambda_1 + \lambda_2 f''(t_{j,m}) + \\ & + C_{2+}C_3] \sum_{j=1}^{n} \left\| \Delta w_j^m \right\|^2 \leq \\ & \leq -[\frac{1}{\eta} - \lambda_1 - \lambda_2 M - C] \sum_{j=1}^{n} \left\| \Delta w_j^m \right\|^2 \leq 0 \end{aligned} \tag{18}$$

This completes the proof to statement (I) of Theorem 1.

Proof to (II) of the Theorem 1. From conclusion (I), we know that $E(w^m)$ is monotone. Additionally, it was bound below. Hence, there must exist $E^* \geq 0$ such that

$$\lim_{k \to \infty} E(w^m) = E^*$$

The proof to (II) is thus completed.

Proof to (III) of the Theorem 1. From (18) and proposition 1 we can write $\gamma = \frac{1}{\eta} - \lambda_1 - M\lambda_2 - C$ and set $\gamma > 0$. We have

$$E(w^{m+1}) \leq E(w^m) - \gamma \sum_{j=1}^{n} \left\| E_{w_j}(w^m) \right\|^2 \leq \cdots \leq$$

$$\leq E(w^0) - \gamma \sum_{m=0}^{k} \sum_{j=1}^{n} \left\| E_{w_j}(w^m) \right\|^2$$

Since $E(w^{k+1}) > 0$ then we have

$$\gamma \sum_{m=0}^{k} \sum_{j=1}^{n} \left\| E_{w_j}(w^m) \right\|^2 \leq E(w^0) < \infty$$

Setting $k \to \infty$, we get

$$\sum_{m=0}^{k} \sum_{j=1}^{n} \left\| E_{w_j}(w^m) \right\|^2 \leq \frac{1}{\gamma} E(w^0) < \infty$$

Thus

$$\lim_{m \to \infty} \sum_{j=1}^{n} \left\| E_{w_j}(w^m) \right\|^2 = 0$$

Observation that

$$0 \leq \left\| E_{w_j}(w^m) \right\|^2 \leq \sum_{j=1}^{n} \left\| E_{w_j}(w^m) \right\|^2 \to \infty$$

Again setting $m \to \infty$, then we see that

$$\lim_{m \to \infty} \left\| E_{w_j}(w^m) \right\| = \lim_{m \to \infty} \left\| \Delta w_j^m \right\| = 0 \tag{19}$$

This completes the proof of (III).

Proof to (IV) of Theorem 1. Returning to the error function $E(w)$ defined in (4) is a continuous and differentiable function. According to Lemma 2, proposition 4, and (19), we can easily get the desired result, that is, there is a point $w^* \in \varphi_0$ like so

$$\lim_{m \to \infty} (w^m) = w^*$$

This completes the proof to (IV).

6. Conclusions

We thank you in advance for the usage carefully of instructions for camera-ready articles, which can be sent for publication with minor modification. In this paper, smoothing double regularization is a novel method based on the gradient descent algorithm for training and pruning PSNN. At the origin, the double regularization term is not differentiable. Using a differentiable continuous function is able to replace the absolute value at the origin and end up by smoothing double regularization. The great advantage of this method is that it allows the optimization to be solved. As well, the numerical show that our new method improved better generalization performance. The monotonicity of the error function in the learning process is proved. And also, the Weak and strong convergence results are presented.

Acknowledgements

The authors like to thank the anonymous referees for their valuable comments and words of encouragement. Their comments helped to improve the clarity of this paper.

References

[1]. S. Das, J. Nayak, S. Nayak, S. Dey, Prediction of life insurance premium during Pre- and Post-Covid-19: A higher-order neural network approach, *Journal of The Institution of Engineers (India): Series B*, 2022, pp. 1-27.

[2]. S. Xu, Adaptive higher order neural network models and their applications in business, in Artificial Higher Order Neural Networks for Economics and Business, *IGI Global*, 2009, pp. 314-329.

[3]. C. L. Dunis, J. Laws, G. Sermpinis, Higher order and recurrent neural architectures for trading the EUR/USD exchange rate, *Quantitative Finance*, Vol. 11, Issue 4, 2011, pp. 615-629.

[4]. Y. Shin, J. Ghosh, The pi-sigma network: An efficient higher-order neural network for pattern classification and function approximation, in *Proceedings of the International Joint Conference on Neural Networks (IJCNN'91)*, Vol. 1, 1991, pp. 13-18.

[5]. E. Bisong, Batch vs. online learning, in Building Machine Learning and Deep Learning Models on Google Cloud Platform, *Apress*, Berkeley, CA, 2019, pp. 199-201.

[6]. Y. Liu, D. Yang, N. Nan, L. Guo, J. Zhang, Strong convergence analysis of batch gradient-based learning algorithm for training pi-sigma network based on TSK fuzzy models, *Neural Processing Letters,* Vol. 43, Issue 3, 2016, pp. 745-758.

[7]. J. Sensors Nayak, B. Naik, H. S. Behera, A hybrid PSO-GA based Pi sigma neural network (PSNN) with standard back propagation gradient descent learning for classification, in *Proceedings of the International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT'14)*, 2014, pp. 878-885.

[8]. E. Egrioglu, U. Yolcu, E. Bas, Intuitionistic high-order fuzzy time series forecasting method based on pi-sigma artificial neural networks trained by artificial bee colony, *Granular Computing*, Vol. 4, Issue 4, 2019, pp. 639-654.

[9]. H. Swapna Rekha, J. Nayak, H. S. Behera, Pi-sigma neural network: Survey of a decade progress, Computational Intelligence in Pattern Recognition, *Springer*, Singapore, 2020, pp. 429-441.

[10]. E. Bas, E. Egrioglu, O. Karahasan, A Pi-Sigma artificial neural network based on sine cosine optimization algorithm, *Granular Computing*, Vol. 7, 2022, pp. 813-820.

[11]. O. Yılmaz, E. Bas, E. Egrioglu, The training of Pi-Sigma artificial neural networks with differential evolution algorithm for forecasting, *Computational Economics*, Vol. 59, Issue 4, 2022, pp. 1699-1711.

[12]. N. Panda, S. K. Majhi, Improved spotted hyena optimizer with space transformational search for training pi-sigma higher order neural network, *Computational Intelligence*, Vol. 36, Issue 1, 2020, pp. 320-350.

[13]. H. Zou. T. Hastie, Regularization and variable selection via the elastic net, *Journal of the Royal Statistical Society: Series B*, Vol. 67, Issue 2, 2005, pp. 301-320.

[14]. W. Shen, J. Wang, S. Ma, Doubly regularized portfolio with risk minimization, in *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*, Vol. 28, Issue 1, June 2014.

[15]. Y. Li, S. Wang, P.X.K. Song, N. Wang, L. Zhou, J. Zhu, Doubly regularized estimation and selection in linear mixed-effects models for high-dimensional longitudinal data, *Statistics and Its Interface*, Vol. 11, Issue 4, 2018, 721.

[16]. A. Zaib, T. Ballal, S. Khattak, T.Y. Al-Naffouri, A doubly regularized linear discriminant analysis classifier with automatic parameter selection, *IEEE Access,* Vol. 9, 2021, pp. 51343-51354.

[17]. Y. Liu, D. Yang, C. Zhang, Relaxed conditions for convergence analysis of online back-propagation algorithm with L2 regularizer for Sigma-Pi-Sigma neural network, *Neurocomputing*, Vol. 272, 2018, pp. 163-169.

[18]. K. Nakamura, B.W. Hong, Adaptive weight decay for deep neural networks, *IEEE Access*, Vol. 7, 2019, pp. 118857-118865.

[19]. T. Van Laarhoven, L2 regularization versus batch and weight normalization, *arXiv Preprint*, 2017, arXiv: 1706.05350.

[20]. A. Cortiella, K. C. Park, A. Doostan, Sparse identification of nonlinear dynamical systems via reweighted ℓ1-regularized least squares, *Computer Methods in Applied Mechanics and Engineering*, Vol. 376, 2021, 113620.

[21]. X. Qian, H. Huang, X. Chen, T. Huang, Efficient construction of sparse radial basis function neural networks using L1-regularization, *Neural Network*, Vol. 94, 2017, pp. 239-254.

[22]. K. P. Burnham, D. R Anderson, Multimodel inference: Understanding AIC and BIC in model selection, *Sociological Methods & Research*, Vol. 33, Issue 2, 2004, pp. 261-304.

[23]. Z. Xu, H. Zhang, Y. Wang, X. Chang, Y. Liang, L1/2 regularization, *Science China Information Sciences*, Vol. 53, Issue 6, 2010, pp. 1159-1169.

[24]. K. S. Mohamed, W. Wu, Y. Liu, A modified higher-order feed forward neural network with smoothing regularization, *Neural Network World*, Vol. 27, Issue 6, 2017, pp. 577-592.

[25]. Y. Liu, Z. Li, D. Yang, K. S. Mohamed, J. Wang, W. Wu, Convergence of batch gradient learning algorithm with smoothing L1/2 regularization for Sigma–Pi–Sigma neural networks, *Neurocomputing*, Vol. 151, 2015, pp. 333-341.

[26]. W. Wu, L. Li, J. Yang, Y. Liu, A modified gradient-based neuro-fuzzy learning algorithm and its convergence, *Information Sciences*, Vol. 180, Issue 9, 2010, pp. 1630-1642.

(0889)

Multiple Unscented Particle Filtering for Dynamic State Estimation in Power Systems

N. Amor [1], M. T. Noman [1] and M. Petru [1]
[1] Department of Machinery Construction, Institute for Nanomaterials, Advanced Technologies and Innovation (CXI) Technical University of Liberec,
461 17, Studentska 1402/2 Liberec 1, Czech Republic
E-mail: nesrine.amor@tul.cz

Summary: The effective control and operation of the power system depend mainly on accurate information about dynamic states. In this paper, we formulate the problem of a dynamic state power system as a nonlinear non-Gaussian state-space system using the Multiple Unscented Particle Filters (MUPF). In the proposed framework, the measurements are a Phasor Measurement Unit (PMU) data, and the system state consist of estimating the dynamic state in the power system when subjected to disturbances. Extensive simulation results carried out on the IEEE New England 39-bus 10-machine test system demonstrates the effectiveness and good performance of the MUPF under fault conditions.

Keywords: Dynamic state estimation, Unscented particle filter, Phasor measurement unit, Power system, Prediction.

1. Introduction

The real-time states of electromechanical dynamics (i.e., the internal states of generators), reflect the current status of a power system and are used to enhance small signal stability and transient problems in power systems such as automatic voltage regulators, power system stabilizers, and under-frequency relays. It can also be utilized to coordinate controllers in a wide area power system. A dynamic model with precise estimated states can closely detect system responses and, thus can be utilized to increase system reliability and stability in the power system. Phasor Measurement Units (PMU) capture the evolution of the dynamic response of the power system under abnormal and normal conditions [1]. Therefore, the Dynamic State Estimation (DSE) of the power system that represents the voltage magnitude and phase angles of the buses is estimated, using the PMU data.

Several approaches have been proposed for tracking dynamic state estimation in power systems subject to large disturbances, including Extended Kalman Filter (EKF), Unscented Kalman Filter (UKF) and Particle Filters (PF), where PF has shown good performance for small systems, and perform better than EKF and UKF [2].

PF provides powerful results for nonlinear and non-Gaussian state-space scenario. However, it is ineffective in high dimensional spaces [3]. Various approaches have been emerged to ameliorate the performance of PF, where UKF or EKF are used to draw the importance distribution [4]. These approaches are the Extended Particle Filter (PF-EKF) and Unscented Particle Filter (UPF). Recently, Multiple Unscented Particle Filters (MUPF) has been proposed in [5], where it has shown promising results in high-dimensional state space system [5]. Therefore, we propose in this paper to use the MUPF for estimating the DSE in the power system.

2. Unscented Particle Filter

We consider a general dynamic system expressed by the following equations:

$$x_k = f(x_{k-1}) + w_k, \tag{1}$$

$$y_k = h(x_{k-1}) + u_k, \tag{2}$$

where x_k and y_k represent the hidden system state and the observation vectors, respectively. f_k and h_k are nonlinear state and measurement functions, respectively. w_k and u_k are respectively the system and measurement noises. In PF, we estimate the state using an ensemble of weighted particles as follows [3]:

$$P(x_k \mid y_{1:k}) = \sum_{i=1}^{N} w_k^{(i)} \delta\left(x_k - x_k^{(i)}\right), \tag{3}$$

where δ is the Dirac delta function.

Ideally, the particles need to be sampled from the true posterior, which is not available. Therefore, another distribution, referred to as the proposal distribution, $q(x_k / x_{k-1}, y_k)$ is used. The importance weights are given by:

$$w_k^{(i)} = w_{k-1}^{(i)} \frac{p(y_k \mid x_k^{(i)})\ p(x_k^{(i)} \mid x_{k-1}^{(i)})}{q(x_k^{(i)} \mid x_{k-1}^{(i)}, y_{1:k})} \tag{4}$$

The mean estimate of the system state is presented as follows:

$$\hat{x}_k = \sum_{i=1}^{N} w_k^{(i)}\ x_k^{(i)} \tag{5}$$

The main steps of UPF are described in [4].

3. Power System Dynamic State Estimation

The main objective is to estimate the rotor angle and rotor speed of each generator in the power system. Therefore, the system state is expressed by:

$$x = [\sigma_1, \dots, \sigma_{ng}, w_1, \dots, w_{ng}] \quad (6)$$

where ng is the total number of generators. The measurement is given by the PMU as:

$$y = [v_1, \dots, v_{ng}, \theta_1, \dots, \theta_{ng}, P_1, \dots, P_{ng}, Q_1, \dots, Q_{ng}], \quad (7)$$

where P and Q are, respectively, real and reactive power injection. v and θ are the voltage magnitude and angle. We used the synchronous machine model described by [2]:

$$\begin{cases} \dot{\sigma} = (w - w_s) \\ \dot{w} = \frac{w_s}{2H_i}[P_{mi} - P_i - D_i(w - w_s)] \end{cases}, \quad (8)$$

where w_i represents the deviation of rotor speed related to generator i. σ_i represents the angular position of rotor related to generator i. w_s defines the synchronous rotor speed. P_{mi} represents the mechanical power input. H_i defines the inertia constant. P_i represents the electrical power injection.

4. Simulation Results

We evaluated the performance of the MUPF and UPF for dynamic state estimation in power system under fault condition. We used for implementation the IEEE New England 39-bus test system [2]. The initial condition for the estimated states is obtained by executing the load flow. We consider the communication link with PMU placed at bus 34, where generator 5 is connected. PMU measurements are assumed to be recorded at a rate of 48 samples per second. The base power was 100 MW for this system. A large disturbance to the system is applied at t = 0.5 s to t = 0.55 s by opening the transmission line between buses 15 and 16, and the system therefore went into an unstable condition. We applied MUPF, in which each UPF is used to track the state of each generator.

Figs. 1 and 2 present the tracking performance of rotor angle deviation and speed angle, respectively, for the generator 5 under the fault condition over time. We noticed that MUPF and UPF are able to estimate the states of parameters (rotor angle deviation and speed angle) using PMU data under an unstable condition.

We computed the average of the mean absolute error (MAE) over 1000 Monte Carlo run to confirm the efficiency of the proposed algorithm for the two estimated states over 10 s, where UPF average = $3.8431e^{-4}$ and MUPF average = $4.5274e^{-7}$ for rotor angle, and UPF average = $2.2817e^{-4}$, MUPF average = $2.2272e^{-5}$ for rotor speed. We observed that the performance of the MUPF is significantly better than UPF in both cases of rotor angle deviation and angle speed. The MUPF outperforms the UPF in tracking both rotor speed and rotor angle with a lower error. In conclusion, MUPF exhibits good tracking capabilities for power system dynamic state estimation under fault conditions compared with UPF.

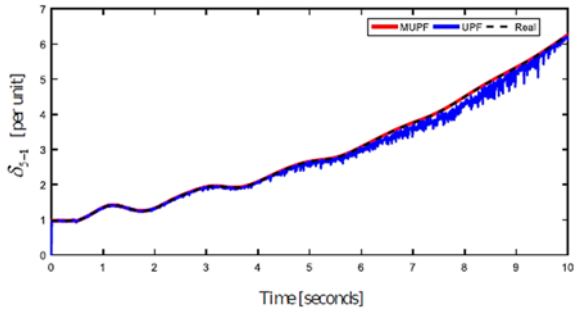


Fig. 1. Time history of the rotor angle of generator-5 tracked during large disturbance using MUPF and UPF.

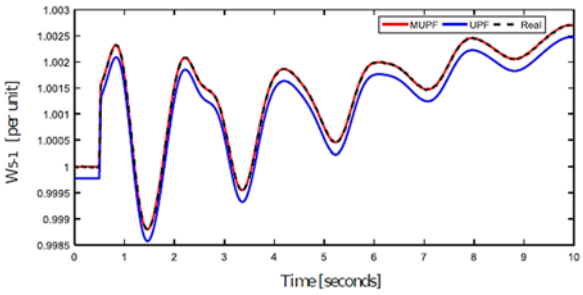


Fig. 2. Time history of the rotor speed of generator-5 tracked during large disturbance using MUPF and UPF.

5. Conclusions

In this paper, we tracked the dynamic state estimation in power system using MUPF. MUPF has been successfully implemented and tested on IEEE New England 39-bus test system to estimate the dynamical state of the generator using PMU data. The simulation results showed that the MUPF was able to track the state of dynamic power system under fault condition with lower error compared to UPF. In conclusion, the MUPF offered a robust algorithm for tracking the state of the power system using PMU data.

Acknowledgements

Financially supported by the Ministry of Education, Youth and Sports of the Czech Republic and the European Union (European Structural and Investment Funds - Operational Programme Research, Development, and Education) in the frames of the project "International Research Laboratories," Reg. No. CZ.02.2.69/0.0/0.0/18_054/0014685.

References

[1]. J. Zhao, A. Gomez-Exposito, M. Netto, L. Mili, A. Abur, V. Terzija, I. Kamwa, B. C. Pal, A. K. Singh, J. Qi, Z. Huang, A. P. S. Meliopoulos, Power system

dynamic state estimation: Motivations, definitions, methodologies and future work, *IEEE Transactions on Power Systems*, Vol. 34, Issue 4, 2019, pp. 3188-3198.

[2]. K. Emami, T. Fernando, H. H.-C. Iu, H. Trinh, K. P. Wong, Particle filter approach to dynamic state estimation of generators in power systems, *IEEE Transaction on Power System*, Vol. 30, 2014, pp. 2665-2675.

[3]. A. Doucet, A. M. Johansen, A tutorial on particle filtering and smoothing: Fifteen years later, *Handbook of Nonlinear Filtering*, Vol. 12, 2009, pp. 656-704.

[4]. S. J. Julier, J. K. Uhlmann, Unscented filtering and nonlinear estimation, *Proceedings of the IEEE*, Vol. 92, Issue 3, 2004, pp. 401-422.

[5]. N. Amor, A. Meddeb, S. Chebbi, EEG source estimation using multiple unscented particle filtering, in *Proceedings of the International Conference on Advanced Systems and Emergent Technologies (IC_ASET'19)*, Hammamet, Tunisia, 9-22 March, 2019, pp. 295-299.

(1156)

A Deep Learning Architecture for Automatic Raga Identification

I. Miliaresi [1] and S. Paschalidou [2]
[1] Ionian University, Corfu, Greece
Tel.: +30 2661087860
E-mail: imiliaresi@ionio.gr
[2] Hellenic Mediterranean University, Rethymno, Greece
Tel.: +30 2831021918
E-mail: pashalidou@hmu.gr

Summary: In this paper we investigate the problem of automatic raga identification in Hindustani music improvisation by applying deep learning techniques. We propose a modular deep learning architecture, composed by convolutional and feed forward neural networks. The proposed method is designed to process data from different modalities, comprising audio recordings and metadata. The methodology focuses on the formulation of the feature input vector with appropriate acoustic descriptors extracted from the audio recordings and metadata parameters. Experiments are carried out on audio recordings as well as manual and automatically extracted annotations for the Hindustani genre from the Saraga open dataset of Indian art music. Our experiments indicate that the integration of two modalities increases the accuracy of raga identification compared to processing only audio features.

Keywords: Raga identification, Deep learning, Convolutional neural networks.

1. Introduction

This work focuses on machine listening identification of ragas in Hindustani music, which is the north Indian style of classical Indian music. Hindustani relies heavily on improvisation and it follows a moveable Do (tonic) system. It is monophonic; hence a performance will typically rely on the melodic development by a single performer. It can be performed either vocally or by an instrument. The soloist is invariably accompanied by a drone instrument (called tanpura), which is used to create a sonic canvas as a tonal reference, and may be also accompanied by a harmonium, which will repeat or imitate phrases of the soloist's melodic line with a slight delay.

Performing a raga involves a number of sections, which can be different according to the specific sub-genre, and may include a non-metrical melodic improvisational introduction, one or more compositions, as well as rhythmic improvisation sections; all adhering to the same raga for a single performance. A raga is a form of melodic mode lying between scale and tune [15]. Broadly speaking, it provides a tonal framework for composition and improvisation. There are more features particular to each raga other than simply the fixed scale that are of paramount importance to the character of the raga. Thus, apart from a fixed set of notes (degrees of the scale), these also include the order and hierarchy of scale degrees, their manner of intonation and ornamentation and their relative duration and importance [17]. A raga can be further characterised by its overall melodic outline in terms of ascent and descent, whereby different degrees of the scale may be used or omitted in ascending versus descending order according to the raga. Finally, characteristic motifs, phrases and ornaments constitute prominent aspects of a raga, and therefore also strong cues for raga identification.

In this paper we formulate a methodology for machine listening raga identification, one that is based on customized features utilizing deep learning architectures. The novelty of the proposed method lies in the combined use of data from different modalities, fusing audio-based features and both automatic and manually annotated metadata into a modular deep learning architecture.

What follows in this paper is organized in six sections. The next section offers an overview and discussion of related work. The third section describes the data collection that was used, while the fourth section details the proposed methodology, mostly focusing on the feature extraction stage and the network architecture. All experiments and results are reported in the fifth section, leading to the concluding part of the work.

2. Related Work

While ragas can be reliably identified by expert listeners, automatic raga classification remains a challenging task under research. Reported research work on Carnatic music ([5, 6, 9, 2, 7, 3, 1, 14, 15, 22]) over numbers that of Hindustani ([16, 13, 10, 8]); the current paper targets the latter genre. Very few studies have been reported using conventional classifiers in Carnatic and Hindstani music. k-Nearest Neighborhood and Support vector machine classifiers have been used in [8]. Gaussian Mixture Model based Hidden Markov Models (HMMs), discrete HMMs and

classification trees have been also utilized in [19]. More recently, machine learning techniques of various neural network architectures were applied, with the integration of convolutional neural networks ([2, 7, 3]), recurrent neural networks and Long-short-term neural networks [1, 10]. Results from transfer learning [15] aggregated methods and supervised learning techniques [4] have been reported. Various features extracted from audio recordings have been tested, mainly Mel-spectrogram, Cepstrum Coefficients, chromogram, pitch and timbre features [13, 10, 11, 22]. Phrases were utilized in [21] using vector space modelling.

We define a two-class raga classification task on recordings of Hindustani vocal music, namely Yaman and Bhairavi. This problem was initially defined in [8], where experiments were carried out on 341 web-extracted audio clips, 60 sec-long each, for both instrumental and vocal music, 194 of which in Yaman and 147 in Bhairavi. Conventional classifiers, K Nearest Neighbor (KNN) and Support vector machine (SVM) classifiers were used in the abovementioned paper, achieving a 92 % classification accuracy.

The novelty of our approach lies in the combined use of data from different modalities, with a fusion of audio-based features and metadata into a modular deep learning architecture, achieving a classification accuracy of 97.41 %.

3. Data Corpus

The most widely used databases of classical Indian music are the CompMusic Art Indian music dataset – and its subset, Dunya Corpora, for Music Information Research in Indian Art Music [20] and the Saraga music research corpus of Indian Art Music [16]. Notwithstanding the mp3-quality of the Saraga music research corpus, it was preferred, as it served the scope of this paper in raising raga classification accuracy by incorporating metadata, which was possible because it includes additional information other than audio. The corpus comprises 107 multitrack (rarely available in Music Information Retrieval) audio recordings of 68 different ragas by 11 singers of Hindustani vocal music. The recordings are accompanied by editorial metadata and time-aligned structural, melodic, and rhythmic (the latter not used for our purposes) annotations, which are either extracted automatically or created manually by expert listeners.

The editorial metadata includes foremost the raga name extracted from the cover-art, manual annotations by Classical Indian Music experts (such as start/end time-stamps of individual sections, section name and melodic transitions, with time-stamped melodic phrase annotations being flagged according to raga-importance as characteristic of the raga or not), and automatically extracted annotations (such as the tonic and the sequence of time-stamped pitches for the lead musician). From the multi-track recordings, the track corresponding to the lead vocalist was only used for the purposes of this work 44100 Hz sampling rate, 16-bit PCM).

For our experiments, we use a subset of subjects of the two ragas under interest, Yaman and Bhairavi. One Hindustani sub-genre was used, ranging from its predominant classical forms, namely Dhrupad and Khayal, to its rather semi-classical forms, namely Thumri.

Given that raga identification is predominantly substantiated upon the melodic structure, the information considered to be most relevant to the task of raga classification among all available meta-data were the c-tonic (automatic extraction), pitch values (automatic extraction) and characteristic melodic phrases (manual annotations).

4. Method Description

4.1. Feature Extraction

At the feature extraction stage, we construct a complex input vector with audio-based features and metadata descriptors. First, audio recordings are preprocessed. The audio file format is converted from mp3 to wav. All signals are normalised and re-sampled to 44100 Hz. The initial recordings are segmented into clips of 3 seconds. Recordings for both classes (Yaman and Bhairavi) vary heavily in duration, ranging between 1081.678 ms and 718.915 ms and as a result, 361 and 202 audio samples per class are produced.

Audio clips are parsed with a 40 ms long moving window with a hop size of 20 ms. A standard zero padding procedure is applied. At each frame, the Discrete Fourier Transform (DFT) is computed and it is given as input to a Mel-filter-bank, with each mel-filter performing a weighted sum of the magnitude of the DFT coefficients that lie inside its frequency range. The logarithm of each filter bank output and the discrete cosine transform of the logarithms is calculated. From the set of Mel frequency cepstrum coefficients (MFCCs), the first coefficient is discarded (as recommended in bibliography), while the following 12 MFCCs are included and the rest are discarded.

In parallel, descriptors from metadata are processed. More specifically the mean pitch, the pitch value of the tonic (ctonic) and the presence of characteristic melodic phrases (mphrase) for each raga are included. All numerical parameters are represented as a number and categorical are transformed to numerical by mapping the state “presence” to numerical vale 0 and “non presence” to 1.

4.2. Network Architecture

The complex input feature vector is processed by a modular deep learning classifier. More specifically, the network is organized into two subnetworks, processing data from the two modalities. The overall model structure is illustrated in Fig. 1.

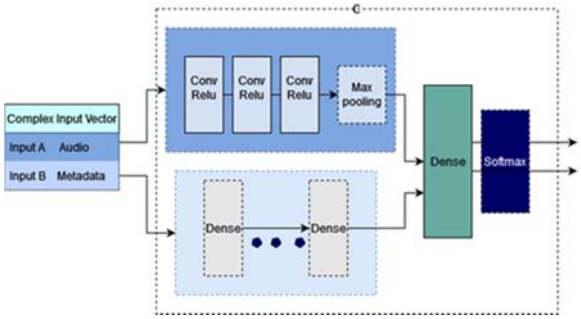


Fig. 1. Overall network architecture consisting of three sub networks. A convolutional branch that processes audio features, a feed forward branch for metadata descriptors and a final feed forward stage that yields the classification decision.

Our model combines convolutional (CNNs) and fully connected (FCN) neural network architectures. The final system setup can be described in detail as follows:

- The first branch consists of three blocks of convolutional – batch normalization layers, with ReLU activation functions consisting of 60 filters of rectangular shape with 57×6 units with 1×1 stride. A max-pooling layer follows with a pool shape of 4×3 and stride of 1×3.
- The second branch consists of two sequentially fully connected layers with 16 and 12 nodes each.
- Finally, the outputs of the two modules are concatenated to a fully connected branch constructed by 1024 nodes and followed by a SoftMax layer of two layers to outcome the classification decision.

The model is trained for 300 epochs with the Adam gradient descent algorithm, with a learning rate of 0.0001. Categorical cross entropy loss was optimized towards validation accuracy. An early-stopping criterion of 30 epochs is applied to reduce the probability of network overfitting. Dropout regularization is also used, with the dropout value set to 0.5 for fully connected and convolutional layers.

The network is evaluated using a 5-fold cross-validation scheme. A 90 % subset of the data collection is used as a training dataset and the remaining 10 % as a testing dataset.

5. Experiments and Results

At an early phase in our experiments, audio parametrization methods, linear prediction cepstrum and mel-frequency cepstrum were compared with regard to raga identification. Multiple streamline audio features, linear, Mel frequency Cepstrum and bark spectrum coefficients were opposed. It was proven that information obtained from Mel frequency Cepstrum coefficients analysis increases the ability of raga discrimination.

To evaluate the proposed methodology, we used two system configurations. First, we evaluate the performance of a converted CNN topology, where only audio features are processed. The model is structured with the convolutional subnetwork and the final fully connected module. Then, we experiment on the overall system configuration with the inclusion of the second fully connected sub network, where metadata descriptors are processed.

The final results are expressed in terms of accuracy, in order to compare with the best reported result in the specified task. The classification accuracies of the audio-based classifier for all five folds in terms of % are [94.44, 96.29, 88.88, 96.29, 95.37], with an average value of 94.25. The classification accuracy results (%) for the overall network architecture are [97.14, 98.07, 99.19, 97.31, 95.37] with an average score of 97.41.

Our experimental results show that the classification accuracy score is increased by 3.16 % units when integrating metadata information. As observed in the reported classification accuracies, our model achieved a competitive accuracy of 97.41 % compared to 92 % of the reference paper [8], even though we experimented on a smaller dataset.

6. Conclusions

Experimental results indicate that fusion of data from multiple modalities into deep learning architectures can contribute to raising the accuracy of machine listening techniques in raga identification tasks. It has been shown, that information obtained from Mel cepstrum prediction analysis and metadata has increased the ability of discriminating between Yaman and Bairavi ragas of Hindustani music, achieving a classification accuracy of 97.41 % in the specific task.

Including a larger number of raga classes, artists and types of instruments are some of our future plans for expanding our current experiments.

References

[1]. S. Madhusudhan, G. Chowdhary, DeepSRGM – sequence classification and ranking in Indian classical music via deep learning, in *Proceedings of Music Information Retrieval Conference (ISMIR'19)*, 2019, pp. 533-540.

[2]. J. Siji, M. S. Sinith, R S. Sudheesh, P. P. Lalu, Classification of Indian classical Hindustani music based on raga using deep learning, in *Proceedings of the IEEE Recent Advances in Intelligent Computational Systems (RAICS'20)*, 2020, pp. 110-113.

[3]. A. Vasudevan, H. Ramasangu, A hybrid cluster-classifier model for Hindustani raga classification, in *Proceedings of the IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT'21)*, 2021, pp. 1-6.

[4]. V. Kaimal, S. S. Barde, Implementation of a supervised learning model for raga identification in Carnatic music, in *Proceedings of the Asian Conference on Innovation in Technology (ASIANCON'21)*, 2021, pp. 1-6.

[5]. A. C. Waghmare, B. A. Sonkamble, Raga identification techniques for classifying Indian classical music: A survey, *International Journal of Signal Processing Systems (IJSPS)*, Vol. 5, Issue 4, 2017, pp. 130-135.

[6]. R. Shastry, A. Yanping Bai, Signal processing and machine learning approaches and evaluation for Indian classical music, *IOP Conference Series: Materials Science and Engineering*, Vol. 1116, 2021, 012188.

[7]. A. Anand, Raga identification using convolutional neural network, in *Proceedings of the Second International Conference on Advanced Computational and Communication Paradigms (ICACCP'19)*, 2019, pp. 1-6.

[8]. D. Joshi, J. Pareek, P. Ambatkar, Indian classical raga identification using machine learning, in *Proceedings of the International Semantic Intelligence Conference (ISIC'21)*, 2021.

[9]. P. Sutar, M. Saitwal, S. Wakchaure, R. Shinde, P. Sawdekar, A survey on raga identification using machine learning techniques, *International Journal of Innovative Research in Technology (IIRT),* Vol. 6, Issue 11, 2020, pp. 691-694.

[10]. S. Chowdhuri, Phononet: Deep learning for raga identification in Indian classical music, in *Proceedings of the International Conference on Multimedia Retrieval (ICMR'19)*, 2019, pp. 197-201.

[11]. H. Dodia, S. Pandey, N. P. Chawande, Identification of raga by machine learning with chromogram, *International Research Journal of Engineering and Technology*, Vol. 07 Issue 05, 2020, pp. 3220-3225.

[12]. D. P. Shah, N. M. Jagtap, P. T. Talekar, K. Gawande, Raga recognition in Indian classical music using deep learning, in *Proceedings of the International Conference on Artificial Intelligence in Music, Sound, Art and Design (Evo-MUSART'21)*, 2021, pp. 248-263.

[13]. K. C. Waghmare, B. A. Sonkamble, Analyzing acoustics of Indian music audio signal using timbre and pitch features for raga identification, in *Proceedings of the 3rd International Conference on Imaging, Signal Processing and Communication (ICISPC'19)*, 2019, pp. 42-46.

[14]. A. Srinivasamurthy, G. K. Koduri, S. Gulati, V. Ishwar, X. Serra, Corpora for music information research in Indian art music, in *Proceedings of the International Computer Music Conference Proceedings (ICMC'14)*, 2014.

[15]. H. S. Powers, R. Widdess, India, subcontinent of, in The New Grove Dictionary of Music and Musicians. *Macmillan*, London, 2001.

[16]. A. Srinivasamurthy, S. Gulati, R. Caro Repetto, X. Serra, Saraga: Open datasets for research on Indian art music, *Empirical Musicology Review*, Vol. 16, Issue 1, 2021, pp. 85-98.

[17]. J. Bor, S. Rao, W. Van der Meer, J. Harvey, The Raga Guide: A Survey of 74 Hindustani Ragas, *Nimbus Records*, 1999.

[18]. S. Sreejith, R. Rajan, Raga recognition in Indian Carnatic music using transfer learning, in *Proceedings of the Fourth International Conference on Microelectronics, Signals & Systems (ICMSS'21)*, 2021, pp. 1-5.

[19]. S. Alekh, Automatic raga recognition in Hindustani classical music, *ArXiv Preprint*, 2017, abs/1708.02322.

[20]. A. Srinivasamurthy, G. K. Koduri, S. Gulati, V. Ishwar, X. Serra, Corpora for music information research in Indian art music, in *Proceedings of the International Computer Music Conference (ICMC/SMC'14),* 2014.

[21]. S. Gulati, J. Serrà, V. Ishwar, S. Sentürk, X. Serra, Phrase-based raga recognition using vector space modeling, in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'16)*, 2016, pp. 66-70.

[22]. P. Dighe, P. Agarwal, H. Karnick, S. Thota, B. Raj, Scale independent raga identification using chromogram patterns and swara based features, in *Proceedings of the IEEE International Conference on Multimedia and Expo Workshops (ICMEW'13)*, 2013, pp. 1-4.

(1252)

Encoder-decoder Network with Guided Transmission Map: Architecture

Le-Anh Tran and Dong-Chul Park
Myongji University, Yongin, South Korea
E-mail: leanhtran@mju.ac.kr, parkd@mju.ac.kr

Summary: An insight into the architecture of the Encoder-Decoder Network with Guided Transmission Map (EDN-GTM), a novel and effective single image dehazing scheme, is presented in this paper. The EDN-GTM takes a conventional RGB hazy image in conjunction with the corresponding transmission map estimated by the dark channel prior (DCP) approach as inputs of the network. The EDN-GTM adopts an enhanced structure of U-Net developed for dehazing tasks and has shown state-of-the-art performances on benchmark dehazing datasets in terms of PSNR and SSIM metrics. In order to give an in-depth understanding of the well-designed architecture which largely contributes to the success of the EDN-GTM, extensive experiments and analysis from selecting the core structure of the scheme to investigating advanced network designs are presented in this paper.

Keywords: Image dehazing, Dark channel prior, Spatial pyramid pooling, U-Net, Generative networks.

1. Introduction

Generally, haze can be considered to be one of the most fundamental phenomena causing image visibility degradation. Accurate estimation of the transmission map in a hazy image, however, has been a major obstacle in performing dehazing [1]. Numerous haze removal approaches have been proposed and most of them have achieved significant progress. Dehazing algorithms can be divided into two types: traditional methods and deep learning-based methods. Traditional approaches apply handcrafted models to perform haze removal tasks while deep learning-based schemes adopt convolutional neural networks (CNNs) in their systems. Both types have their own advantages.

In order to take the advantages of both types of dehazing algorithms, the Encoder-Decoder Network with Guided Transmission Map (EDN-GTM) has been proposed in our preliminary work which utilizes the transmission map extracted by the dark channel prior (DCP) as an additional input channel of a CNN model in order to achieve an improved dehazing performance [1]. To further provide an insight into the well-designed architecture that leads to the EDN-GTM's improvement, various experiments and analysis starting from selecting the core structure of the EDN-GTM scheme to evaluating the effect of every single modification on the network are presented in this paper.

2. The EDN-GTM Scheme

The EDN-GTM scheme is illustrated in Fig. 1. In terms of the generator network, the EDN-GTM utilizes the transmission map estimated by DCP as an additional input channel of a U-Net-based generative network [2]. To further customize U-Net for dehazing task, the EDN-GTM applies three main modifications: 1) a spatial pyramid pooling (SPP) module is plugged into the bottleneck of U-Net; 2) ReLU activation is replaced with Swish activation; and 3) one convolution layer with the filter size of 3×3 is appended in each of the main convolution stages to increase the receptive field and capture more high-level features. In terms of the discriminator design, the encoding part of U-Net is utilized as the base network in order to encourage the discriminator to be equally capable of extracting and analyzing advanced features with the generator such that the two networks compete with each other to boost their performances.

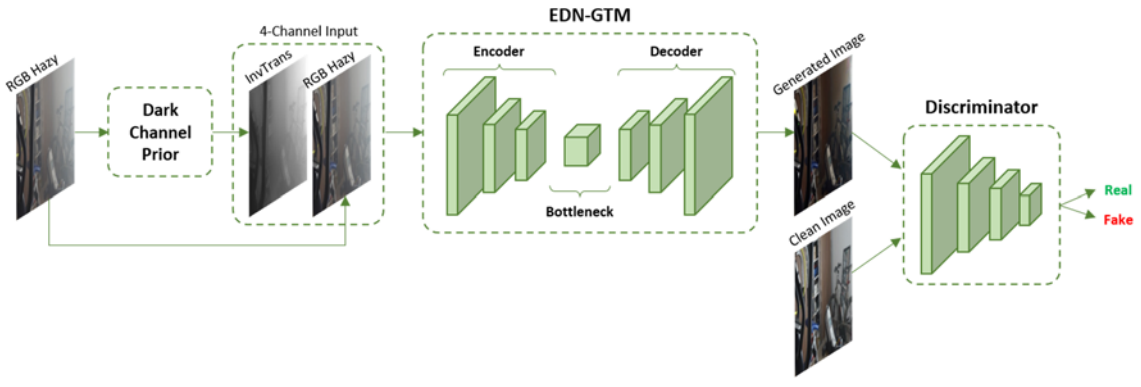


Fig. 1. The EDN-GTM scheme.

3. Architectural Analysis

In order to determine more acceptable architecture for the EDN-GTM, extensive experiments and analysis on different network configurations are conducted for evaluating the influence of each change of the network architecture on the dehazing performance. Based on the experimental results and evaluations, the optimal network design and the most effective data augmentation methods are selected to present the primary performance of the EDN-GTM scheme on benchmark dehazing datasets.

Overall, we have examined 12 different network configurations and 4 data augmentation methods: 1) core structures of the scheme including segmentation-like structure (S-U-Net) and generative structure (G-U-Net); 2) impact of the transmission map channel (G-U-Net 4-C); 3) advanced structures such as cross-stage partial module (CSP G-U-Net 4-C) and spatial pyramid pooling module (SPP G-U-Net 4-C); 4) attention mechanisms including spatial attention (SPP G-U-Net 4-C SAM) and channel attention (SPP G-U-Net 4-C CAM); 5) effects of various data augmentation methods such as random crop, horizontal flip, cutout, and mosaic; 6) effect of different activations including ReLU, Leaky ReLU, Swish, and Mish; and 7) impact of receptive field. Those configurations have been examined on 4 benchmark dehazing datasets (I-HAZE, O-HAZE, Dense-HAZE, and NH-HAZE [1]) in order to give comprehensive and concrete assessments of the EDN-GTM's design.

Typical visual results are illustrated in Fig. 2 while the quantitative evaluations are summarized in Table 1. Note that only 6 representative network designs, specifically S-U-Net, G-U-Net, G-U-Net 4-C, SPP G-U-Net 4-C (ReLU), SPP G-U-Net 4-C (Swish), and EDN-GTM, are chosen to provide the visual and quantitative results. The results demonstrate that a dramatic improvement in performance has been achieved when comparing the outcome of the original U-Net with that of the EDN-GTM scheme.

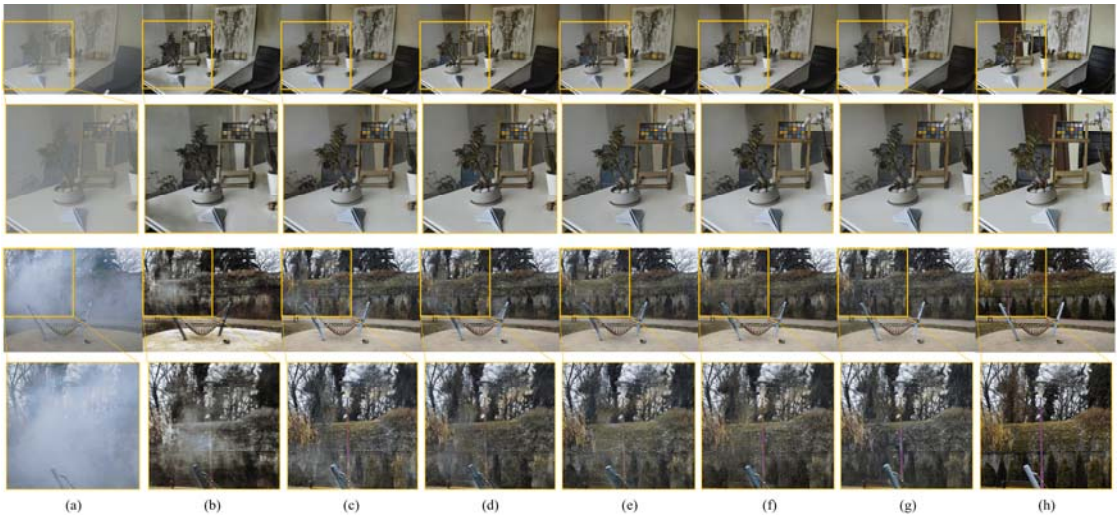


Fig. 2. Results across different network designs: (a) Input, (b) S-U-Net, (c) G-U-Net, (d) G-U-Net 4-C, (e) SPP G-U-Net 4-C (ReLU), (f) SPP G-U-Net 4-C (Swish), (g) EDN-GTM, and (h) Ground Truth (top: indoor, bottom: outdoor).

Table 1. Quantitative dehazing results of different network designs on 4 benchmark datasets.

	S-U-Net		G-U-Net		G-U-Net 4-C		SPP G-U-Net 4-C (ReLU)		SPP G-U-Net 4-C (Swish)		EDN-GTM	
Dataset	**PSNR**	**SSIM**	**PSNR**	**SSIM**	**PSNR**	**SSIM**	**PSNR**	**SSIM**	**PSNR**	**SSIM**	**PSNR**	**SSIM**
I-HAZE	19.82	0.7805	21.04	0.8056	21.39	0.8115	22.20	0.8092	22.66	**0.8311**	**22.90**	0.8270
O-HAZE	20.96	0.7413	23.10	0.8099	23.15	0.8159	23.27	0.8174	23.43	**0.8283**	**23.46**	0.8198
Dense-HAZE	14.18	0.2954	14.96	0.4952	15.19	0.5062	15.43	0.5147	**15.46**	**0.5359**	15.43	0.5200
NH-HAZE	16.79	0.6368	19.18	0.6892	19.52	0.6877	19.73	0.7011	19.80	0.7064	**20.24**	**0.7178**

4. Results on Benchmark Datasets

In this section, we present the performances of the EDN-GTM scheme on various benchmark datasets for image dehazing tasks including I-HAZE, O-HAZE, Dense-HAZE, and NH-HAZE.

The quantitative results of the EDN-GTM scheme on I-HAZE and O-HAZE datasets compared with those of other modern dehazing approaches are first summarized in Table 2, while typical visual dehazing results produced by various dehazing methods on I-HAZE and O-HAZE datasets are presented in

Fig. 3(a) and Fig. 3(b), respectively. As shown in Table 2, the EDN-GTM scheme achieves the best dehazing performance in PSNR metric (22.90 dB) while showing the second-best result in SSIM (0.8270) on I-HAZE dataset. On the other hand, on O-HAZE dataset, the EDN-GTM scheme produces the second-best result in PNSR measure (23.46 dB) while showing the best performance in terms of SSIM (0.8198). Further quantitative results of various haze removal algorithms on I-HAZE and O-HAZE datasets are summarized in Table 2, where the best and the second-best results are shown in red and blue colors, respectively.

The quantitative results on Dense-HAZE and NH-HAZE datasets produced by the EDN-GTM scheme are also compared with those of other approaches and summarized in Table 3, where the best and the second-best results are indicated in red and blue colors, respectively. Note that Dense-HAZE and NH-HAZE datasets are more challenging than I-HAZE and O-HAZE datasets. On Dense-HAZE dataset, the EDN-GTM scheme provides the second-best result in PNSR measure (15.43 dB) while giving the best result in SSIM metric (0.5200). On NH-HAZE dataset, the EDN-GTM scheme convincingly achieves the best performances in both PSNR (20.24 dB) and SSIM (0.7178) measures. Some visual dehazing results from various methods on Dense-HAZE and NH-HAZE datasets are presented in Fig. 3(c) and Fig. 3(d), respectively.

Table 2. Quantitative dehazing results on I-HAZE and O-HAZE datasets.

	I-HAZE		O-HAZE	
Method	**PSNR**	**SSIM**	**PSNR**	**SSIM**
DCP (TPAMI'10) [3]	14.43	0.7516	16.78	0.6532
CAP (TIP'15) [4]	12.24	0.6065	16.08	0.5965
MSCNN (ECCV'16) [7]	15.22	0.7545	17.56	0.6495
AOD-Net (ICCV'17) [6]	13.98	0.7323	15.03	0.5385
PPD-Net (CVPRW'18) [8]	22.53	0.8705	24.24	0.7205
EDN-GTM	22.90	0.8270	23.46	0.8198

Table 3. Quantitative dehazing results on Dense-HAZE and NH-HAZE datasets.

	Dense-HAZE		NH-HAZE	
Method	**PSNR**	**SSIM**	**PSNR**	**SSIM**
DCP (TPAMI'10) [3]	10.06	0.3856	10.57	0.5196
DehazeNet (TIP'16) [9]	13.84	0.4252	16.62	0.5238
AOD-Net (ICCV'17) [6]	13.14	0.4144	15.40	0.5693
MSBDN (CVPR'20) [10]	15.37	0.4858	19.23	0.7056
AECR-Net (CVPR'21) [11]	15.80	0.4660	19.88	0.7173
EDN-GTM	15.43	0.5200	20.24	0.7178

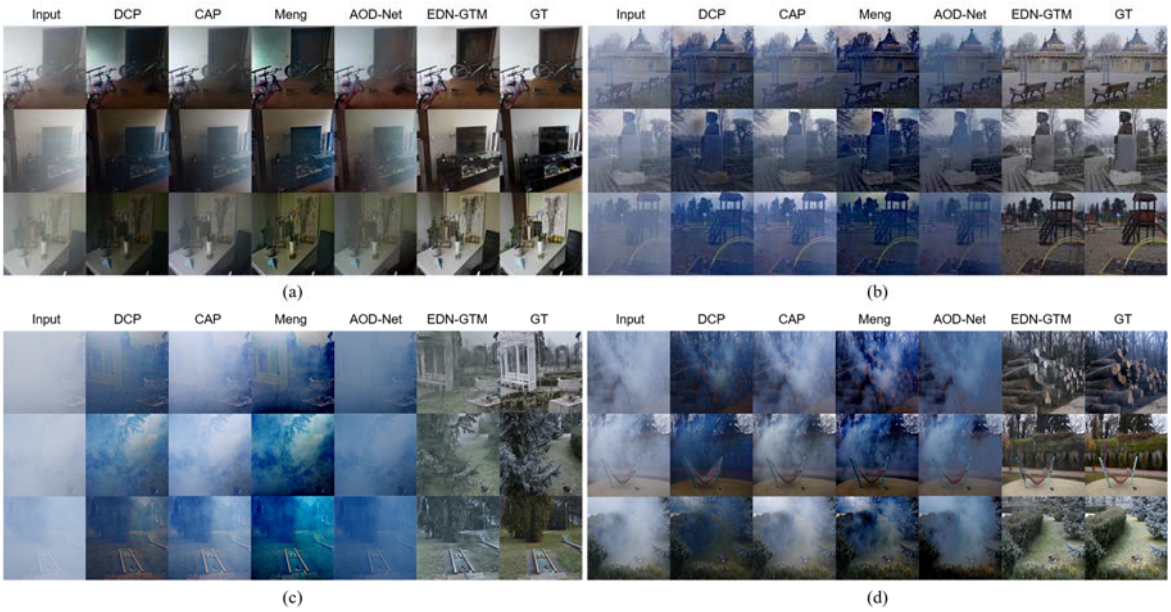


Fig. 3. Results of various single image dehazing methods (DCP [3], CAP [4], Meng et al. [5], AOD-Net [6], and EDN-GTM) on high-resolution datasets: (a) I-HAZE, (b) O-HAZE, (c) Dense-HAZE, and (d) NH-HAZE.

As summarized in Table 2, Table 3, and Fig. 3, the EDN-GTM scheme is able to achieve favorable results on all the datasets examined in our experiments when compared with other conventional and recent dehazing methods. It implies that the EDN-GTM scheme's architecture has been well-designed as the scheme can perform efficiently on haze removal problems. Moreover, we notice that the transmission map plays a critical role in guiding the network to achieve promising results in image dehazing tasks.

5. Conclusions

An extensive analysis of the architectural features of the Encoder-Decoder Network with Guided Transmission Map (EDN-GTM), an effective single image dehazing scheme, is presented in this paper. In order to find an optimal architecture of the EDN-GTM, various features of architecture including core structure, transmission map channel, spatial pyramid pooling, activation functions, and receptive field variations are thoroughly studied. Various experiments examining the effect of every single change on the architectural features provide an in-depth understanding of the optimal design of the EDN-GTM scheme. Future work includes a study on the applicability of the EDN-GTM to an image pre-processing tool in hazy-weather circumstances to remove haze efficiently for autonomous driving systems.

References

[1]. L.-A. Tran, S. Moon, D.-C. Park, A novel encoder-decoder network with guided transmission map for single image dehazing, *Procedia Computer Science*, Vol. 204, 2022, pp. 682-689.

[2]. O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional networks for biomedical image segmentation, in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI'15)*, Munich, Germany, 2015, pp. 234-241.

[3]. K. He, J. Sun, X. Tang. Single image haze removal using dark channel prior, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 33, Issue 12, 2011, pp. 2341-2353.

[4]. Q. Zhu, J. Mai, L. Shao, A fast single image haze removal algorithm using color attenuation prior, *IEEE Transactions on Image Processing*, Vol. 24, Issue 11, 2015, pp. 3522-3533.

[5]. G. Meng, Y. Wang, J. Duan, S. Xiang, C. Pan, Efficient image dehazing with boundary constraint and contextual regularization, in *Proceedings of the IEEE International Conference on Computer Vision (ICCV'13)*, Sydney, Australia, 2013, pp. 617-624.

[6]. B. Li, X. Peng, Z. Wang, J. Xu, D. Feng, AOD-Net: All-in-one dehazing network, in *Proceedings of the IEEE International Conference on Computer Vision (ICCV'17)*, Venice, Italy, 2017, pp. 4770-4778.

[7]. W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, M.-H. Yang, Single image dehazing via multi-scale convolutional neural networks, in *Proceedings of the European Conference on Computer Vision (ECCV'16),* 2016, pp. 154-169.

[8]. H. Zhang, V. Sindagi, V. M. Patel, Multi-scale single image dehazing using perceptual pyramid deep network, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW'18)*, Salt Lake City, USA, 2018, pp. 902-911.

[9]. B. Cai, X. Xu, K. Jia, C. Qing, D. Tao, Dehazenet: An end-to-end system for single image haze removal, *IEEE Transactions on Image Processing*, Vol. 25, Issue 11, 2016, pp. 5187-5198.

[10]. H. Dong, J. Pan, L. Xiang, Z. Hu, X. Zhang, F. Wang, M.-H. Yang, Multi-scale boosted dehazing network with dense feature fusion, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR'20)*, Seattle, USA, 2020, pp. 2157-2167.

[11]. H. Wu, Y. Qu, S. Lin, J. Zhou, R. Qiao, Z. Zhang, Y. Xie, L. Ma, Contrastive learning for compact single image dehazing, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR'21)*, Nashville, USA, 2021, pp. 10551-10560.

(1641)

Variable Projection Support Vector Machines

Tamás Dózsa [1,2] and Péter Kovács [1]
[1] Eötvös Loránd University, Faculty of Informatics, Department of Numerical Analysis, Budapest, Hungary
[2] Institute for Computer Science and Control, Systems and Control Laboratory, Budapest, Hungary
E-mail: dotuaai@inf.elte.hu

Summary: We introduce an extension of the classical support vector machine classification algorithm with adaptive orthogonal transformations. The proposed transformations are realized through so-called variable projection operators. This approach allows the classifier to learn an informative representation of the data during the training process. Furthermore, choosing the underlying adaptive transformations correctly allows for learning interpretable parameters. Since the gradients of the proposed transformations are known with respect to the learnable parameters, we focus on training the primal form the modified SVM objectives using a stochastic subgradient method. We consider the possibility of using Mercer kernels with the proposed algorithms. We construct a case study using the linear combinations of adaptive Hermite functions where the proposed classification scheme outperforms the classical support vector machine approach. The proposed variable projection support vector machines provide a lightweight alternative to deep learning methods which incorporate automatic feature extraction.

Keywords: SVM, Variable projection, Adaptive orthogonal transformations, Classification, Kernel methods.

1. Introduction

Classical machine learning approaches often rely on predefined feature extraction steps applied to the data before the training of the classification or regression algorithms. Commonly used feature extraction steps range from simple statistical methods like principal component analysis (PCA) [5] to more sophisticated adaptive transformations [1, 6, 7, 8, 19]. Many classification tasks can be successfully solved this way (for example in biological signal processing [6, 7]), however the main limitation of this classical approach is that the resulting representation of the data and the weights of the underlying classifier are optimized separately. Practically this means that ensuring an appropriate representation requires apriori information about the structure of the data.

The introduction of convolutional neural networks [9, 10] addresses this limitation by binding the optimization of data representation to the training of the classifier's weights. Feature extraction in this case is done in an adaptive manner and the resulting data representation is optimal *for the classification task.* This idea gave rise to many popular machine learning algorithms especially in image processing [10].

The application of such approaches to real life problems suffers from the fact that the learned weights of the convolution kernels have no physical meaning [1]. The recent introduction of model based neural network architectures, especially the so-called VP-Net [1] aims to remedy this problem. In a VP-Net, the first few layers (VP-Layers) of a neural network implement adaptive orthogonal transformations. Similarly, to convolutional networks, these layers learn informative features of the data. Training a VP-Net however, often involves the optimization of less parameters than in the case of convolution layers. Another advantage of VP-Net is that if the underlying orthogonal transformations were chosen correctly, then the learned parameters of the VP-layers can be interpreted. VP-Net has already been shown to be an appropriate classifier choice for some problems arising in biological signal processing [1, 11] and autonomous vehicle control [8].

In this work, we investigate the possibility of extending the popular SVM classification algorithm with adaptive orthogonal transformations similarly to the idea of VP-Net. Such an extension can be useful in cases, where computational capacity is limited [8], or the classification task does not require the construction of deep neural networks.

The rest of this paper is organized as follows. In Section 2, we discuss the general form of the adaptive orthogonal transformations which will be used to perform automatic feature extraction. In Section 3 we extend the SVM objective functions using these transformations. In Section 4 we discuss the training of the introduced classifiers using stochastic subgradient descent. Section 5 discusses a numerical experiment. Finally, in Section 6 we draw our conclusions and discuss future steps.

2. Variable Projection Operators

We are going to assume that our data consists of $\boldsymbol{x} \in R^N$ vectors. Our aim is to represent $\boldsymbol{x}$ by $n \ll N$ numbers:

$$\boldsymbol{x} \approx \mathrm{P}_{\Phi(\boldsymbol{\eta})}\boldsymbol{x} = \Phi(\boldsymbol{\eta})(\Phi^{+}(\boldsymbol{\eta})x), \qquad (1)$$

where $\Phi(\boldsymbol{\eta}) \in R^{N\times n}$, $\boldsymbol{\eta} \in R^m$, $\Phi^{+}(\boldsymbol{\eta})$ refers to the Moore-Penrose pseudoinverse of the matrix $\Phi(\boldsymbol{\eta})$ and N, n, m are natural numbers. Thus, we represent the data $\boldsymbol{x}$ with the vector $\Phi^{+}(\boldsymbol{\eta})\boldsymbol{x} \in R^n$ and (1) defines the projection of $\boldsymbol{x}$ onto the column space of $\Phi(\boldsymbol{\eta})$. The

columns of $\Phi(\boldsymbol{\eta})$ consist of discrete samplings of a (usually complete and orthonormal) function system in the Lebesgue space $L_2(R)$. The function system that defines $\Phi(\boldsymbol{\eta})$ depends on the parameter vector $\boldsymbol{\eta}$ in a nonlinear fashion and the operator $\mathrm{P}_{\Phi(\boldsymbol{\eta})}$ is referred to as a variable projection operator [2, 19]. Clearly, the feature extraction scheme (1) depends on the parameter $\boldsymbol{\eta}$. A common procedure to obtain a good representation is to solve

$$\min_{\eta \in R^m} r_2(\boldsymbol{\eta}; \boldsymbol{x}) = \min_{\eta \in R^m} ||\boldsymbol{x} - P_{\Phi(\boldsymbol{\eta})}\boldsymbol{x}||_2^2. \quad (2)$$

Provided that the partial derivatives of the functions generating the columns of $\Phi(\boldsymbol{\eta})$ are known with respect to $\boldsymbol{\eta}$, (2) can be solved using a gradient based method [2, 19].

We note that solving the optimization task (2) allows for several adaptive representations for the vector $\boldsymbol{x}$. Suppose the columns of $\Phi(\boldsymbol{\eta})$ consist of smooth functions. Then, once the parameter vector $\boldsymbol{\eta}$ that minimizes (2) has been determined, one could replace $\boldsymbol{x}$ with its smooth approximation $\mathrm{P}_{\Phi(\boldsymbol{\eta})}\boldsymbol{x}$ or the residual transformation $\boldsymbol{x} - \mathrm{P}_{\Phi(\boldsymbol{\eta})}\boldsymbol{x}$. In some applications [6, 8] such data representations are preferable, however in this work, for simplicity, we are only going to discuss the extension of SVM classifiers with the transformation $\boldsymbol{x} \to \Phi^+(\boldsymbol{\eta})(\boldsymbol{x})$.

3. VP-SVM Objectives

Given a training set with q number of examples $\{(\boldsymbol{x}_i, y_i)\}_{1 \le i \le q}$, $\boldsymbol{x}_i \in R^N, y_i \in \{-1,1\}$, the purpose of an SVM classifier is to identify a hyperplane which separates the examples $\boldsymbol{x}_i$. In the simplest (linear) case, the hyperplane is given by its normal vector $\boldsymbol{w} \in R^N$, whose components are found through an optimization process known as training. Most commonly, the training of SVM classifiers is posed as a linear programming problem and convex optimization tools are used to find the optimal hyperplane [12-14]. In real life applications, SVM classifiers often rely on Mercer kernels [3, 4, 12, 13]. These allow us to transform the data examples $\boldsymbol{x}_i$ to a high dimensional reproducing kernel Hilbert space (RKHS), where a separating hyperplane might more easily be identified. In this (nonlinear) case, dual formulations of the above-mentioned linear programming problem are solved using convex optimization. When training such nonlinear SVM classifiers, the dual problems are preferable, as they allow to express the above-mentioned transformations as inner products of the examples $\boldsymbol{x_i}$.

For the proposed VP-SVM classifiers, we consider gradient based algorithms for training. This is because, the adaptive transformations (2) by which we enhance the SVM objectives are optimized using gradient based methods. Fortunately, extensive literature exists on how to train SVM objectives with such methods [3, 4], focusing mostly on the primal form of the optimization problem. For this reason, we take as starting points the SVM objectives discussed in [3]. It is well-known [3], that the primal form of the linear SVM classifier objective can be reduced to minimizing the following expression with respect to $\boldsymbol{w}$ and b:

$$C \sum_{i=1}^{q} \max(0, 1 - y_i (\boldsymbol{w}^T \boldsymbol{x_i} + b)) + ||\boldsymbol{w}||^2, \quad (3)$$

where $\boldsymbol{w} \in R^N, C, b \in R$.

As mentioned before, several strategies [3, 4] exist for solving the optimization problem (3), however for simplicity, in this work we focus on obtaining a solution using a subgradient based method (see Section 4). An adaptive feature extraction step can be easily added to (3) by

$$C \sum_{i=1}^{q} \max(0, 1 - y_i (\boldsymbol{w}^T (\Phi^+(\boldsymbol{\eta})\boldsymbol{x}_i) + b) + + ||\boldsymbol{w}||^2 + R(\boldsymbol{\eta}), \quad (4)$$

where this time $\boldsymbol{w} \in R^n$ and $R(\boldsymbol{\eta})$ is a regulatory term to ensure a good representation of the data and to avoid the problem of vanishing (sub)gradients with respect to $\boldsymbol{\eta}$:

$$R(\boldsymbol{\eta}) := \frac{\alpha}{q} \sum_{i=1}^{q} \frac{||\boldsymbol{x}_i - P_{\Phi(\boldsymbol{\eta})}\boldsymbol{x_i}||_2^2}{||\boldsymbol{x}_i||_2^2} \quad (5)$$

In (5), $\alpha \in R$ is a penalty parameter for the regulatory term. In Section 4 we provide the subgradients of (4) with respect to $\boldsymbol{\eta}$. These can be calculated provided that the partial derivatives of $\Phi(\boldsymbol{\eta})$ are known [1, 2].

One of the main advantages of support vector machines is their ability to be used with Mercer kernels [3]. Even though in this case, usually the dual form of the SVM objective is considered, popular algorithms exist which optimize the primal objective as well [3, 4]. In this work, we consider the objective presented in [3]:

$$\min_{f \in \mathcal{H}} \lambda ||f||_{\mathcal{H}}^2 + \sum_{i=1}^{q} \max\big(0, 1 - y_i f(\boldsymbol{x}_i)\big), \quad (6)$$

where $\lambda = 1/C, \mathcal{H}$ is a reproducing kernel Hilbert space and by the representer theorem [20] we can look for the solution f in the form

$$f(\boldsymbol{x}) = \sum_{i=1}^{q} \beta_i k(\boldsymbol{x}, \boldsymbol{x}_i) \ (\boldsymbol{\beta} \in \mathrm{R}^q), \quad (7)$$

provided that the inner product $k(\cdot,\cdot)$ of $\mathcal{H}$ is known. We omitted the bias parameter b from (6) to simplify the expression, but it could easily be incorporated as well. Because of (7), supplementing the objective (6) with orthogonal transformations naively would yield a problem that has a complexity of $\mathcal{O}(q^2)$ for every update of the parameters. This is because evaluating the term $||f||_{\mathcal{H}}^2$ using (7) requires us to calculate the inner products $k(\boldsymbol{x_i}, \boldsymbol{x_j})$ for every possible index i, j. To overcome this, we propose using instead the modified objective

$$\min_{\boldsymbol{\beta} \in R^q, \boldsymbol{\eta} \in R^m} \sum_{i=1}^{q} \max\big(0, 1 - y_i \sum_{j=1}^{q} \beta_j k(\tilde{\boldsymbol{x}}_i, \tilde{\boldsymbol{x}}_j)\big) + R(\boldsymbol{\eta}), \quad (9)$$

where $R(\boldsymbol{\eta})$ is the regulatory term from (5) and $\tilde{x}_i$ denotes the transformed vector $\Phi^+(\boldsymbol{\eta})\boldsymbol{x}_i$. This notation will be used throughout the rest of the paper. Similarly, to the linear case, this regulatory term is responsible for ensuring that the projections $\mathrm{P}_{\Phi(\boldsymbol{\eta})}\boldsymbol{x}_i$ approximate the examples $\boldsymbol{x_i}$ well, while keeping the problem of vanishing gradients at bay. In our experiments (Section 5) we show that problems can be constructed when the simplified data representation through the proposed method allows for easier separation of the classes.

4. Training the VP-SVM Classifiers

We note that the proposed VP-SVM objectives (4) and (9) are not differentiable everywhere. Because of this, we can only utilize subgradient based methods to perform the minimization task. The subgradients of a function $f: R^N \to R$ are defined at the point $\boldsymbol{x}$ as the set of vectors $\boldsymbol{g} \in R^N$ for which

$$f(\boldsymbol{z}) \geq f(\boldsymbol{x}) + \boldsymbol{g}^T(\boldsymbol{z} - \boldsymbol{x}), \tag{10}$$

for all $\boldsymbol{z}$ in the domain of f. If the function f is convex and differentiable, then the subgradient coincideces with the gradient. The above-mentioned objective functions are subdifferentiable with respect to the learnable parameters. In addition, $\boldsymbol{x}$ is a minimizer point of a convex function f if and only if $0 \in \partial f(\boldsymbol{x})$, or in other words 0 is a subgradient of f at $\boldsymbol{x}$. This property allows for the construction of optimization algorithms using the notion of subgradients.

In [4] an efficient subgradient based algorithm is presented to minimize the objective functions (3) and (6). The adaptation of this algorithm to train VP-SVM will be part of our future work, however in this study we used the stochastic subgradient descent (SSGD) [14] method for training. In each step of the training process, we randomly select a single example $\boldsymbol{x_i}$, calculate the subgradient of the objective function with respect to the learnable parameters, then update the parameters. This in practice means that the objective functions being minimized change from the form presented in section 3. For example, instead of calculating the subgradients of objective (3), in single step of SSGD we would subdifferentiate the modified loss

$$q \cdot C \cdot \max(0, 1 - y_i(\boldsymbol{w}^T\boldsymbol{x_i})) + ||\boldsymbol{w}||^2, \tag{11}$$

where q denotes the total number of examples and C is a hyper-parameter that controls the tradeoff between the margin and the hinge loss. Again, the bias parameter was omitted for simplicity. We can modify the proposed linear VP-SVM objective (4) similarly by

$$J(\boldsymbol{w}, \boldsymbol{\eta})_{\boldsymbol{i}} := q \cdot C \cdot \max(0, 1 - y_i(\boldsymbol{w}^T\tilde{x}_i) + \\ + ||\boldsymbol{w}||^2 + R(\boldsymbol{\eta}, \boldsymbol{x}_i), \tag{12}$$

where for better scalability we also change the regularization term (5) to

$$R(\boldsymbol{\eta}, \boldsymbol{x}_i) := \alpha \cdot \frac{||\boldsymbol{x_i} - P_{\Phi(\boldsymbol{\eta})}\boldsymbol{x_i}||_2^2}{||\boldsymbol{x_i}||_2^2} \tag{13}$$

The subgradients of (12) exist with respect to $\boldsymbol{\eta}$ and $\boldsymbol{w}$, and can be expressed as

$$\frac{\partial J(\boldsymbol{w},\boldsymbol{\eta})_i}{\partial \boldsymbol{w}} = \begin{cases} \boldsymbol{w} - C \cdot q \cdot y_i\tilde{x}_i, if\ s > 0 \\ \boldsymbol{w}, else \end{cases}, \tag{14}$$

where $s := 1 - y_i(\boldsymbol{w}^T\tilde{x}_i)$ and

$$\frac{\partial J(\boldsymbol{w},\boldsymbol{\eta})_i}{\partial \boldsymbol{\eta}} = \begin{cases} -C \cdot q \cdot y_i \cdot \frac{\partial \tilde{x}_i}{\partial \boldsymbol{\eta}} + \frac{\partial R(\boldsymbol{\eta},\boldsymbol{x}_i)}{\partial \boldsymbol{\eta}}, if\ s > 0 \\ \frac{\partial R(\boldsymbol{\eta},\boldsymbol{x}_i)}{\partial \boldsymbol{\eta}}, else \end{cases}, \tag{15}$$

In (15), the gradients $\frac{\partial \Phi^+(\boldsymbol{\eta})x_i}{\partial \boldsymbol{\eta}} = \frac{\partial \tilde{x}_i}{\partial \eta}$ and $\frac{\partial R(\boldsymbol{\eta},\boldsymbol{x}_i)}{\partial \boldsymbol{\eta}}$ can be calculated provided that the partial derivatives of $\Phi(\boldsymbol{\eta})$ are known with respect to $\boldsymbol{\eta}$ [1, 2]. For the exact formulas we refer to [1]. Once (14) and (15) have been calculated, the stochastic subgradient descent algorithm updates the parameters with

$$\boldsymbol{\eta} \to \boldsymbol{\eta} - \gamma_i \cdot \frac{\partial J(\boldsymbol{w},\boldsymbol{\eta})_i}{\partial \boldsymbol{\eta}} \text{ and} \\ \boldsymbol{w} \to \boldsymbol{w} - \gamma_i \cdot \frac{\partial J(\boldsymbol{w},\boldsymbol{\eta})_i}{\partial \boldsymbol{w}} \tag{16}$$

If the learning rate γ_i is sufficiently small, then the above-described stochastic subgradient descent algorithm is guaranteed to converge [14].

Similarly, to the linear case, the proposed nonlinear version of the VP-SVM objective (9) can be trained using the SSGD algorithm. In this case, the objective to be optimized can be expressed as

$$J(\boldsymbol{\beta}, \boldsymbol{\eta})_i := qC \max\left(0, 1 - y_i \sum_{j=1}^{q} \beta_j k(\tilde{x}_i, \tilde{x}_j)\right) + \\ + R(\boldsymbol{\eta}, \boldsymbol{x}_i), \tag{17}$$

where $R(\boldsymbol{\eta}, \boldsymbol{x}_i)$ is defined by (13). The subgradients of (17) with respect to the learnable parameters are given as

$$\frac{\partial J(\boldsymbol{\beta},\boldsymbol{\eta})_i}{\partial \boldsymbol{\beta_j}} = \begin{cases} -Cq\ y_i\ k(\tilde{x}_i, \tilde{x}_j), if\ u > 0 \\ 0, else \end{cases}, \tag{18}$$

and

$$\frac{\partial J(\boldsymbol{\beta},\boldsymbol{\eta})_i}{\partial \boldsymbol{\eta}} = \begin{cases} \frac{\partial R(\boldsymbol{\eta},\boldsymbol{x}_i)}{\partial \boldsymbol{\eta}} - Cq\ y_i \sum_{j=1}^{q} \beta_j \frac{\partial k(\tilde{x}_i,\tilde{x}_j)}{\partial \boldsymbol{\eta}}, if\ u > 0 \\ \frac{\partial R(\boldsymbol{\eta},\boldsymbol{x}_i)}{\partial \boldsymbol{\eta}}, else \end{cases} \tag{19}$$

In (18) and (19) the subgradients are determined by the magnitude of

$$u := 1 - y_i \sum_{j=1}^{q} \beta_j k(\tilde{x}_i, \tilde{x}_j)$$

Furthermore, in (19) the partial derivatives with respect to $\boldsymbol{\eta}$ clearly depend on the choice of the kernel $k(\cdot,\cdot)$. If, for example one chooses the popular radial basis kernel function defined as

$$k(x, y) := e^{\frac{-||x-y||_2^2}{\sigma^2}},$$

then $k(\tilde{x}_i, \tilde{x}_j) := k(\Phi^+(\boldsymbol{\eta})\boldsymbol{x}_i, \Phi^+(\boldsymbol{\eta})\boldsymbol{x}_j)$ becomes

$$e^{\frac{-\left\|\Phi^+(\boldsymbol{\eta})x_i - \Phi^+(\boldsymbol{\eta})x_j\right\|_2^2}{\sigma^2}} = e^{\frac{-\left\|\Phi^+(\boldsymbol{\eta})\,(x_i - x_j)\right\|_2^2}{\sigma^2}}$$

and $\frac{\partial k(\Phi^+(\boldsymbol{\eta})\boldsymbol{x}_i, \Phi^+(\boldsymbol{\eta})\boldsymbol{x}_j)}{\partial \boldsymbol{\eta}}$ can be given as

$$-e^{\frac{-\left\|\Phi^+(\boldsymbol{\eta})(x_i - x_j)\right\|_2^2}{\sigma^2}} \cdot 2 \cdot \Phi^+(\boldsymbol{\eta})(\boldsymbol{x}_i - \boldsymbol{x}_j) \cdot \frac{1}{\sigma^2} \cdot \frac{\partial \Phi^+(\boldsymbol{\eta})(x_i - x_j)}{\partial \boldsymbol{\eta}}$$

For a fixed kernel, in each step of the SSGD algorithm the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$ can be updated similarly to the linear case (16) once the subgradients (18) and (19) have been calculated.

5. Experiments

We demonstrate the utility of the proposed VP-SVM classifier (9) through an example problem. In our case study, the training examples $\boldsymbol{x}_i \in R^N$ $(i = 1, \dots, q)$ consisted of the linear combinations of the first two so-called adaptive Hermite functions [17]. These functions are closely related to the classical Hermite orthogonal polynomials [15, 16]. Namely, let

$$h_k(x), \quad (x \in R)$$

denote the k-th Hermite polynomial. These form a complete orthogonal system in the weighted Lebesgue space $L_2^w(R)$, where the weight function is given as $w(x) = e^{-x^2}$. We can then introduce the so called Hermite functions by

$$\varphi_k(x) := \\ = h_k(x) \cdot e^{-x^2/2} \cdot \sqrt{\pi^{1/2} 2^k k!} \quad (k \in N, x \in R) \tag{20}$$

The functions defined in (20) form a complete system in $L_2(R)$ and are orthonormal with respect to the usual inner product:

$$\langle \varphi_k, \varphi_j \rangle = \int_{-\infty}^{\infty} \varphi_k(x)\varphi_j(x)dx = \delta_{kj}$$

These functions and their subsequent generalizations form the basis of many signal processing applications [1, 6, 8, 11, 17, 18]. To properly describe the training set for our case study, we need to mention the affine argument transformations of (20)

$$\varphi_k^{\lambda,\tau}(x) := \\ = \sqrt{\lambda}\varphi_k(\lambda(x - \tau)) \; (x, \tau \in R, \lambda > 0, k \in N), \tag{21}$$

known as adaptive Hermite functions [17]. Adaptive Hermite functions have been especially useful for modeling quasi-periodic signals with quasi-compact support.

We constructed each example for our dataset by fixing the affine parameters $\boldsymbol{\eta} := (\lambda, \tau)$ and taking the linear combinations of the first two adaptive Hermite functions:

$$\boldsymbol{x}_i := c_{0,i}\varphi_0(\boldsymbol{\eta}) + c_{1,i}\varphi_1(\boldsymbol{\eta}), \\ (c_{0,i}, c_{1,i} \in R, i = 1, \dots, q) \tag{22}$$

where $\varphi_k(\boldsymbol{\eta})$ is an equidistant sampling of (21). The parameter $\boldsymbol{\eta} := (\lambda, \tau)$ remained constant for every example $\boldsymbol{x}_i$, and two classes were introduced through linear parameters $c_{0,i}$ and $c_{1,i}$. This was done in a way, that the points given as $(c_{0,i}, c_{1,i})$ were randomly chosen from two concentric circles as shown in Fig. 1.

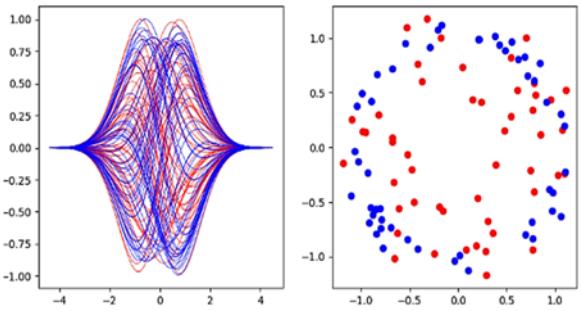


Fig. 1. LEFT: $\boldsymbol{x_i}$ training examples generated by adaptive Hermite functions. RIGHT: Corresponding $(\boldsymbol{c_{0,i}}, \boldsymbol{c_{1,i}})$ parameters from (22).

In the VP-SVM, the columns of the matrix $\Phi(\boldsymbol{\eta})$ also contained the adaptive Hermite functions (21). The idea in this case was to demonstrate, that if the correct $\boldsymbol{\eta} := (\lambda, \tau)$ are learned during the training process, then the examples shown on left side of Fig. 1 become more easily separable (the right side of Fig. 1). We minimized the objective function (17) using the methods described in Section 4. We compared the accuracy of VP-SVM on the set with a classical nonlinear SVM classifier also trained using SSGD. Both classifiers used an RBF kernel with the same σ parameter. Accuracy scores on the training and on the test, sets are shown in Fig. 2 during each epoch of the training.

By Fig. 2, the above-described experiment provides a proof of concept for the utility of the proposed VP-SVM classification algorithm. Since many anomalies detection and classification tasks have been studied, where features are extracted by orthogonal transformations [6, 8, 19], we are hopeful that the proposed algorithms can also be used to solve real world problems. VP-SVM provides a lightweight model-based machine learning approach for classification problems. This can be especially useful when model driven machine learning methods have been shown to provide good results, however the available computational capacity does not allow for using deep neural networks [8].

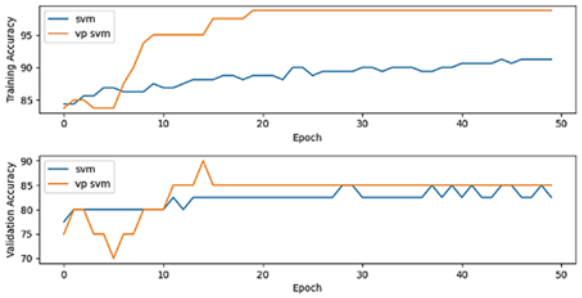


Fig. 2. TOP: training accuracies of VP-SVM (orange) and SVM (blue) using RBF kernels. BOTTOM: accuracy scores on unseen (validation) data at different epochs.

6. Conclusion

By utilizing variable projections, we proposed novel extensions of the usual SVM objectives. We discussed a training scheme using stochastic subgradient method for the proposed classifiers. We demonstrated the efficiency of our method through a simulated test scenario. The proposed VP-SVM extends the classical SVM classifiers with an automatic feature extraction scheme defined by adaptive orthogonal transformations. The learned parameters of the feature extraction can be interpreted [1].

The latter property is of utmost importance when applying the proposed classification scheme to biomedical signal processing tasks (such as illness recognition), which will be part of our future work. In addition, VP-SVM provides a lightweight alternative to similar deep learning-based feature learning and classification schemes such as [1, 10]. Using VP-SVM can be helpful in situations where computational capacity is limited [8].

Acknowledgements

This project was supported by the NVKDP Cooperative Doctoral Program by the Hungarian Ministry of National Development and the National Research, Development and Innovation Fund. This project was supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences.

References

[1]. P. Kovács, et al., VPNET: Variable projection networks, *International Journal of Neural Systems*, Vol. 32, Issue 1, 2022, 21500544.

[2]. G. H. Golub, V. Pereyra, Separable nonlinear least squares: The variable projection method and its applications, *Inverse Problems*, Vol. 19, Issue 2, 2003, R1.

[3]. O. Chapelle, Training a support vector machine in the primal, *Neural Computation*, Vol. 19, Issue 5, 2007.

[4]. S. Shalev-Shwartz, et al., Pegasos: Primal estimated subgradient solver for SVM, *Mathematical Programming*, Vol. 127, Issue 1, 2011, pp. 3-30.

[5]. H. Abdi, L. J. Williams, Principal component analysis, *Wiley Interdisciplinary Reviews: Computational Statistics*, Vol. 2, Issue 4, 2010, pp. 433-459.

[6]. T. Dózsa, G. Bognár, P. Kovács, Ensemble learning for heartbeat classification using adaptive orthogonal transformations, in *Proceedings of the International Conference on Computer Aided Systems Theory (EUROCAST'19)*, 2019, pp. 355-363.

[7]. P. Addison, Wavelet transforms and the ECG: A review, *Physiological Measurement*, Vol. 26, Issue 5, 2005, R155-99.

[8]. T. Dózsa, J. Radó, J. Volk, Á. Kisari, A. Soumelidis, P. Kovács, Road abnormality detection using piezoresistive force sensors and adaptive signal models, *IEEE Transactions on Instrumentation and Measurement*, Vol. 71, 2022, pp. 1-11.

[9]. J. Schmidhuber, Deep learning in neural networks: An overview, *Neural Networks*, Vol. 61, 2015, pp. 85-117,

[10]. A. Krizhevsky., I. Sutskever, G. Hinton, ImageNet classification with deep convolutional networks, in *Proceedings of the Conference Neural Information Processing Systems (NIPS'12)*, 2012, pp. 1-9.

[11]. T. Dózsa, C. Böck, G. Bognár, J. Meier, P. Kovács, Color classification of visually evoked potentials by means of Hermite functions, in *Proceedings of the 55th Asilomar Conference on Signals, Systems, and Computers*, 2021, pp. 251-255.

[12]. S. Boyd, L. Vandenberghe, Convex Optimization, *Cambridge University Press*, 2004.

[13]. S. Fine, K. Scheinberg, Efficient SVM training using low-rank kernel representations, *J. of Mach. Learning Res.*, Vol. 2, 2001, pp. 242-264.

[14]. J. Shawe-Taylor, S. Shiliang, A review of optimization methodologies in support vector machines, *Neurocomputing*, Vol. 74, Issue 17, 2011, pp. 3609-3618.

[15]. G. Szegő, Orthogonal Polynomials, 3rd Ed., *AMS Colloquium Publications*, 1967.

[16]. W. Gautschi, Orthogonal Polynomials, Computation and approximation, in Numerical Mathematics and Scientific Computation, *Oxford University Press*, 2004.

[17]. T. Dózsa, P. Kovács, ECG signal compression using adaptive Hermite functions, in *Proceedings of the International Conference on ICT Innovations*, 2015, pp. 245-254.

[18]. P. Kovács, C. Böck, T. Dózsa, J. Meier, M. Huemer, Waveform modeling by adaptive weighted Hermite functions., in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'19)*, 2019, pp. 1080-1084.

[19]. G. H. Golub, V. Pereyra, Separable nonlinear least squares: The variable projection method and its applications, *Inverse Problems*, Vol. 19, 2003, R1.

[20]. G. S. Kimeldorf, G. Wahba, A correspondence between Bayesian estimation on stochastic processes and smoothing by splines, *The Annals of Mathematical Statistics*, Vol. 41, Issue 2, 1970, pp. 495-502.

(1761)

Synchronization in Multi-channel Burst Mode Demodulator for DVB-RCS Satellite Receiver: A Multi-core DSP Implementation

Laxmaiah P., Dileep K. G. and Hari Prasad S. V.
Centre for Development Of Telematics (C-DOT),
Electronic City Phase-1, Bangalore-560100, India
E-mail: laxman_p@cdot.in, dileepkg@cdot.in, svhari@cdot.in

Summary: This paper presents the design and implementation of a multi-channel burst mode demodulator for DVB-RCS satellite receivers. As satellite terminals are power limited and have a small antenna aperture, the use of powerful error correction codes at very low SNR is mandatory. The implemented demodulator is able to achieve a good synchronization of time, frequency and phase under very low SNR. The proposed design efficiently uses the multi-core DSP for multi-channel implementation. As each channel requires complex signal processing stages, the implementation is carried out by considering proper trade-off between implementation complexity and processing time. Simulations and functional verification of the demodulation algorithm are performed in Matlab and algorithm implementation is based on fixed-point C. The receiver design is prototyped for 10 MF-TDMA channels in Texas Instruments (TI) multi-core C66x DSPs.

Keywords: Satellite, Multi-core DSP, DVB-RCS, RCST, MF-TDMA, Synchronization, Design prototyped in DSP.

1. Introduction

Satellite communication systems are generally operated for large downlink coverage. However, recent applications like DVB-RCS standard [1-2], an interactive on-demand multimedia satellite communication system, demands an uplink channel from the terminal via the satellite to the service provider's gateway/hub. The satellite terminals installed at the user location are in general power limited and have a small antenna aperture. Because of this antenna power limitation, strong error-correcting codes, like turbo codes, need to be applied and because of small antenna aperture, the signal to noise ratio (SNR) remains relatively rather low. Hence the use of powerful error-correcting coding schemes, permitting operation at very low bit energy to noise density ratios (E_b/N_0), is almost mandatory. The turbo codes require almost ideal synchronization of time, frequency and phase at the receiver to achieve maximum coding gain. Moreover, the phase noise performance of the components used in relatively cheap terminals is not up to the standards. So, because of all these factors, in the current DVB-RCS demodulator designs, synchronization has been proven to be one of the major limitations on achievable performance.

For a receiver to estimate all the channel parameters, DVB-RCS standard provides a preamble along with each data burst. To increase the spectral efficiency, the DVB-RCS standard kept the length of this preamble rather short. As DVB-RCS standard is using FDMA/TDMA modulation techniques in order to further improve the utilization of satellite capacity, the hub simultaneously receives MF-TDMA signals from several terminals. So, the hub architecture is generally based on multi-channel demodulator which simultaneously demodulates a number of bursts in different channels having different data rates. In this paper, we implemented a multi-channel burst mode demodulator algorithm with an estimation of all the synchronization parameters using the preamble.

Various types of frequency offset and timing offset estimations using correlation methods are covered in [4-8]. The main thresholds for the design of estimator in real applications are its estimation accuracy at low E_b/N_0, its estimation range and the most important, trade-off between implementation complexity and computational time. In this design, multi-channel demodulator having complex and multi-domain signal processing stages is realized in multi-core DSP by efficiently utilizing the eight parallel cores of the DSP. A data-aided feedforward approach is adopted here for synchronization of all parameters. The design proposes the implementation of a multi-channel receiver with 10 MF-TDMA channels that are processed in parallel. As each channel is processed independently, the synchronization of a single channel is explained in this paper. The functional simulation of proposed design is carried out in MATLAB.

This paper is organized as follows: Section 2 of this paper introduces the system model and burst structure. Section 3 discusses matched filtering. The timing synchronization of QPSK burst in the presence of frequency and phase offset is explained in Section 4. Section 5 deals with frequency and phase offset estimations. Section 6 describes hardware architecture and DSP-based implementation of the algorithm. In Section 7, the prototype test setup is explained. Finally, conclusions are given in Section 8.

2. System Model

The burst structure of different types of burst formats in DVB-RCS system is shown in Fig. 1. After the forward link synchronization [3], return channel

satellite terminal (RCST) sends a logon request in a common signaling channel (CSC) burst to initiate return link synchronization. The acquisition (ACQ) burst is transmitted to achieve coarse synchronization. After logon, synchronization (SYNC) burst is used to maintain the synchronization. Traffic (TRF) bursts are used for carrying useful data (asynchronous transfer mode (ATM) cells or MPEG2 transport stream (MPEG2-TS)) from the RCST to the gateway(s)/RCST. Each burst consists of the preamble (P) and data (D) except ACQ burst, where ACQ burst contains fixed frequency sequence in place of data. Every burst is surrounded by a guard time which allows for RCST power switch-off transient and system timing errors. Preambles are used for the start of the burst (SOB) detection, symbol timing recovery (STR), carrier frequency recovery (CFR) and carrier phase recovery (CPR). A few generic parameters used in the design are given in Table.1.

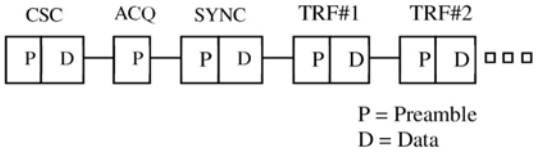


Fig. 1. Burst structure in DVB-RCS systems.

The complex envelope of the received signal sampled at the output of the matched filter is given by

$$\tilde{r}(nT_s) = \sum_k \tilde{s}_k \, g(nT_s - kT - \delta T) \, e^{j(2\pi f_d nT_s + \theta)} + \tilde{v}(nT_s) \quad (1)$$

where $\tilde{s}(t)$ (tilde denotes complex quantities) is the complex envelope of the QPSK signal, *g(t)* is the impulse response of the convolution of the transmitted pulse shaping filter (root- raised cosine, roll-off factor = 0.35) with channel impulse response and matched filter. T is the symbol duration and $T_s = T / 4$ is a sampling time. f_d and θ are the frequency offset and the unknown carrier phase respectively. δ denotes the error due to random timing phase, which is assumed to be uniformly distributed in [0, T). $\tilde{v}(t)$ is the complex-valued white Gaussian noise with two-sided power spectral density $N_0/2$.

Table 1. System model specifications.

Parameter	Value
FEC	Double binary turbo code
Pulse shaping	Root raised cosine, roll-off = 0.35
Traffic burst payload	ATM or optional MPEG2-TS
Modulation	QPSK
Access scheme	MF-TDMA

Fig. 2 shows the block diagram of the demodulator implemented in DSP. A data burst is sent to the digital automatic gain control (AGC) to make each sample to rms level of 2^10, as required by the algorithm. The AGC output is given to the main algorithm, which contains up sampling and matched filtering, synchronization and demodulation. After demodulation, the demodulator sends the soft decision bits to the turbo decoder for decoding.

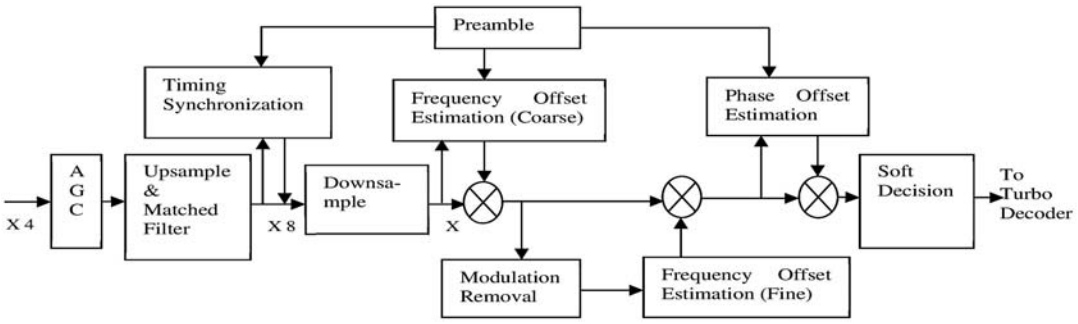


Fig. 2. Demodulator block diagram.

3. Matched Filtering

In real-time practical applications, it is very difficult to design the exact matched filter at the receiver as the receiver does not know the value of δ, the random timing phase. This issue can be resolved by up sampling the output of the digital down converter (DDC) by a factor and then feeding it to the matched filter and the output is sampled at higher rate. Here the symmetric property of the filter is utilized for reducing the computational complexity.

4. Timing Synchronization

A maximum likelihood (ML) method is used for timing offset estimation. Let L_p be the length of the preamble pattern, L_d is the length of data, L_g is the length of guard and $L = L_p + L_d + L_g$. If the preamble pattern starts at the τ-th position, $\tau \in [0, L\text{-}1]$, then ML estimate $\hat{\tau}$ is an integer that maximizes the conditional probability density $f(\bar{\tilde{r}}|\tau)$ of the received signal vector $\bar{\tilde{r}} = (\tilde{r}_0, \tilde{r}_1, \tilde{r}_2, \ldots, \tilde{r}_{L_p+L_d+L_g-1})$. The result of ML with double correlation [7] is given in (2)

$$Sob(n)=\sum_{i=1}^{Lp-1}\left\{\left|\sum_{k=i}^{Lp-1}\tilde{r}^{*}_{n+k}\,\tilde{\beta}_{k}\,\tilde{r}_{n+k-i}\,\tilde{\beta}^{*}_{k-i}\right|-\left|-\sum_{k=n+i}^{n+Lp-1}|\tilde{r}_{k}|\,|\tilde{r}_{k-i}|\right|\right\},\quad 0\le n\le L_{ted}-1 \tag{2}$$

where the first term in the right-hand side is referred to as double correlation, which evaluates a correlation after properly multiplying the received signal with the preamble, the second term is the random data correction and $\tilde{\beta}$ is the QPSK- modulated known preamble symbols of length L_p. L_{ted} is the search length for start of burst finding. The start of burst is estimated using (2) by choosing the value of $\hat{\tau} = n$, which maximizes *Sob(n)*. Since the computational complexity of timing estimation using (2) is more, *i* is restricted to vary from 1 to 2.

The mean square error (MSE) of timing offset estimation versus E_b/N_0 curve with random normalized frequency offset (f_dT) is shown in Fig. 3. The result is measured on 10,000 bursts of different data rates with f_dT that varies randomly in the range of [–0.18, 0.18] and phase variance = 0.3. The figure shows that high estimation accuracy of <10e-6 is achieved through the proposed design even at low E_b/N_0 close to –1 dB. The high estimation accuracy is achieved due to random correction term. Because of the correction term, correlation amplitude calculated using (2) results in a negative value most of the time. So, it is very easy to implement (2) with less computational complexity by applying some threshold. Moreover, a short preamble is used for synchronization, again resulting in less computational time.

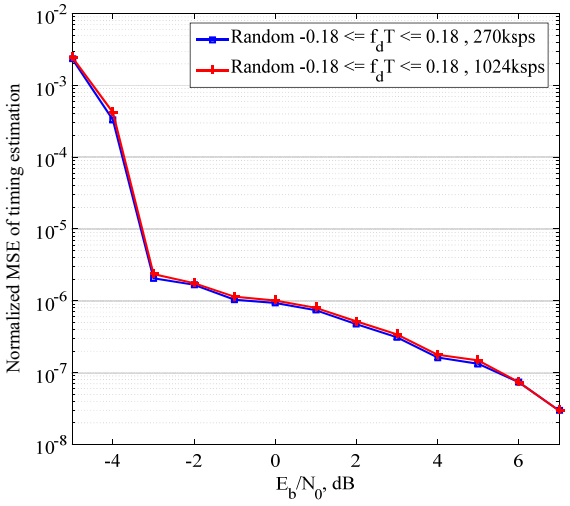


Fig. 3. Normalized MSE of timing estimation versus E_b/N_0 curve with different f_dT and a phase variance = 0.3.

5. Frequency and Phase Offset Estimation

The time-domain frequency estimators proposed in [4-6] exploit the sample correlations

$$R(m)=\frac{1}{L_p-m}\sum_{n=m}^{L_p-1}\tilde{z}_n\,\tilde{z}^{*}_{n-m},\ 1\le m\le N_1, \tag{3}$$

$$\tilde{z}_n=\tilde{r}_{\hat{\tau}+n}\tilde{\beta}^{*}_{n},\ 0\le n\le L_p-1, \tag{4}$$

$$\hat{f}_{dc}=\frac{1}{2\pi T}\sum_{m=1}^{N_1}w(m)\times\Big[\arg\{R(m)\}-\arg\{R(m-1)\}\Big]_{2\pi}, \tag{5}$$

where *w(m)* is a smoothing function given by

$$w(m)=\frac{3\left[(L_p-m)(L_p-m+1)-N_1(L_p-N_1)\right]}{N_1(4N_1^2-6N_1L_p+3L_p^2-1)} \tag{6}$$

The received signal is correlated with the conjugate of $\tilde{\beta}$ from the start of burst ($\hat{\tau}$) onwards to get $\tilde{z}_n$ and $N_1 \le L_p/2$ is the number of correlators, which is a design parameter.

As the DVB-RCS systems require operating at very low SNR and moreover, a short preamble is used for synchronization, it is very difficult to provide complete carrier recovery by only using M&M algorithm [5]. So, in this design M&M algorithm is used for coarse frequency offset f_{dc} estimation. By choosing $N_1 = 12$, it is able to estimate the frequency offset of 20 % of the symbol rate with the expense of some computational time. After correction of the estimated coarse frequency offset $\hat{f}_{dc}$ on the down sampled signal to get $\tilde{r}'_n$ from which, the fine frequency offset f_{df} estimation is calculated. To find f_{df}, which is $f_d - \hat{f}_{dc}$, ML method is used. For the ML method to be more accurate, the modulation in each sample must be removed. The modulation removal method used in this design is $\tilde{s}_n^4 = 1$, so that the received signal contains only unmodulated carrier and noise. The fourth power of $\tilde{r}'_n$ is

$$\tilde{r}'^{4}_{n} = e^{j(8\pi(f_d-\hat{f}_{dc})nT+4\theta)}+\tilde{v}''_n, \tag{7}$$

where $\tilde{v}''_n$ is the noise term resulting from noise interactions. After the coarse frequency offset correction, let the remaining frequency error in the received signal have a range in between $-f_m$ and $+f_m$. For minimizing the computational complexity of fine frequency offset estimation, an iterative method is used. First, the interval (Δf) from $-f_m$ to $+f_m$ is divided into *R* equal parts. So, the ML step size can be given as $f_{step} = (\Delta f / R)$. ML rule for estimating fine frequency offset is given in (8), with f_f varying from $-f_m$ to $+f_m$ with a step size of f_{step}

$$\left|\sum_{n=0}^{L_{ml}-1}\tilde{r}'^{4}_{\hat{\tau}+n}\;e^{-j(2\pi f_f nT)}\right|, \tag{8}$$

where L_{ml} is ML estimation length. One of the R parts, which maximizes (8), is considered as the new interval for ML method. Likewise, the output obtained after

three to four ML iterations is treated as the estimated value of fine frequency offset $\hat{f}_{df}$. The total estimated frequency offset $\hat{f}_{dc} + \hat{f}_{df}$ is corrected on the downsampled signal.

The received signal is then down sampled and corrected by estimated frequency offset. The estimation of carrier phase is performed on the, time-synchronized, frequency-corrected, down sampled received signal. The carrier phase estimate $\hat{\theta}$ is calculated using (9) as

$$\hat{\theta} = \tan^{-1}\left[\sum_{n=0}^{Lp-1} \tilde{r}''_{\hat{\tau}+n}\, \tilde{\beta}^*_n\right], \qquad (9)$$

where $\tilde{r}''_n$ is the frequency-corrected, time-synchronized, down sampled received signal. The received signal is demodulated after phase offset correction, and hard and soft decision outputs are determined. After the demodulation, the soft-decision outputs are transfers to the turbo decoder for decoding.

Fig. 4 shows normalized MSE of frequency offset estimation versus E_b/N_0 curve for different data rates. The result is measured on 10,000 bursts using f_dT which varies randomly in the range of [–0.18, 0.18] and phase variance = 0.3. The figure shows that high estimation accuracy <10e-6 is achieved through the proposed design even at low E_b/N_0 close to –1 dB.

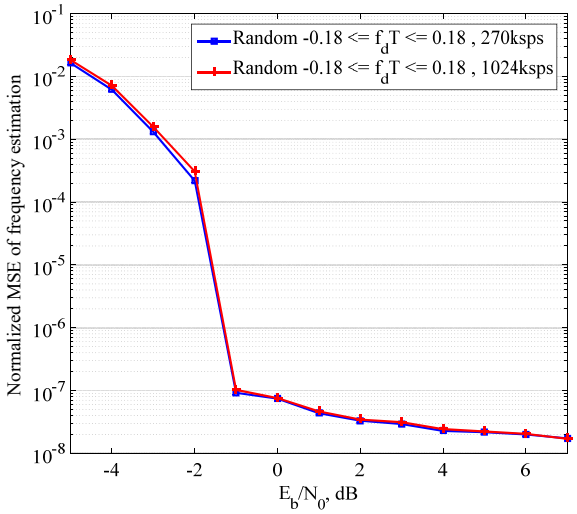


Fig. 4. Normalized MSE of frequency estimation versus E_b/N_0 curve after timing recovery with a phase variance = 0.3.

6. Hardware Architecture and DSP-based Implementation of Algorithm

The prototype hardware module of the receiver, as shown in Fig. 5, is having one Xilinx Kintex FPGA, two TI multi-core TMS320C6678 DSPs and Vitesse Ethernet switch. The C66x multi-core DSP contains eight cores and each core runs at 1.25 GHz. The multi-core shared memory controller (MSMC) provides a shared on-chip SRAM that is accessible by all the eight C66x cores. The two DSPs support two Ethernet media access controllers (EMACs) each and are connected to the Ethernet switch to support the user datagram protocol (UDP) packet transmission. The Kintex FPGA interfaces with DSPs using Serial RapidIO (SRIO) and external memory interface (EMIF16). Kintex FPGA performs DDC and FEC decoding whereas DSP performs synchronization and demodulation.

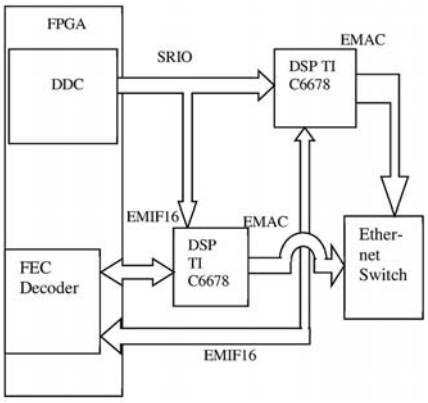


Fig. 5. Prototype hardware module architecture.

Flow chart for the demodulator implementation in DSP is based on multiple threads as shown in Fig. 6. Multiple threads are used to process various signal processing units of the demodulator, where core 0 process packet processing, core1 process interrupt handler and data buffering, core 2 to 6 process synchronization and demodulation and core 7 process send/receive data to/from FEC decoder threads.

FPGA stores I and Q samples of one burst of all the channels after DDC and send an interrupt to core 1 of DSP. Core 1 receives all the bursts of multiple channels from FPGA through the SRIO interface and transfers the data burst of the respective channels to the other cores (core 2 to 6, where each core processes single channel) for synchronization and demodulation through MSMC memory. After the demodulation, each core will transfer the demodulated data to core 7 if the burst correlation level (length of the preamble) condition is satisfied. If the correlation level condition is not met, data will not be transferred to core 7 so that processing cycles will be saved. After receiving demodulated data at core 7, it transfers the data to FPGA for decoding over dedicated EMIF16 interface. After the completion of the decoding process, FPGA transfers the data back to core 7. After reception of the decoded data, core 7 transfers the data to core 0 through MSMC memory. Core 0 packetizes the decoded data and transmits it to the network through Ethernet switch.

7. Prototype Test Setup

Fig. 7 shows the prototype test setup, which is used for verification, development, and testing of the DVB-RCS demodulator. Test vector files are

generated for multiple channels with different data rates in MATLAB and are loaded to Keysight vector signal generator (VSG). The combiner's functionality is to combine the signal from VSG and AWGN from Noise generator and it generates noise added signal at 70 MHz IF. The prototype hardware platform receives the noise-added 70 MHz multichannel signal. The output of the prototype hardware platform is demodulated and decoded packets can be captured in PC (Wireshark).

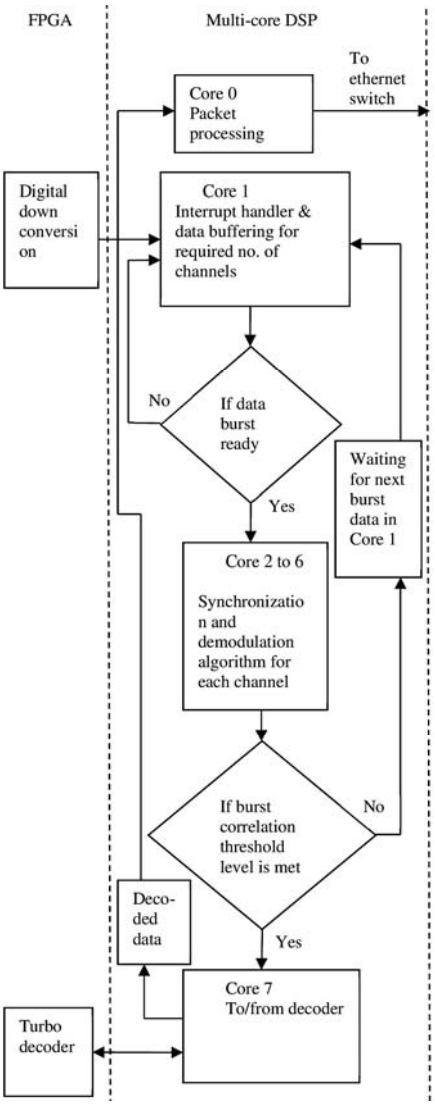


Fig. 6. Flow chart of demodulator implementation.

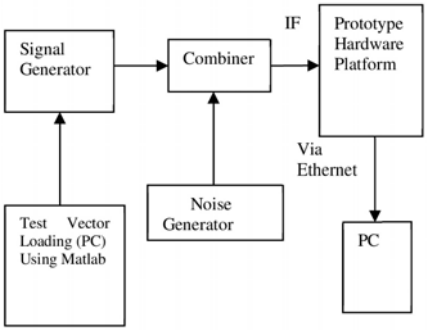


Fig. 7. DVB-RCS receiver test setup.

Fig. 8 depicts bit error rate (BER) performance for an uncoded traffic burst which contains ATM or MPEG2-TS, and is evaluated at 270 ksps and 1024 ksps symbol rates. A normalized carrier frequency offset, which randomly varies per burst in the range of [–0.18, 0.18], is considered. BER of the proposed demodulator is very close to theoretical performance.

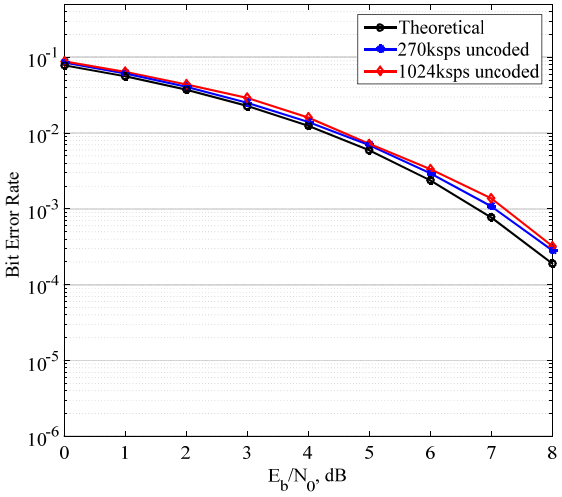


Fig. 8. BER performance (uncoded).

Fig. 9 shows the performance of the DVB-RCS turbo coding. BER curve vs. E_b/N_0 using AWGN channel is influenced by frame size in couples and code rate. Two frame sizes of length 212 and 752 couples are considered for performance evaluation. The Max-Log-MAP decoding algorithm [9] with four iterations is used for turbo decoding.

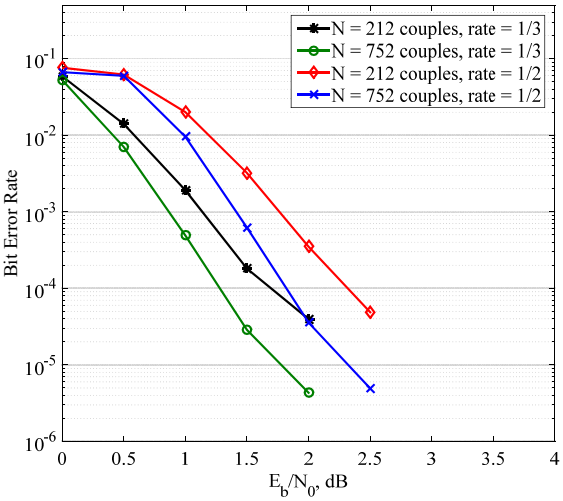


Fig. 9. BER performance with turbo codes.

8. Conclusions

DVB-RCS standard MF-TDMA demodulator for satellite receiver is implemented in multi-core DSP. The implemented receiver uses a multi-channel burst

mode demodulator algorithm, which successfully estimates all synchronization parameters even at low E_b/N_0 close to 0 dB. The proposed multi-channel demodulator design effectively utilizes the eight parallel cores of the DSP to compute different signal processing stages. The proposed design efficiently demodulates 10 MF-TDMA channels in parallel.

References

[1]. ETSI EN 301 790, Digital Video Broadcasting (DVB); Interaction Channel for Satellite Distribution Systems, v1.5.1, *ETSI*, May 2009.

[2]. ETSI TR 101 790, Digital Video Broadcasting (DVB); Interaction Channel for Satellite Distribution Systems, Guidelines for the Use of EN 301 790, v1.4.1, *ETSI*, July 2009.

[3]. ETSI EN 302 307, Digital Video Broadcasting (DVB); Second Generation Framing Structure, Channel Coding and Modulation Systems for Broadcasting, Interactive Services, News Gathering and Other Broadband Satellite Applications (DVB-S2), v1.2.1, *ETSI*, August 2009.

[4]. M. Luise, R. Reggiannini, Carrier frequency recovery in all-digital modems for burst mode transmissions, *IEEE Transaction on Communications*, Vol. 43, Issues 2-4, 1995, pp. 1169-1178.

[5]. U. Mengali, M. Morelli, Data-aided frequency estimation for burst digital transmission, *IEEE Transaction on Communications*, Vol. 45, Issue 1, January 1997, pp. 23-25.

[6]. U. Mengali, A. N. D'Adrea, Synchronization Techniques for Digital Receivers*, Plenum Press,* 1997.

[7]. Z. Y. Choi, Y. H. Lee, Frame synchronization in the presence of frequency offset, *IEEE Transaction on Communications*, Vol. 50, Issue 7, July 2002, pp. 1062-1065

[8]. J. Li, J. Kim, Data-aided synchronization for MF-TDMA multi-carrier demultiplexer/demodulator (MCDD), *IEEE Transaction on Broadcasting*, Vol. 55, Issue 3, September 2009, pp. 623-632.

[9]. P. Robertson, P. Hoeher, E. Villebrun, Optimal and suboptimal maximum a Posteriori algorithm suitable for turbo decoding, *European Transaction on Telecommunications*, Vol. 8, Issue 2, 1997, pp. 119-125.

(1901)

From Neutral Human Face to Persuasive Virtual Face, a New Automatic Tool to Generate a Persuasive Attitude

Afef Cherni [1,2], Roxane Bertrand [2] and Magalie Ochs [1]
[1] Aix Marseille Univ., Universite de Toulon, CNRS, LIS, Marseille, France´
[2] Aix Marseille Univ., CNRS, LPL, Aix En Provence, France

Summary: In order to motivate the user to change her/his behavior or attitudes, for instance to practice physical activities to improve her/his well-being, virtual agents should have persuasive capabilities. The persuasiveness of the virtual agent not only depends on its speech but also on its non-verbal behavioral cues. In this paper, we propose the new tool called THRUST (from neuTral Human face to peRsUaSive virTual face), to automatically generate the head movements and facial expressions of a persuasive virtual character from a video of a human. Combining a machine learning approach on a corpus of persuasive human speech and a convolution-based method, we propose a model, based on real data of persuasive human message that transforms the non-verbal behavior of the human expressed in a video to a persuasive non-verbal behavior replicated on a virtual face.

Keywords: Multimodal cues, Persuasion, Embodied conversational agent, Machine learning methods, Convolution.

1. Introduction

A key challenge in intelligent virtual agent research concerns the automatic generation of the embodied conversational agent's behaviors, in particular related to social and emotional dimensions. In this article, we focus more particularly on the social skills of *persuasion*. Persuasion can be defined as "*any message that is intended to shape, reinforce or change the responses of another or others*" [1]. As highlighted in [2, 3], the persuasiveness of a message does not only depend on its content but crucially depends on all the multimodal components involving the different verbal, vocal and mimo-gestural levels (facial expressions, gestures, pitch). In this article, we particularly focus on the social signals expressed by non-verbal cues (facial expressions, head movements), that a virtual agent could express to be more persuasive. We do not consider the argumentative aspects (e.g., the identification of the arguments to convince, the order of the presentation of the arguments, the responses to the arguments of the persuade). As a first step, we concentrate our study on *non-verbal cues* of persuasion.

The final goal of our project is to develop a *persuasive Embodied Conversational Agent* (ECA) to motivate elderly population to practice physical activity. In the intelligent virtual agent domain, several persuasive virtual agents have already been developed (e.g. [4-6]). The main approach to model persuasive behavior consists in identifying in the literature the behavioral cues that have an impact on persuasiveness and to integrate these cues in artificial agents. Indeed, the literature highlights some human's behavioral cues related to persuasion, as for instance the body movements [2] or the prosody [5]. In the domain of virtual agents, some empirical research works have shown the importance of certain verbal and non-verbal cues to improve the virtual agent's persuasiveness [7, 3]. However, as far as we know, no multimodal behavioral model to determine the verbal and non-verbal cues that a persuasive virtual agent should express during an interaction with a user, has been yet proposed.

In this article, we propose a new tool called *THRUST: from neuTral Human face to peRsUaSive virTual face*. This tool automatically transforms the video of a human to a video of a virtual character with a persuasive non-verbal behavior. More precisely, the tool extracts automatically the head movements and the facial expressions of the human, modifies them based on a computational model that we proposed, and replay the computed head and facial movements on a virtual face. The main contribution of the paper concerns the computational model that automatically modified the head and facial movements extracted from the human face to persuasive movements replicated on the virtual face.

One first challenge that has to be tackled to create such model is to identify more precisely the cues related to persuasion; i.e., the cues that we will modify to simulate persuasive behavior. For this purpose, in a first step, we propose to explore the relevant behavioral cues of persuasion using machine learning methods applied on a corpus of human videos. We consider the POM corpus [8] which is the only multimedia corpus created with annotations for studying persuasiveness to our knowledge. It contains web videos of speakers talking about different subjects in front of a camera.

Based on machine learning classifiers, we have explored the behavioral cues of persuasion. In this work, special attention has been paid to create explainable models with interpretable features (i.e., features that we can understand contrary to raw data). Indeed, our objective in this first step is not to create a classification model to recognize persuasiveness, but to identify the relevant features that we have to

consider to create persuasive behavior on a virtual character.

The second step to create the tool is to transform the relevant signals extracted from the human face to persuasive one. Since the POM corpus contains the real measurements of different non-verbal cues of human expression with neutral and persuasive attitude, we create a dictionary from these real data to define the references reflecting persuasive non-verbal behavior. Based on this dictionary, a convolution-based method has been applied. Note that in this work, we define the neutral attitude as to speak without being persuasive or without making the effort to be so.

This article is organized as follows. In the next section, we present related works, i.e., the theoretical and empirical research works exploring the behavioral cues related to persuasion. We then introduce an overview of our proposed architecture to create persuasive behavior in Section 3. We detail the machine learning framework in Section 4. Then, Section 5 is dedicated to the convolution-based model that we propose based on the POM corpus. We present the implementation of our new automatic tool to generate a persuasive attitude in Section 6. We conclude in Section 7 by presenting the limits of the study and by discussing future works.

2. Related Work

The study of the persuasiveness of specific behavioral cues has been the main interest of various works, especially in the context of human-human interaction. The research works show the importance of several multimodal behavioral [2, 5, 9]. In this article, we focus on the non-verbal cues of persuasiveness. [2] proved that gestural and facial activity (e.g., gestures, body movements, facial expressions and smiles) improve the persuasion. At the interactional level, several works studied the positive impact of mimicry on persuasion [10]. In this article, we analyse corpora of monologue excluding the possibility of studying the interactional level. Other contextual elements, such as the appearance of the persuader [2], may impact the perceived persuasion. In this article, given the size of the considered corpus and the lack of contextual variability, as a first step, we do not consider the influence of the context. Based on the research showing the importance of face and head movements for persuasion [2], we consider in our study the facial expressions through the study of action units and the head movements. These behavioral cues considered as features of the learned models are presented in more details in Section 4.3.

In the Intelligent Virtual Agent domain, to generate automatically the behavior of a virtual agent, two main approaches are identified in the literature. The first approach relies on rule-based systems that exploit linguistic information from the text and the meaning of gestures, facial expressions or head movements to determine the appropriate signals to express (e.g., [11, 12]). Rule-based approaches remain very limited, given the variability of human expressions across modalities. In a much more recent approach, machine learning methods are used to automatically generate co-verbal gestures (e.g. [13]), facial expressions and body movements from speech (e.g. [14]) or from speech and text to take into account both acoustic and semantic information (e.g. [15, 16]). Most studies are based on deep neural networks (e.g. [13, 17, 16]) and, more recently, on the use of GAN architectures (e.g. [15, 14]). Our presented work differs from the existing models on different aspects: (1) we generate non-verbal behavior, not from speech or text, but from a video of a human with a neutral attitude; (2) we generate the facial and head movements whereas most of the existing models consider the body and head movements and (3) we explore the automatic generation of *persuasive* behavior whereas a limitation of the existing work is that the proposed models do not allow for the generation of different social-emotional behaviours.

From a *machine leaning perspective*, few research works have investigated persuasion. The main work has been conducted by Park *et al.* [8, 18, 19] on the Persuasive Opinion Multimedia (POM) corpus consisting of 1000 movie review videos obtained from a social multimedia website called ExpoTV.com. As proposed by Park *et al.* [8, 18, 19], we use machine learning algorithms to explore persuasiveness. However, our work differs from the latter in several aspects:

- Contrary to Park *et al.*, in order to obtain *explainable models*, we do not use deep learning methods but "white box" classifiers such as SVM and Random Forest;
- Still in our perspective of interpretability, we consider nonverbal features that can be simulated on a virtual character[1];
- Last but not least, our final objective is not to create a prediction model but to explore the non-verbal cues and use machine learning-based methods in order to create a persuasive artificial agent.

3. General Architecture

In this section, we present the general architecture of our system illustrated on the Fig. 1.

Input. The system takes as input a video of a speaker talking about a specific topic in a neutral way. The input is not limited to real-time data video, it can be webcam video, recorded video files or sequences of images. The important aspect is to be able to extract the facial landmarks, the head poses, the eye gaze and the facial Action Unit (AUs) from the video. For this

[1] https://github.com/isir/greta/wiki

purpose, we use the OpenFace tool[1]. We note these N measures as $(U_i)_{i=1...N}$ where we index them by i from 1 to N and we design by each vector U the measured feature (for example AU1, AU2, AU12, head position according to x, y or z−axis, ...) and i its index in the variables set. In other words, the vector $(U_i)_{i=1...N}$ represents the values of the features characterizing the head and face movements extracted from the video of the human. These variables set $(U_i)_{i=1...N}$ represents the input of the "Model" box.

Model: As illustrated in Fig. 1, the proposed model takes as input the set of features $(U_i)_{i=1...N}$ characterizing the face and head movements of a neutral human speaker and produces as output a set of features $(W_i)_{i=1...N}$ characterizing the head and face movements of a persuasive speaker. To compute how to modify the set of features to be persuasive, we combined a machine learning approach and a convolution-based method. The first steps consist in using machine learning methods on an existing corpus to identify the important relevant features to consider to simulate persuasiveness (Step 1, 2, and 3, Fig. 1). These steps are detailed in Section 4. Note that in these steps, we learn a classification model to automatically determine if the face and head movements are persuasive or not. This classification model is also used to verify that the transformed vector of characteristics $(W_i)_{i=1...N}$ is indeed considered as persuasive (illustrated on the Fig. 1 by the dotted arrow from the "output" box to the "classification" box). The second steps consist in modeling a convolution-based method, from the data of the corpus, in order to determine how to modify the features to be persuasive (Step 4 and 5, Fig. 1). These steps are detailed in Section 5.

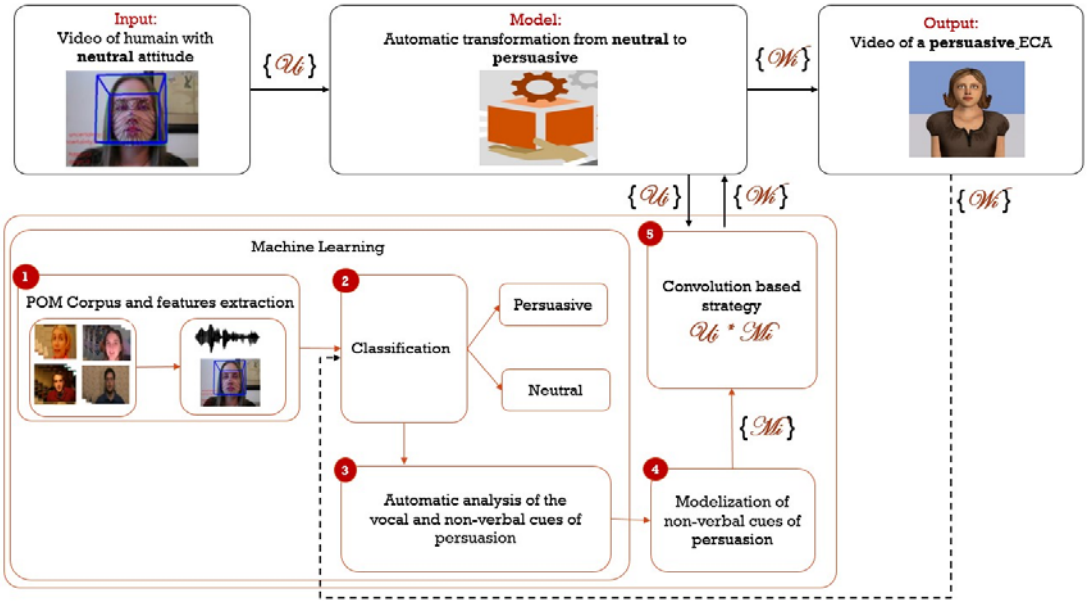


Fig. 1. Global architecture of the system to automatically transform a neutral human video to a persuasive virtual character video. *Input*: a video of a human with automatic extraction of head and facial movements using OpenFace. *Model*: a computational model to automatically transform neutral non-verbal features to persuasive non-verbal features *Output*: a video of a virtual character replicating the behavior of the human but with persuasive head and face movements.

Output. As output, the system produces a video of a persuasive virtual character. The virtual character replicates the same speech but with persuasive face and head movements. For this purpose, the "Output" box simulates the variables set $(W_i)_{i=1...N}$ on the embodied conversational agent Greta, and generates the target video. The vector $(W_i)_{i=1...N}$ represents the value of the head and face movements extracted from the video of the human and transformed to be persuasive.

4. Machine Learning Framework

In this section, we describe the steps 1, 2 and 3 illustrated on the Fig. 1.

4.1. Corpus and Features Extraction

In the step 1 (Fig. 1), we consider a specific corpus and extract the features form the video of the corpus. Concerning the choice of the corpus, nowadays, few corpora in the research community are available to study persuasiveness. In this work, we consider the Persuasive Opinion Multimedia (POM) corpus [8]. This corpus is freely available and contains videos of speakers trying to convince on different subjects. POM corpus consisting of 1000 movie review videos obtained from a social multimedia website called ExpoTV.com. It contains different conversational videos cut into a total of 1096 thin slices. Each cut was annotated by different native English-speaking

[1] https://www.cl.cam.ac.uk/research/rainbow/projects/openface/

workers of the United States. The research works conducted by Park *et al.* on this corpus [8, 19], show that behavioral cues can be used to predict *extreme value* of persuasion. This work has been conducted to automatically classify the persuasiveness of a human speech. In this article, we aim at exploring the corpus to identify the relevant features of persuasiveness, i.e., the non-verbal cues that enable a human to be perceived as persuasive. Then, we propose to determine the set of relevant features to consider to simulate persuasive message. For this purpose, we explore different sets of features and their impact on the performances of classifiers to evaluate their importance.

Based on the theoretical and empirical research on persuasion presented above (Section 2), we consider the following groups of features:

- *Group 1: facial expressions*: facial action units (AU1, AU2, AU4, AU5, AU6, AU7, AU9, AU10, AU12, AU14, AU15, AU17, AU18, AU20, AU23, AU24, AU25, AU16, AU28, AU43);
- *Group 2: emotions*: Anger, Contempt, Disgust, Joy, Fear, Surprise, Confusion, Frustration. The emotions are computed based on vocal and non-verbal cues [18];
- *Group 3: head movements*: head position (displacement and rotation in (*x, y, z*) axes, speed of the head movement and its acceleration according to (*x, y, z*) axis;
- *Group 4: acoustic descriptors*: fundamental frequency f_0, peak slope.

In order to explore the importance of these features, we propose to compute statistical functions for each feature: the mean, median, maximum, minimum, standard deviation and the variance. Note that even if we do not consider the acoustic features for the simulation on the virtual character, at this step, we consider relevant to evaluate the importance of these acoustic features in comparison to non-verbal ones.

4.2. Formalization of the Classification Problem

Our objective by using machine learning methods is to investigate the behavioral cues related to persuasion. As illustrated by the step 2 (Fig. 1), to identify the importance of the features in the perception of the persuasion, we consider a classification task: based on the features as input, the classifiers have to predict if the features are persuasive as output.

Different classification methods could be considered (e.g., binary classification, multi-class classification, or regression). As a first step, we consider a binary classification to simplify the learning problem (i.e., prediction if persuasive or not). The accuracy of the prediction may depend on the chosen definition of the output classes. Indeed, we can choose to predict only extreme values (as proposed in [8]) or to split in two balanced classes without excluding middle values. In order to compare the different approaches, we propose to explore 2 strategies to define the clustering of the two classes:

Strategy 1: This strategy considers the extreme values of the annotated persuasion to create two classes as proposed in [8]: one class clusters the values equal to or greater than 5.5 and the other class clusters the values equal to or less than 2.5 (for values ranging from 1 to 7);

Strategy 2: With the above *Strategy 1*, the classes are imbalanced. Therefore, we propose in *Strategy 2* to explore oversampling methods to increase the amount of data and to obtain balanced classes. The over-sampling methods generate new samples of the minority class based on the existing dataset, in order to remove class imbalance. For this purpose, we propose to use SMOTE (Synthetic Minority Over-sampling Technique) algorithm [20].

4.3. Automatic Analysis of the Vocal and Nonverbal Cues of Persuasion

In the step 3 (Fig. 1), the objective is to test the clustering strategies proposed in Section 4.2 to compare the performances of the classifiers and then to select the most important features that ensure the highest prediction performances. We consider at this step both vocal and non-verbal cues.

Classifiers: We propose to experiment different classifiers: the *Naives Bayes* (NB), the *System Vector Machine* (SVM) and the *Random Forest* (RF). These methods, among the best classifiers [21], have the advantage, compared with other statistical models such as RNN, to handle high-dimensional data with a high generalization power [22, 23]. They are also well suited for handling small datasets. All experiments were performed with 10-fold cross-validation (CV) where each CV was tested 10 times.

Baselines: In order to estimate the performances of the different classifiers, we compute scores from classifiers returning random predictions, to establish *baselines*. We consider three different strategies: uniform (generates predictions uniformly at random) (noted BR), stratified (generates predictions with respect to the training set's class distribution) (noted BU0) and most frequent (always predicts the most frequent class in the training set) (noted BU1). For each fold of the cross-validation, the random classifiers are fitted on the training set and used to generate predictions on the validation set, for each strategy.

Prediction model: Each classifier will be fitted on the training set (80 %) and testing set on 20 % of the corpus. This experiment was performed with 10-fold cross-validation (CV) where each CV was tested 10 times. The performances of the classifiers are evaluated though the classical metrics of accuracy and F1 weighted score (to cope with the unbalanced classes for the *Strategy 1*). Table 1 summarizes the performances of the different classifiers. We moreover compute the statistically significant differences of the obtained F-scores. The *Student's t-test* is performed to

compute the statistical differences between the F1-scores of the classifiers and of the baselines obtained by the k-fold-cross-validation. This test is one of the recommended methods to compare the performance of machine learning algorithms [24].

Prediction results: In order to evaluate the importance of each group of features (facial expressions, emotions, head movements and acoustic descriptors) to predict the persuasion, we compute the performance scores of the classifiers considering each group of features as input and combinations of several groups of features. Table 1 summarizes the performances of the best classifiers and the significant differences with the baselines considering the different groups of features as input (the scores significantly different from the average scores of the 3 baselines are presented with gray cells on the table).

Table 1. Performance scores of different classifiers (*Support Vector Machine* (SVM), *Random Forest* (RF), *Naive Bayes* (NB)) using two strategies. We design by *Group 1*: facial expressions, *Group 2*: emotions, *Group 3*: head movements, *Group 4*: acoustic descriptors. The highest scores are written in bold and the scores significantly different from the average scores of the 3 baselines are presented with gray cells.

Considered features	Classifiers	***Strategy 1***		***Strategy 2***	
		Accuracy score	F1 weighted score	Accuracy score	F1 weighted score
Group 1	SVM	0.64	0.66	0.64	0.73
	RF	**0.71**	**0.74**	0.67	0.67
	NB	0.66	0.63	0.66	0.64
Group 2	SVM	0.54	0.71	0.54	0.96
	RF	0.55	0.58	0.55	0.53
	NB	0.54	0.55	0.54	0.56
Group 3	SVM	0.63	0.65	0.63	0.62
	RF	0.72	0.074	0.72	0.71
	NB	0.57	0.55	0.57	0.45
Group 4	SVM	0.61	0.65	0.61	0.65
	RF	0.69	0.51	0.69	0.51
	NB	0.72	0.62	0.72	0.62
Group 1 + 3	SVM	0.64	0.69	0.64	0.67
	RF	0.69	0.83	**0.74**	**0.82**
	NB	0.64	0.55	0.64	0.50
Group 1 + 4	SVM	0.45	0.55	0.45	0.56
	RF	0.52	0.70	0.54	0.68
	NB	0.63	0.68	0.63	0.68
Group 3 + 4	SVM	0.71	0.72	0.71	0.71
	RF	0.74	0.76	0.74	0.78
	NB	0.54	0.57	0.59	0.65
Group 1 + 3 + 4	SVM	0.65	0.73	0.65	0.73
	RF	0.76	0.74	**0.81**	**0.72**
	NB	0.64	0.56	0.64	0.56
Group 1 + 2 + 3 + 4	SVM	0.63	0.74	0.63	0.74
	RF	0.76	0.76	0.77	0.68
	NB	0.64	0.53	0.64	0.56

The results (Table 1) show that the emotions (*Group 2*) do not enable us to obtain significant differences with the baselines. In others words, the emotions are not sufficient to predict persuasion. In the same way, the group of features containing only head movements (*Group 3*) or only acoustic features (*Group 4*) lead to performances not significantly different from the baselines. However, the facial expressions (features of the *Group 1*) provide good performance scores with significant differences with the baselines.

Considering combinations of groups of features, the result reveals than the combination of non-verbal and vocal cues improves significantly the accuracy score. These results are in line with the research on persuasion showing the importance of multimodality for perceived persuasion. Finally, the best accuracy score is obtained by combining facial expressions features, head movements and vocal features.

4.4. Modelization of the Non-verbal Cues of Persuasion

In the previous section, we have considered both verbal and non-verbal features. The results show that the set of features with the facial expressions (AUs), the head movements and the vocal features ensures efficiently the persuasion prediction with the highest accuracy score (0.81). However, we obtain similar results considering only facial expressions features (*Group 1*) (accuracy score equal to 0.71 and F1 weighted score equal to 0.74) or the combination of facial expressions features and head movement (*Group 1 + 3*) (accuracy score equal to 0.74 and F1 weighted score equal to 0.82) (Table 1)).

In this step, we focus on the non-verbal cues that can be simulated on the embodied conversational agent Greta: AU1, AU2, AU4, AU5, AU6, AU7, AU9,

AU10, AU12, AU14, AU15, AU20, AU23, AU25, AU26, AU45. Note that we do not consider the AUs related to lips movements. We plan to use a specific tool to ensure the lips synchronisation on the speech because the AUs are not sufficient to synchronize speech and lips movements. Moreover, since the POM corpus is based on different videos in which the speakers give their point of views or feelings about a particular subject, we propose to avoid the disgust expressions (described in our case by the facial expression AU9 and AU10). Consequently, in order to ensure the transformation from neutral to persuasive non-verbal movements, as a first step, we consider only face and head movements by modifying the following features AU1, AU2, AU4, AU5, AU6, AU7, AU12 (in *Group 1*) and head movement according to (*x, y, z*)-axis (in *Group 3*). Note that we have run our prediction model presented in Section 4.3 with these considered features and we have obtained a good performance score with an accuracy score equal to 0.63 and F1 weighted score equal to 0.73.

5. Convolution-based Model

In the steps 4 and 5 (Fig. 1), we define how to transform the non-verbal features (considered as important for persuasion in the previous steps) to increase the persuasiveness of a virtual speaker. These steps are based on the data of the POM corpus. Indeed, for each important non-verbal selected feature, we propose to generate a signal that describes on average its dynamic according to all the sequences of POM corpus classified as persuasive or neutral. More precisely, since each video in POM corpus is annotated with respect to thin slice method [25], we consider each slice as a sample and the average value of each non-verbal behavior dynamic as a reference. After considering all the slices, we obtain a signal that describes the average values that non-verbal cue takes with respect to a persuasive attitude. This makes a reference for each non-verbal cue (i.e., feature) to follow in order to build a persuasive attitude. These references will be noted $(M_i)_{i=1,..,N}$. We remind that N is the number of considered features, i its index and M presents the reference of each feature (expression facial units (AU1, AU2, ...), head movements).

Concerning the convolution-based strategy, at this step, we have two types of data: (1) the features set of the neutral human video noted U_i, (2) the reference of each non-verbal cues M_i build in the previous step based on the POM corpus. In order to adapt the dynamic of the values U_i and make it persuasive, we apply a convolution product between U_i and M_i. This step can be considered as an average filtering where the U_i input follows the specific properties of the M_i function. Since the size of each M_i depends on the used corpus, an inadequacy with the input size U_i may occur. To avoid this problem, a re-sampling treatment of each M_i according to the size of its corresponding input U_i is highly recommended at this step. Moreover, we propose to apply the convolution product to the i-th variables (U_i and M_i) with respect to windows with size w to keep the same level of reference evolution and avoid the outliers. The results are noted W_i. We remind that our convolution-based strategy proposed in this paper is applied only on the non-verbal cues that Greta takes into consideration (head movement according (*x, y, z*)-axis, and specific AUs).

6. Implementation and Evaluation

The entire process of our proposed tool illustrated in Fig. 1 has been implemented. The tool is called THRUST: from neuTral Human face to peRsUaSive virTual face. The entire code of the tool was provided in open source on GitHub[1]. We present examples of outputs of the system in Fig. 2 considering two different embodied conversational agents of the Greta platform: one female and one male. The resulting videos show that the proposed tool can be used on virtual characters of various appearances. In the Fig. 2, we compare (1) the videos replicating directly the features extracted from the human video on an ECA, i.e., without any transformation and (2) the videos after the transformation of the model to create persuasive non-verbal behavior. The videos are available in THRUST channel[2].

The resulting videos show a significant difference of the face and head movements of the two videos. While the movements in the video without transformation is quite stable, in the video after the transformation, we can notice eyebrow movements and smiles. In order to evaluate these results, we first propose an objective method based on the persuasion classifier. Indeed, since we have built an efficient classifier that ensures the persuasion prediction (Section 4.2, Table 1), we propose to use it to test if the output of the model is correctly classified. We use the best identified classifier (*Random Forest*). To evaluate the model objectively, we have generated 50 videos (25 recorded human faces, replayed on virtual characters of the same gender and transformed to 25 videos of persuasive virtual faces). These videos correspond to 5 different speeches, each speech produced by 5 different participants (3 female and 2 male) and lasts around 10 seconds. In total, we have generated 25 videos of *neutral* virtual faces (replay of the recorded human features on a virtual face) and 25 videos of *persuasive* virtual faces (output of the model). Using the classifier, the results show that all the videos transformed by our model are classified as persuasive whereas those before the transformation are classified as neutral. This first objective evaluation constitutes a first validation step of the proposed tool.

[1] https://github.com/CherniAfef/THRUST-Tool

[2] https://www.youtube.com/channel/UC87g8UeHbMJync8n8DjLe8g/videos

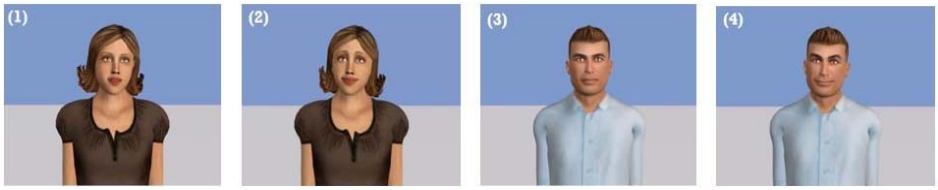


Fig. 2. Screenshots at the same time of two studied cases: (1) and (2) played with the virtual female character "Emma" of Greta, (3) and (4) played with the virtual male character "John" of Greta. (1) and (3) present the input data (neutral attitude), (2) and (4) present the output of our model (persuasive attitude).

The next step is to conduct a subjective evaluation of the generated videos. In a perceptive study, we plan to ask participants to evaluate the believability and naturalness of the generated behavior (as proposed in [26]) but also the perceived persuasiveness of the virtual characters: both from the video before and after transformation. Different questionnaires would be considered to evaluate the level of persuasiveness such as Godspeed questionnaire [27] with some modifications in order to adapt it to persuasiveness context. Note that the video will be played without sound to avoid the lip synchronization problem and the effects of the speech on the perception.

7. Conclusion and Perspectives

In this article, we have proposed a new tool, called THRUST (from neuTral Human face to peRsUaSive virTual face), to automatically transform a video of a neutral human face to a video of a virtual character face expressing persuasive head and face movements. To create such a tool, we have based our work on real videos of human with different levels of persuasiveness. Combining a machine learning approach and a convolution-based model, the proposed tool modifies automatically the relevant features of the face expressions and head movements to increase the persuasiveness of the behavior. The objective evaluation of the resulting video shows that the video generated by the tool as in fact automatically classified as persuasive.

The presented work presents some limits. We have limited the considered features according to the available corpus (POM corpus, the only corpus that contains persuasion annotations) and given the freely accessible toolbox (Greta and OpenFace used in our work). In a second step, we aim at extending the study to other multimodal features and in particular to vocal ones in order to improve the persuasive model and build an automatic artificial agent able to speak and express itself persuasively.

References

[1]. G. R. Miller, On Being Persuaded: Some Basic Distinctions, *Sage Publications, Inc.*, 2013.

[2]. J. K. Burgoon, T. Birk, M. Pfau, Nonverbal behaviors, persuasion, and credibility, *Human Communication Research*, Vol. 17, Issue 1, 1990, pp. 140-169.

[3]. V. Chidambaram, Y.-H. Chiang, B. Mutlu, Designing persuasive robots: how robots might persuade people using vocal and nonverbal cues, in *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI'12)*, 2012, pp. 293-300.

[4]. C. Lisetti, R. Amini, U. Yasavur, N. Rishe, I can help you change! An empathic virtual agent delivers behavior change health interventions, *ACM Transactions on Management Information Systems (TMIS'13)*, Vol. 4, Issue 4, 2013, pp. 1-28.

[5]. V. Petukhova, M. Raju, H. Bunt, Multimodal markers of persuasive speech: Designing a virtual debate coach, in *Proceedings of the 18th Annual Conference of the International Speech Communication Association (INTERSPEECH'17)*, 2017, pp. 142-146.

[6]. H. Nguyen, J. Masthoff, P. Edwards, Persuasive effects of embodied conversational agent teams, in *Proceedings of the International Conference on Human Computer Interaction (HCI'07)*, 2007, pp. 176-185.

[7]. A. S. Ghazali, J. Ham, E. I. Barakova, P. Markopoulos, Poker face influence: persuasive robot with minimal social cues triggers less psychological reactance, in *Proceedings of the 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN'18)*, 2018, pp. 940-946.

[8]. S. Park, H. S. Shim, M. Chatterjee, K. Sagae, L.-P. Morency, Computational analysis of persuasiveness in social multimedia: A novel dataset and multimodal prediction approach, in *Proceedings of the 16th International Conference on Multimodal Interaction (ICMI'14)*, 2014, pp. 50-57.

[9]. N. Miller, G. Maruyama, R. J. Beaber, K. Valone, Speed of speech and persuasion, *Journal of Personality and Social Psychology*, Vol. 34, Issue 4, 1976, pp. 615-624.

[10]. R. Tanner, T. Chartrand, The Convincing Chameleon: The Impact of Mimicry on Persuasion, *ACR North American Advances*, 2006.

[11]. J. Cassell, H. Vilhjilmsson, T. Bickmore, BEAT: The Behavior Expression Animation Toolkit, in *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH-Oi'01)*, 2001, pp. 477-486.

[12]. S. Marsella, Y. Xu, M. Lhommet, A. Feng, S. Scherer, A. Shapiro, Virtual character performance from speech, in *Proceedings of the 12th ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, 2013, pp. 25-35.

[13]. C.-C. Chiu, S. Marsella, Gesture generation with low-dimensional embeddings, in *Proceedings of the International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS'14)*, 2014, pp. 781-788.

[14]. I. Habibie, W. Xu, D. Mehta, L. Liu, H.-P. Seidel, G. Pons-Moll, M. Elgharib, C. Theobalt, Learning speech-driven 3D conversational gestures from video, in *Proceedings of the 21st ACM International Conference on Intelligent Virtual Agents (IVA'21)*, 2021, pp. 101-108.

[15]. C. Ahuja, D. W. Lee, R. Ishii, L.-P. Morency, No gestures left behind: Learning relationships between spoken language and freeform gestures, in Findings of the Association for Computational Linguistics: EMNLP 2020, *Association for Computational Linguistics*, 2020, pp. 1884-1895.

[16]. T. Kucherenko, P. Jonell, S. van Waveren, G. E. Henter, S. Alexandersson, I. Leite, H. Kjellstrom, Gesticulator: A framework for semantically-aware speech-driven gesture generation, in *Proceedings of the International Conference on Multimodal Interaction (ICMI'20)*, 2020, pp. 242-250.

[17]. D. Hasegawa, N. Kaneko, S. Shirakawa, H. Sakuta, K. Sumi, Evaluation of speech-to-gesture generation using bi-directional LSTM network, in *Proceedings of the 18th International Conference on Intelligent Virtual Agents (IVA'18)*, 2018, pp. 79-86.

[18]. S. Park, H. S. Shim, M. Chatterjee, K. Sagae, L.-P. Morency, Multimodal analysis and prediction of persuasiveness in online social multimedia, *ACM Transactions on Interactive Intelligent Systems (TiiS)*, Vol. 6, Issue 3, 2016, pp. 1-25.

[19]. B. Nojavanasghari, D. Gopinath, J. Koushik, T. Baltrusaitis, L.-P.˘ Morency, Deep multimodal fusion for persuasiveness prediction, in *Proceedings of the 18th ACM International Conference on Multimodal Interaction (ICMI'16)*, 2016, pp. 284-288.

[20]. N. V. Chawla, K. W. Bowyer, L. O. Hall, W. P. Kegelmeyer, Smote: synthetic minority over-sampling technique, *Journal of Artificial Intelligence Research,* Vol. 16, 2002, pp. 321-357.

[21]. M. Fernandez-Delgado, E. Cernadas, S. Barro, D. Amorim, Do we need hundreds of classifiers to solve real world classification problems?, *The Journal of Machine Learning Research*, Vol. 15, Issue 1, 2014, pp. 3133-3181.

[22]. G. Forman, I. Cohen, Learning from little: Comparison of classifiers given little training, in *Proceedings of the European Conference on Principles of Data Mining and Knowledge Discovery (PKDD'04)*, 2004, pp. 161-172.

[23]. C. Salperwyck, V. Lemaire, D. U. D. P. de Bois, Impact de la taille de l'ensemble d'apprentissage: Une etude empirique, in *Proceedings of the Confrence Internationale Francophone sur l'Extraction et la Gestion de Connaissance*, 2011.

[24]. T. G. Dietterich, Approximate statistical tests for comparing supervised classification learning algorithms, *Neural Computation*, Vol. 10, Issue 7, 1998, pp. 1895-1923.

[25]. N. Ambady, R. Rosenthal, Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis, *Psychological Bulletin*, Vol. 111, Issue 2, 1992, pp. 256-274.

[26]. T. Kucherenko, P. Jonell, Y. Yoon, P. Wolfert, G. E. Henter, A large, crowdsourced evaluation of gesture generation systems on common data: The genea challenge 2020, in *Proceedings of the 26th international Conference on Intelligent User Interfaces (IUI'21)*, 2021, pp. 11-21.

[27]. C. Bartneck, E. Croft, D. Kulić, Measuring the anthropomorphism, animacy, likeability, perceived intelligence and perceived safety of robots, *International Journal of Social Robotics*, Vol. 1, 2009, pp. 71-81.

(2021)

Machine Learning-based Identification of Data-driven Virtual Channels for Automotive Proving Ground Testing

M. Elkafafy, E. Risaliti and B. Cornelis
[1] Siemens Digital Industries Software, 68 Interleuvenlaaan, B-3001 Leuven, Belgium
E-mail: mahmoud.el-kafafy@siemens.com

Summary: This paper introduces machine learning (ML)-based identification of data-driven virtual channels to predict vehicle dynamics and durability-related quantities that require instrumentation of expensive and/or hard-to-install sensors (e.g., wheel force transducers) when performing a proving ground track test to characterize both the ride & handling dynamic behavior of a new vehicle and to confirm its conformity with standard metrics. In such a test, replacing expensive and/or hard-to-install physical sensors by data-driven virtual channels will certainly reduce the test time on the proving ground track, and hence decrease the total cost required. The identification and validation of the obtained data-driven virtual channels are illustrated by means of a real-life data set measured from a full vehicle that was undergoing different handling maneuvers (e.g., sine sweep maneuver). The obtained results showed that the estimations retrieved from the identified data-driven channels agree well with the measured quantities.

Keywords: Machine learning, Neural network, Wheel center loads, Data-driven virtual channel, Vehicle dynamics, Proving ground, Automotive.

1. Introduction

In automotive proving ground testing, the vehicle is traditionally heavily instrumented with different types of physical sensors that are needed to characterize the dynamical behavior of a newly designed variant of a certain vehicle. The physical sensors used in such type of tests could include some expensive and/or hard-to-install sensors. For instance, in vehicle dynamics and endurance tests, wheel force transducers are needed to measure the dynamic forces acting on the center of the wheel hub of the tested vehicle. The wheel center loads (WCL) are important inputs when evaluating the ride & handling and/or the durability performance of a car. Wheel force transducers are known to be costly and cumbersome-to-install which makes it difficult to conduct test campaigns on multiple vehicles. Data-driven virtual channels aim at replacing such type of transducers to increase the reliability and to reduce the cost of measurement campaigns. The basic idea is to utilize other quantities that are cheaper and easier to be measured in order to estimate the expensive and/or cumbersome-to-install physical sensor's signals.

A current trend in the automotive industry is that autonomous driving and driving assistance functions are gaining importance. This trend is accelerating the development of virtual channels, not only due to the increasing usage of control units to improve the overall dynamical behavior of the vehicle and to ensure stability, but also due to the continuing increase of the number of vehicle variants and the high number of conditions to be tested in the design phase. Data-driven virtual channels can be constructed using different degrees of complexity, namely white, grey, and black boxes, as such delivering a different set of values [1-4]. Black box models have the advantage of being reduced order models (ROM) in comparison with the white and grey box models. Being ROM makes them (i.e., the black-box models) convenient and compact digital twins for the modelled system. This contribution will introduce the identification of machine learning (ML)-based data-driven black box models that aim at estimating the wheel center loads (WCL) of a full vehicle as outputs, given other signals that can be measured by cheap and easy-to-install sensors as inputs. Both linear and nonlinear time series regression models are considered in this work.

2. Theory: Linear ARX & Nonlinear ANN

In this section, a brief introduction to the adopted time series regression machine learning models in this study will be given. Two time series regression models namely the linear autoregressive with exogenous input (ARX) model and the nonlinear artificial neural networks (ANN) will be used to identify the targeted data-driven virtual channels. The linear ARX model describes the observed outputs as the sum of a regression on the previous input and output observations and of a white noise process that describes the equation error [5]. This gives the model represented by Eq. (1) (assuming a single input single output (SISO) system) with $y(t)$, $u(t)$, and $e(t)$ the output, the input, and the white noise disturbance (the equation error) respectively at time instant t, where n_a and n_b are the ARX model orders.

$$y(t) = \\ = \sum_{i=1}^{n_b} b_i u(t-i) - \sum_{j=1}^{n_a} a_j y(t-j) + e(t) \quad (1)$$

The adjustable parameters are $\theta = [a_1 \quad a_2 \quad \cdots \quad a_{n_a} \quad b_1 \quad \cdots \quad b_{n_b}]^T$. In a

special case of that model where $n_a = 0$ the output $y(t)$ is modelled as a finite impulse response (FIR). For a multi-input multi output (MIMO) case, the output $y(t)$ is a N_o-dimensional vector with N_o the number of outputs, the input $u(t)$ is a N_i-dimensional vector with N_i the number of inputs, the coefficients b_i are $N_o \times N_i$ matrices, and the a_o are $N_o \times N_o$ matrices.

Artificial neural network (ANN) is a powerful tool that has become a very popular choice of model structure in recent years when identifying a nonlinear dynamical system from time series input-output data [6]. The advantage of the ANN is that it does not rely on prior knowledge of the type of the nonlinearity or on a preconceived mathematical model for the nonlinear system under test. The ANN is composed of a set of connected input/output units, where each connection has a weight associated with it. These connections involve a mapping function that is in general a nonlinear function called the activation function (e.g., hyperbolic tangent sigmoid function). Assuming a neural network with one hidden layer, the general multi-inputs single output (MISO) relation of a time-regression NN can be written as follows:

$$\begin{aligned} y_o(t) = & \sum_{j=1}^{N_n} \Big(b_o + w_j^o \varphi \big(b_j + \\ & + \sum_{i=1}^{N_i} \sum_{d=0}^{N_d} w_{id}^j u_{id}(t) + \\ & + \sum_{j=1}^{N_n} \sum_{k=0}^{N_{nd}} w_j^k x_{jk} \big) \Big), \end{aligned} \tag{2}$$

with N_n the number of neurons in the hidden layer, N_i the number of the inputs to the network, N_d the number of input delays, N_{nd} the number of the feedback delays (i.e. the delays imposed on the hidden layer outputs and sent back as input to the hidden layer), $u_{id}(t)$ the value of the i^{th} input delayed by d time samples, x_{jk} the output of the neuron j at the past time sample k, φ the used activation function, w_{id}^j the weight associated to the delayed input $u_{id}(t)$, w_j^k the weight associated to the delayed hidden layer output x_{jk}, b_j the bias of a hidden neuron j, w_j^o the weight associated to a neuron o in the output layer and a neuron j in the previous hidden layer, b_o the bias of a neuron o in the output layer. From Eq. (2), the 3 basic architectures of the time series regression NN can be deduced. The basic feed forward NN (FFNN) architecture is deduced by setting N_d and N_{nd} equal to zero. Feed Forward NN can be used for any kind of input to output mapping. However, it is preferred when modelling static input-output data meaning that there is no time-dependency in the data sequence. In the FFNN, the information from the input layer to the output layer passing through the hidden layers always moves in one direction, and it never goes backwards. The time delay NN (TDNN) is imposed by setting N_d to a certain positive integer value and N_{nd} to zero. In the TDNN, the input weight has a tap delay line associated with it. This means that both the instantaneous and the past input samples will be used to predict the instantaneous output sample. The recurrent NN (RNN) is reached by setting N_d to zero and N_{nd} to a certain positive integer value. In the RNN each hidden layer has a recurrent (or a feedback) connection with a tap delay (i.e., N_{nd} in Eq. (2)) associated with it. This recurrent connection with a tap delay allows the RNN to acquire state representation. Having this state representation allows the RNN to not only learn from the provided inputs but also from the hidden state of the network. The Time delay NN (TDNN) and recurrent NN (RNN) are meant to model dynamic data where the data sequence has a time-dependency. The ANN learns from the provided input-output data. In this learning process, the parameters of the neural network (cf. biases and weights in Eq. (2)) are optimized using a backpropagation method (e.g., Levenberg-Marquardt (LM) optimization algorithm) over a predefined number of iterations (or epochs). When setting up the NN architecture, the basic hyperparameters to be set are the number of hidden layers and the number of neurons per layer. In case of TDNN and RNN the number of delays imposed on the inputs and internal states respectively also has to be set.

3. Proving Ground Test

The data used in this work originate from a vehicle dynamic proving ground test of a full electric sport car depicted in Fig. 1-top. The proving ground test track is located in Aldenhoven, near Aachen in Germany. The car was fully instrumented with different physical sensors types and was driven over the proving ground to perform certain vehicle dynamics-related maneuvers. The measured signals were acquired with a sampling rate of 1024 Hz for about 100 s. The four suspension systems (lower and upper control arms, steering knuckle, shock absorber) were instrumented with several (+ 50) strain gauges to measure the strains at specific locations and in different directions. The longitudinal, lateral, and vertical acceleration responses of the wheel hub center, the top-mounts connection points (connections between suspension system and car body), and the center of gravity of the car were measured via tri-axial accelerometers. The wheel center loads (WCL) acting on the wheel hub center were measured at each wheel using wheel force transducers (cf. blue circles on each wheel in Fig. 1-top). Each wheel force transducer acquires the 3 forces acting along the x-axis (the backward longitudinal direction), y-axis (i.e., the lateral direction), and the z-axis (i.e., the vertical direction) and their corresponding force moments around those 3-axes. These force transducers are known to be costly and cumbersome-to-install which makes it difficult to conduct test campaigns on multiple vehicles. In addition to the acquired strain, acceleration, WCL signals, some driver action signals (e.g., steering wheel angle, steering torque, engine throttle, brake pressure) and the displacements of the 4 suspension systems were also acquired.

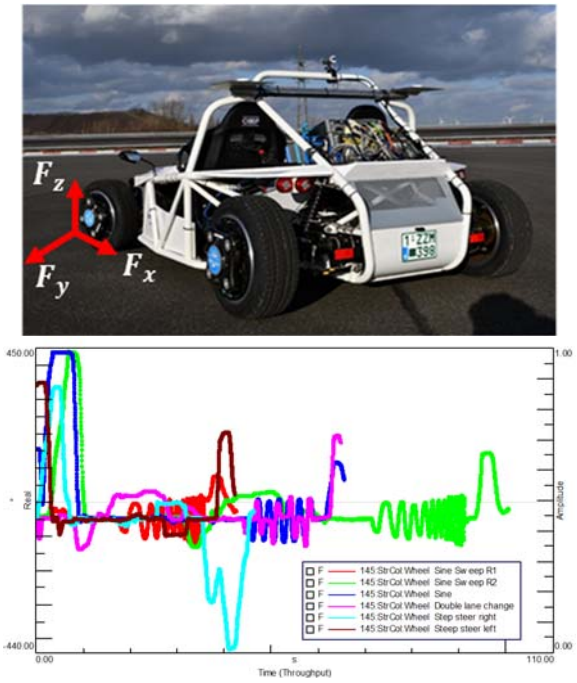


Fig. 1. Instrumented vehicle (top) and the time histories of the steering angle applied to the steering wheel representing the performed lateral maneuvers.

Together with those signals acquired by external sensors, +140 CAN bus signals were also acquired with a sampling rate of 512 Hz. To assess the ride & handling performance of the tested car, five different vehicle dynamic-related maneuvers are performed where the car was mainly under a lateral excitation. Fig.1-bottom shows the time histories of the applied steering wheel angle for the five maneuvers performed. The performed maneuvers in this test were sine sweep, straight driving plus small sine, double lane change, step right, and step left maneuvers. The different signals measured while performing those maneuvers are traditionally used to calculate some key performance indicators (KPIs) that are used to assess the ride & handling performance of the tested car [7]. In the vehicle dynamic context, the measured wheel center loads (i.e., 3 forces & 3 force moments) acting on the wheel hub center are used to determine important handling -related parameters. For instance, during cornering, lateral force of the vehicle is exerted to the four tires and thus they experience side-slips. The angle between the direction of tire's heading and travel direction is known as tire slip angle α, and there exits $F_y = C_\alpha \alpha$, where F_y is the lateral wheel center force and where the proportional constant C_α is known as the cornering stiffness, defined as the slope of the curve for F_y versus α at $\alpha = 0$. The side slip angle α varies under different lateral forces, and therefore needs to be estimated in real time for control applications [8]. Hence, replacing those costly and cumbersome-to-install physical force transducers needed to measure the different wheel center loads by accurate black-box data-driven virtual channels would allow to have cheaper and fast instrumentation which however still delivers high value data. Having this cheap black-box data-driven virtual channel would accelerate the testing campaigns when testing multiple vehicles, and also increase the feasibility of the application of the control strategies.

4. Results and Discussion

This section of the paper deals with the identification and validation of linear and nonlinear black-box data-driven virtual channels that can be used to replace the physical wheel force transducers. The different steps performed to set up the linear & nonlinear models' structures will be detailed in the Subsection 4.1. In the Subsection 4.2, the prediction accuracy and the generalizability of the identified data-driven models will be tested using unseen input data that have not been seen by the models in the training phase. The prediction accuracy in the time-domain will be quantified by the following goodness-of-fit criterion [9]:

$$GoF_j = 100 \times \times \left(1 - \frac{\|L_{j,meas} - L_j\|}{\|L_{j,meas} - mean(L_{j,meas})\|}\right) \%, \quad (3)$$

with $L_{j,meas}$ and L_j the measured and estimated j^{th} wheel center load channel and where $\|.\|$ indicates the 2-norm of a vector. The GoF_j % can take values $-\infty \leq GoF_j \leq 100$ with $GoF_j = 100$ % indicating perfect fit and $GoF_j = -\infty$ an extremely poor fit.

4.1. Inputs Selection & Model Training

4.1.1. Training Data

As it was mentioned in Section 3, the available data consist of five different lateral maneuvers whereby each of them has repeated runs (i.e., different realizations). Two realizations of the sine sweep maneuver are considered for the training of the intended data-driven virtual channel while the other maneuvers (sine, double lane change, step right, and step left) are kept as unseen data to test the generalizability of the identified data-driven models. The sine sweep maneuver, compared to the sine and double lane change maneuvers, has the advantages that the excitation frequency increases over time. Compared to the step steer maneuver, the sine sweep has the advantage of exciting in both the positive and negative directions in the same measurement run. Considering those advantages of the sine sweep maneuver over the other ones, makes it logical to consider it for the training of the models. The first realization of the sine sweep is used to optimize the parameters of the trained model while the second realization is used as a validation set to avoid the overfitting during the training (i.e., the training process will stop in case the validation fit error starts to increase with respect to the training fit error).

4.1.2. Input Channels Selection

The Data-driven virtual channel approach aims to build a time series regression model that estimates a quantity that is costly and cumbersome-to-measure. The inputs to that model need to be carefully selected in order to reach the objective of the intended data-driven virtual channel. The inputs to the virtual channel are supposed to be measured by means of cheap and easy-to-install sensors, and they need to show a high correlation with the targeted outputs (i.e., WCL in this work) in order to maximize the estimation accuracy. Therefore, a main step in building a data-driven virtual channel is to decide about which measured signals will be used as inputs to estimate the intended output quantity. In this work, the input channels are selected in such a way that they fulfill the abovementioned two criteria.

As it was previously mentioned in Section 3, several strain and acceleration signals were acquired during this measurement campaign. From the practical point of view, accelerometers are cheaper and easier-to-install than the strain gauges. Moreover, the 12-wheel center accelerations (3 accelerations × 4 wheels) measured at the center of the wheel hub showed quite good correlations with the 24-wheel center loads (6 WCL × 4 wheels).

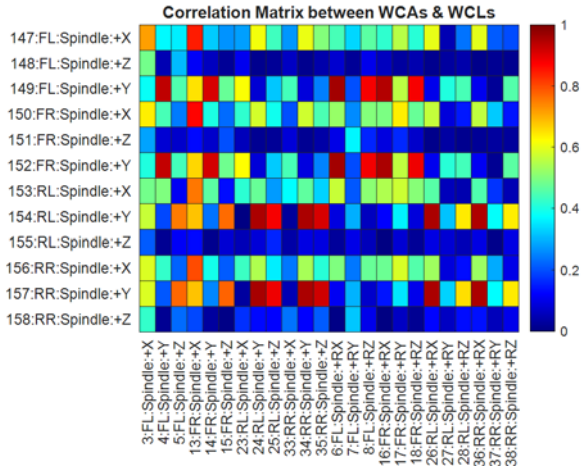


Fig. 3. Pearson's linear correlation coefficient calculated between the 12-wheel center accelerations (vertical axis) and the 24-wheel center loads (horizontal axis). FL: front left wheel, FR: front right wheel, RL: rear left wheel, RR: rear right wheel.

Fig. 3 shows the cross-correlation matrix that represents the values of the Pearson's linear correlation coefficients calculated between the 12-wheel center accelerations (WCA) and the 24-wheel center loads (WCL). From Fig.3, one can see that the WCA show quite good correlations with the 24 WCL. Considering these good correlations with the WCL, the low cost of the accelerometers as physical sensors, and their easiness-to-install, the wheel center accelerations (WCA) have been selected as input channels to the intended black-box data-driven model.

4.1.3. Model Architecture Selection

In this work and as previously introduced in Section 2, the linear ARX model and the non-linear artificial neural network (ANN) will be used to identify the intended data-driven virtual channel. For the linear ARX model, the model orders n_a and n_b (cf. Eq. (1)) have to be set. For the ANN model, the network architecture has to be set in terms of network type (i.e., feedforward (FFNN), time-delay (TDNN), or recurrent (RNN), number of hidden layers, number of neurons per hidden layer, and the number of input or feedback delays in case of TDNN or RNN respectively. To achieve that, a grid search optimization approach is adopted for both the linear ARX and the nonlinear ANN. In case of the linear ARX model, a range of model orders n_a and n_b is tried while making sure that the number of unknown parameters to be estimated does not exceed the number of the available data samples. The model order combination n_a and n_b that gives the lowest mean square validation error is then selected as the best model orders. The best validation performance was found for the case where $n_a = 0$ and $n_b = 55$.

To set the neural network architecture, the number of hidden layers and the maximum number of neurons are firstly set by considering a feed forward NN in which the number of hidden layers is varying from one to three. The first hidden layer is allowed to have a number of neurons as 1, 5, 12, 24, 30, or 50, the second hidden layer is allowed to have a number of neurons as 1,5, or 10, and the third hidden layer is allowed to have a number of neurons as 1 or 5. Based on the best validation performance (i.e., the lowest mean square validation error), this first hyperparameters optimization loop delivers a feed forward NN that has one hidden layer with 30 neurons. Then, extra complexity is added to the architecture by considering TDNN and RNN architectures where in both architectures one hidden layer with number of neurons as 1, 5, 12, 24, or 30 is considered together with a number of inputs (for the TDNN) or feedback (for the RNN) delays is allowed to be 1, 2, or 3 delays. The NN architecture that gave the best validation performance is found to be a recurrent neural network (RNN) with 1 hidden layer, 30 neurons, and 2 feedback delays (cf. Fig. 4). For the training of the ANN, the Bayesian regularization backpropagation learning algorithm is used with number of epochs set to 50 and number of stopping fails set to 6 (i.e., the training process will automatically be terminated in case the validation error continues to increase for 6 consecutive epochs).

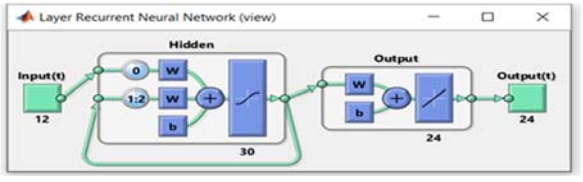


Fig. 4. The optimized neural network architecture.

For the training of both the linear ARX model and the ANN, the inputs and outputs data are normalized such that the minimum and the maximum values of each signal are mapped to the interval [-1,1].

4.2. Model Inference & Validation Results

In this section, the performance of the derived black box data-driven virtual channel (i.e., linear ARX model and the nonlinear RNN model) will be evaluated on unseen validation data sets (i.e., data sets that were not used in the models' training phase) regarding to four lateral maneuvers (i.e., the sinusoidal, double lane change, step-right steer, and step-left steer maneuvers). For each maneuver, the identified data-driven models will be fed by the 12 measured wheel center accelerations (WCA) as inputs, and the models' estimation of the 24-wheel center loads (WCL) will be compared to the real measured ones. The value of the goodness-of-fit calculated according to Eq. (3) is used to quantify the quality of the fit between the measured and the estimated quantities in the time-domain. In this exercise, a total of 96-wheel center loads (6 per wheel $\times$ 4 wheels $\times$ 4 unseen maneuvers) are estimated. Since all the performed maneuvers are lateral ones, the lateral wheel center forces (F_y) and their corresponding resulting force moments (e.g., M_x that causes the vehicle to roll) could be the most important ones among the wheel center loads in this context. Due to the huge number of the estimated channels (i.e., 96 channels) and due to the fact that the performed test campaign was focused on the lateral dynamics, the estimation results of only four load channels (2 lateral force F_y and 2 force moments around the x-axis M_x) are shown in Fig. 5 as an example. The numbers in the figures' legends represent the goodness-of-fit between the estimated and the measured quantity calculated by Eq. (3).

From Fig. 5, one can observe that a good correspondence between the measurements and the estimates is achieved by both the linear ARX model and the RNN where both follow the general trend of the measured signals. By having a deeper look on the time traces (cf., zoom-in shoots in Fig. 5) and by examining the goodness-of-fit values, one can clearly observe that the RNN model outperforms the linear ARX model. The RNN estimation (i.e., the black curve) more closely fits the measurements than the linear ARX model's estimation. This can be explained by the fact that the nonlinear modelling capability of the RNN model enables it to capture the possibly nonlinear dynamics that are induced in the system under test due to the high excitation levels in the lateral directions.

The same observations can also be made when examining the estimation results in the frequency-domain. In Fig. 6, the power spectral density (PSD) of the estimated and the measured quantities are compared for the same wheel center load channels presented in Fig. 5. From Fig. 6, one can notice the superior performance of the RNN-based data-driven virtual channel over the linear ARX -based one where the PSD of the RNN estimation follows well the measured PSD. This can clearly be seen in the sinusoidal maneuver's plot (i.e., Fig. 6 / top plot) where the RNN model follows well both the fundamental frequency peak (i.e., around 0.47 Hz → the excitation frequency) and the second peak (i.e., around 0.94 Hz) which is basically caused by nonlinear dynamics of the system.

The given results both in time and frequency domains show that the obtained RNN-based data-driven virtual channel generalizes well given the fact that the validation data sets (the sinusoidal, double lane change, step-right steer, and step-left steer maneuvers) have not been seen by the model in the training phase. Estimating the WCL for 4 different unseen maneuvers with this good accuracy using the RNN-based data-driven virtual channels which were obtained by training on data measured from only one performed maneuver indicates that this relatively simple approach has the capability to accelerate the testing campaign when a fleet of N_v vehicles need to be tested for N_m various maneuvers, as the testing time can be reduced by a factor of N_m.

5. Conclusions

In this paper, an identification approach of machine learning-based data-driven virtual channels that can accurately estimate the costly and cumbersome-to-measure wheel center loads (WCL) of a full vehicle is demonstrated. Owing to their high correlation with the wheel center loads and their easiness to be measured via inexpensive sensors, the wheel center accelerations (WCA) are selected to be used as the input channels to the developed data-driven virtual channel. A linear ARX model together with different architectures of neural networks are trained using the measured wheel center accelerations and wheel center loads data from a sine sweep maneuver, while the same type of data of 4 different maneuvers are kept as unseen data to test the generalizability of the trained models. In the training phase, the complexity of the linear ARX model and the optimal architecture of the neural network were set by adopting a grid search optimization approach in which the best validation performance was used as a model selection criterion.

It was shown that both the trained linear ARX model and the recurrent neural network (RNN) are capable of estimating the wheel center loads with a reasonable accuracy even for data never seen during the training phase. However, the time and frequency-domains' results showed that the RNN-based virtual channel outperforms the linear ARX-based one owing to its nonlinear modelling capability that helped to capture the existing nonlinear dynamics of the system under test. The presented research delineates the high potential of using the time series regression neural network to build black-box data-driven virtual channels which can be used to replace the costly and cumbersome-to-install sensors in order to make the proving ground test campaigns more efficient.

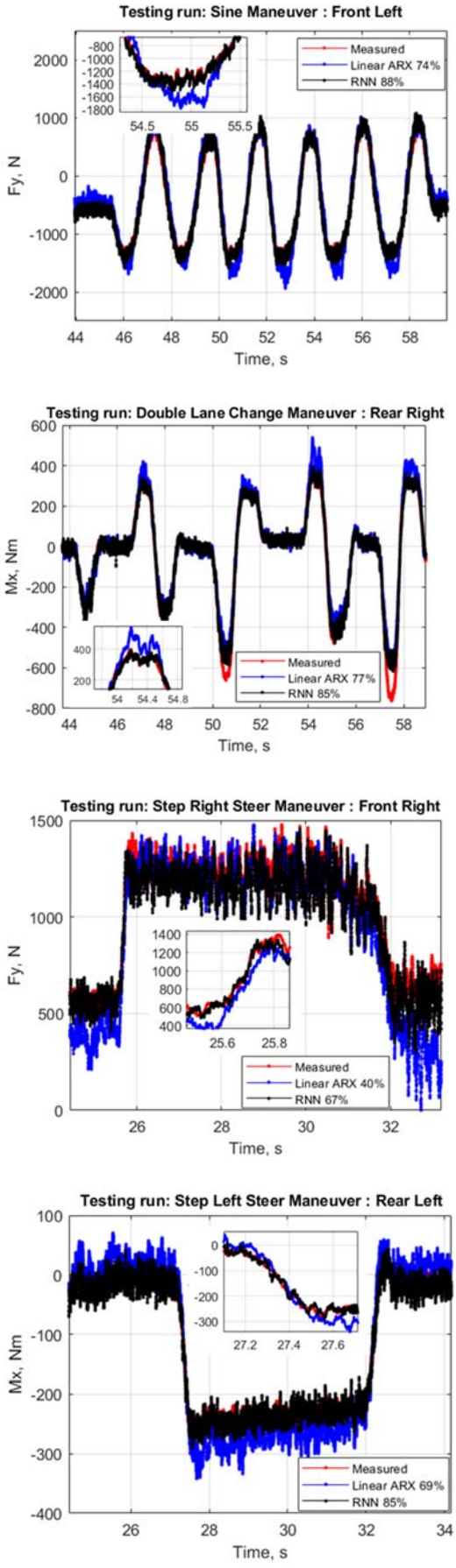


Fig. 5. Time histories: the estimated wheel center load by the linear ARX and nonlinear RNN data-driven virtual channels compared to the measured ones.

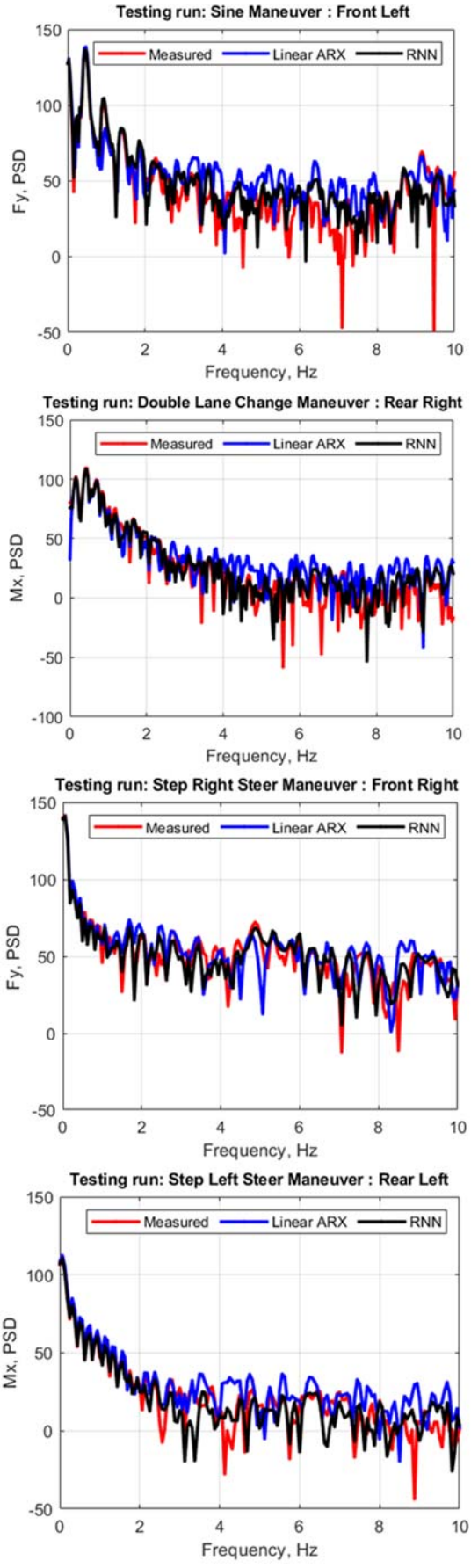


Fig. 6. Power spectral density: the estimated wheel center load by the linear ARX and nonlinear RNN data- driven virtual channels compared to the measured ones.

Acknowledgements

The authors gratefully acknowledge the support of VLAIO (Flemish Innovation & Entrepreneurship) through the O&O project AI4Test (HBC.2022.0005). This project is financed through the European Recovery and Resilience Facility (RRF).

References

[1]. E. Risaliti, T. Tamarozzi, M. Vermaut, B. Cornelis, W. Desmet, Multibody model-based estimation of multiple loads and strain field on a vehicle suspension system, *Mechanical Systems and Signal Processing*, Vol. 123, 2019, pp. 1-25.

[2]. J. D. Willard, et al., Integrating physics-based modeling with machine learning: A Survey, *ArXiv Preprint*, 2020, abs/2003.04919.

[3]. M. Elkafafy, P. Z. Csurcsia, B. Cornelis, E. Risaliti, K. Janssens, Machine learning and system identification for the estimation of data-driven models: An experimental case study illustrated on a tire-suspension system, in *Proceedings of the International Conference on Noise and Vibration Engineering (ISMA'20)*, Leuven, Belgium, 2020, pp. 3287-3301.

[4]. H. Van der Auweraer, D. Hartmann, The executable digital twin: Merging the digital and physical worlds, in *Proceedings of the International Conference on Noise and Vibration Engineering (ISMA'22)*, Leuven, Belgium, 2022.

[5]. L. Ljung, System Identification – Theory for the User, *Prentice-Hall*, Englewood Cliffs, NJ, 1999.

[6]. P. Mas, S. M. Babu, F. Santos, C. Sobie, H. V. d. Auweraer, The application of artificial neural networks in mechatronics system development, in *Proceedings of the International Conference on Noise and Vibration Engineering (ISMA'18)*, Leuven, Belgium, 2018, pp. 2431-2443.

[7]. N. Pagliarecci, F. Zimmer, A. Birouche, M. Basset, Test methodology for the vehicle-tire handling performance evaluation: objectification of driver's subjective assessment, *IFAC-PapersOnLine*, Vol. 53, Issue 2, 2020, pp. 14394-14400.

[8]. Y. Wang, H. Fujimoto. Dynamics control for EVs, Chapter 9, in Modeling Dynamics & Control of Electrified Vehicles (H. Zhang, D. Cao, H. Du, Eds.), *Woodhead Publishing*, New Delhi, India, 2018, pp. 309-337.

[9]. MATLAB 2021a, System Identification Toolbox, *The MathWorks, Inc.*, Natick, Massachusetts, United State, 2021.

(2225)

Embedded Artificial Neural Network for Data Prediction in Energy Efficient Wireless Sensors Networks

Imourane Abdoulaye, Laurent Rodriguez, Cecile Belleudy and Benoît Miramond
Université Côte d'Azur, CNRS, LEAT UMR 7248, 930 Route des Colles, 06903 Sophia Antipolis, France
E-mails: imourane.abdoulaye@etu.univ-cotedazur.fr, {name.surname, cecile.belleudy}@univ-cotedazur.fr

Abstract: In this paper, we propose a novel strategy for minimizing energy consumption in a Wireless Sensor Network using embedded machine learning techniques. Our approach uses a Convolutional Neural Network model embedded in a cluster head node to predict nodes' data to minimize communication instead of transmission of real one. Thanks to this method, the sensor nodes can remain in idle mode and save 90 % of their energy in comparison to normal transmission. In order to validate our algorithm, we have developed a wireless network consisting of electronic boards that can communicate in LoRaWAN, which are compatible with The Things Network server. The results of this test bench show that with this method, it becomes possible to reduce the number of transmissions in the wireless sensor network and to put the sensor nodes in idle mode much longer. This technique allows us to reduce Energy Consumption, and therefore, increases the overall lifetime of the wireless sensor network.

Keywords: Embedded artificial neural network, Wireless sensor networks (WSN), Energy efficiency (EE), Data prediction.

1. Introduction

Nowadays, there are multiple applications using Wireless Sensor Networks (WSN) like smart cities, smart farming, health care, manufacturing, and many others. Wireless nodes are deployed in remote places to collect information from sensors and transmit data to base stations. However, the limited amount of energy in batteries does not allow for long-term operation of these applications. Minimizing the Energy Consumption (EC) of a WSN is, therefore, a crucial problem.

The development of Artificial Intelligence (AI) has allowed the emergence of many innovative approaches to address a wide range of problems [3]. Thanks to the evolution of technology, their integration is made possible in an embedded system, i.e., as close as possible to the data, and therefore to sensors. This allows a reduction in the need for communication and provides a potential solution to minimize the EC of WSN. Some research has been done on power consumption models for WSN by predicting sensor nodes' data instead of communicating them. For example, in [1], the authors worked on a Linear Mean Square (LMS) error-based algorithm, implemented on every node of the network. The sensor nodes don't transmit their data as long as the difference between the prediction and the captured data is less than a threshold. But in this method, the LMS algorithm makes only one prediction for the next instant and the sensor nodes must wake up from the idle mode for capturing for every sample period. In [2], the authors used an Artificial Neural Network (ANN) approach. Their algorithm predicts nodes data. Then, nodes can avoid data transmission and stay in idle mode. The prediction is performed on a central computer in a star topology.

In this paper, we propose a novel strategy by deporting the prediction algorithm in the network architecture. The proposed algorithm is deployed at the level of the cluster head elected among a set of nodes. With this method, it becomes possible to place nodes in idle mode more regularly, reduce the number of transmissions in the network and thus save energy at various levels. The WSN, therefore, gains autonomy.

2. Proposed Prediction Algorithm

The field of embedded AI and in particular on very limited computing resources such as microcontrollers (TinyML) is experiencing significant advances and finds its full meaning in our application. The algorithm can be deployed to provide predictive intelligence based on collected and analyzed data. The principle of the strategy is to place the ANN model for node data prediction at the cluster head level in a hierarchical star topology (Fig. 1).

In this proposed network, the node sends the collected data to the cluster head. First, the cluster head with this data predicts the data for the next sample periods. In this step, predicted data for one sample period mixed with old ones (previous predictions and collected) are used for another prediction. Since the next values can be predicted, we won't need to transmit them. Thus, the sensor nodes can go into idle mode and remain there for more than one sample period. In the second step, the data are transmitted to the base station for display and visualization. In this way, we reduce the number of transmissions and increase the overall lifetime of the network.

Our baseline ANN model is a Convolutional Neural Network (CNN). The architecture is detailed in Table 1. The CNN uses a 1D kernels. The notation **c**

denotes the use of a convolution layer, **p** the use of maxpooling layer and **fc** the use of fully-connected layer. For example, **20-c2** for the first layer means that it is composed of 20 convolution kernels of size 2. The activation function in our network is ReLU and we have a final fully-connected layer for prediction. According to the topology of Fig. 1, we do the training for the three nodes with the same CNN. We concatenate 15 consecutive values for each sensor in an input vector of the CNN, and at the output, we recover a vector of the next ones for each of them.

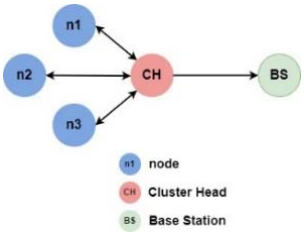


Fig. 1. Network topology.

Table 1. Architecture of the model studied

model	Layer1	Layer2	Layer3	Layer4	Layer5
CNN	20-c2	p2	15-c6	p2	60-fc

3. Experiment and Results

In our project, the Mean Squared Error (MSE) is used as a prediction quality metric. Thus, the weaker it is, the more we can potentially turn off the nodes. In order to evaluate the quality of our method, we designed a network consisting of three sensor nodes, one cluster head node, and a base station (Fig. 1). The nodes are electronic boards that can communicate in LoRaWAN. These boards have been developed at the Laboratoire d'Electronique Antennes et de Télécommunications (LEAT) in partnership with RFThings. The station is a The Things Network server. In the following experiment, the data used to represent the sensors are Ozone concentrations, measured by the Laboratoire Central de Surveillance de la Qualité de l'Air (LCSQA) in France with their stations in Nice Arson, Cannes and Antibes [4].

Fig. 2 represents the comparison between the real data (in blue) and the predicted one (in orange) of node 2. In this figure, the curves of the predicted and the real data are very close.

In Fig. 3, we can see the evolution of average MSE. The CNN model converges with an average MSE of 177.6205 giving 116.85, 191.49, and 224.52 respectively for each node. That means that the data prediction algorithm performed well. The Fig. 4 presents the evolution of the current consumption while sending a message (on LoRaWAN) for a node. By keeping a node in idle mode for a sample period, we can save 90% of the energy in comparison to normal transmission.

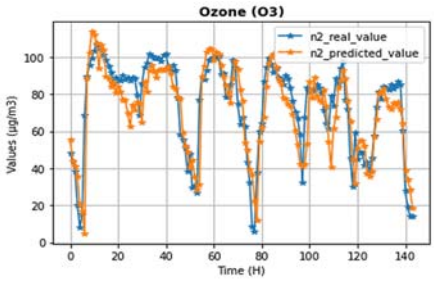


Fig. 2. Evolution of real and predicted data for node 2.

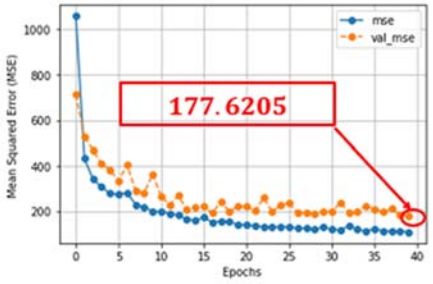


Fig. 3. Evolution of model's MSE.

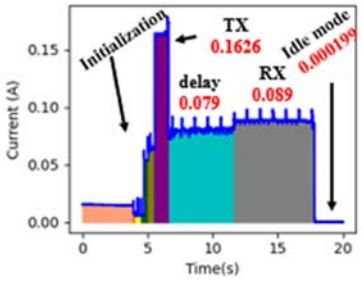


Fig. 4. Current consumption of a sensor node

4. Conclusion

In this work, a new strategy to reduce power consumption in a WSN by making data predictions to minimize communication with an embedded CNN on a cluster head node has been proposed. We checked the performance of the developed prediction model, and the results show that the prediction data is very close to the real one. In this first experiment, our method allowed a reduction of the energy consumption by a factor 2. As a future scope of work, our algorithm will be fully deployed with a suitable routing protocol and multi-sensors to evaluate more realistic values of the number of transmissions that can be avoided and the power saving.

References

[1]. B. Stojkoska, D. Solev, and D. Davcev, Data prediction in WSN using Variable Step Size LMS Algorithm, in *Proceedings of the 5th International Conference on Sensor Technologies and Applications, 2011 (SENSORCOMM' 2011),* Nice, Saint Laurent du Var, France, 21-27 August 2011, pp. 191-196.

[2]. R. Kromes, A. Russo, B. Miramond, and F. Verdier, Energy consumption minimization on LoRaWAN network by using an Artificial Neural Network based application, in *Proceedings of the IEEE Sensors Applications Symposium (SAS' 2019),* Sophia Antipolis, France, 11-13 March 2019, pp. 1-6.

[3]. S. Dong, P. Wang, and K. Abbas, A survey on deep learning and its applications, *Computer Science Review,* Vol. 40, Issue C, May 2021, Article ID 100379.

[4]. Data Gouv Web Portal (https://www.data.gouv.fr/fr/datasets/)

(2289)

Better Exploiting the Redundancy of Serial Manipulators with Fuzzy Logic: A Planar 3-DOF Example

L. Baron and N. Hassantabar
Polytechnique Montréal, PO 6079, station CV, Montréal, QC, Canada, H3C 3A7
Tel.: + 15143404711 ext. 4744, fax: + 15143405170
E-mail: Luc.Baron@polymtl.ca

Summary: For tasks, like milling, welding, water-jet cutting or deburring, industrial 6-joint manipulators are redundant, as they have more mobilities than required by the task. The extra mobility can be exploited to avoid the joint limits. Exploring the self-motion of the manipulator at each end-effector position requires moving onto the null-space of the instantaneous Jacobian matrix, a very time-consuming procedure. However, this exploration can be precomputed, and saved as the bounds of the self-motion for a grid of workspace end-point positions. The ANFIS fuzzy logic system is used to predict these bounds of the feasible region of joint trajectories. Therefore, the complexity of solving redundancy is taken off-line in the training dataset, including its knowledge of redundancy. Once the self-motion bounds are known for a specific joint, the user or a decision-making system can select the most appropriate solution within the feasible region. By using an ANFIS system of 8 fuzzy sets on the x and y inputs, good numerical results are obtained on the prediction of self-motion bounds.

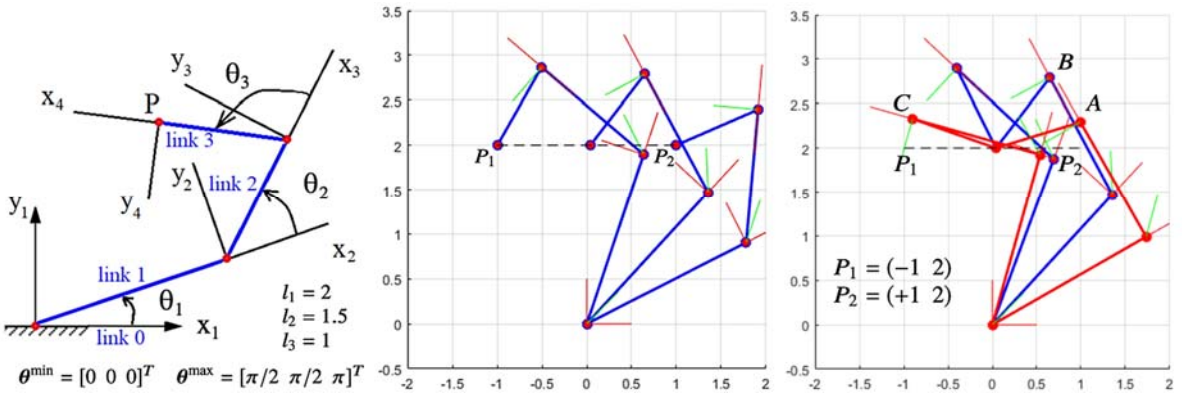
Keywords: Inverse kinematic, Redundant manipulator, Self-motion, Fuzzy logic, ANFIS.

1. Introduction

Most of the robotic tasks, like milling, welding, water-jet cutting or deburring, require less than 6 degree-of-freedom (DOF). When perform with an industrial 6-joint manipulator, the latter is said redundant, and the extra mobility can be used to avoid joint limits. This is obviously an issue of major relevance for motion planning and control purposes [1]. The geometrical methods solve the inverse kinematics at the displacement level with the nonlinear equations with possibly multiple solutions. They usually set arbitrarily one joint and verify if the limits are satisfied on the other joints. Analytical methods solve the same problem at the velocity level by projecting a joint motion onto the null-space of the instantaneous Jacobian matrix. They both suffer from the lack of continuity, the possible bad conditioning of the Jacobian, and a heavy computational load [2, 3]. Artificial intelligence methods rely on forward kinematics to compute a training dataset and use an interpolating tool, like fuzzy logic, to predict solutions with smaller computational load [4]. For example, a fuzzy neural network is used to compute a joint hint as the minimum-norm solution of the instantaneous Jacobian matrix and used in an iterative procedure to converge to an accurate solution [5]. However, they suffer from a lack of knowledge of redundancy [6].

1.1. Self-motion Behavior

A planar 3-DOF redundant manipulation is shown in Fig. 1, while performing a straight-line trajectory from P_1 to P_2. At the mid-point, configuration B is the one selected by the user. Configurations A and C are both at the limit of its joint displacement.



(a) Manipulator geometry (b) Straight line from P_1 to P_2 (c) Self-motion at mid-point of P_1–P_2

Fig. 1. The planar 3-DOF redundant serial manipulator.

The null-space manifold of trajectory P_1-P_2 is shown in Fig. 2. The joint limits are represented by the six planes of the feasible box. Each black curve is the self-motion of the manipulator, i.e., maintaining the EE position while moving the manipulator joint. The minimum-norm solution is shown in green, while any solution curve on the surface of the manifold from the self-motion P_1 to the self-motion P_2 is a feasible solution. For this trajectory, the upper bounds of self-motions initially come from θ_1^{max}, then from θ_2^{max}, and finally, from θ_3^{max}, while the lower bounds come from θ_2^{max}. Since the null-space manifold is only composed of feasible self-motions, it is bounded by the first joint which reaches its limit.

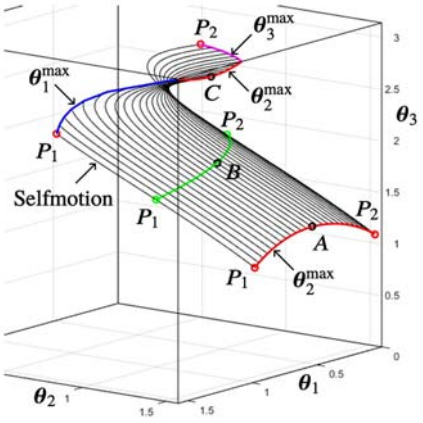


Fig. 2. Null-space manifold from P_1 to P_2.

Fig. 3 shows the projection of the null-space manifold onto each individual joint. Apparently, all the curves that cross the self-motion from P_1 to P_2 are feasible solutions. For θ_1, we obtain the bounds as described previously. For θ_2, both bounds are equal to θ_2^{max} between steps 6 to 18. However, the minimum-norm solution is a curve that intersects the self-motions, but not always between the bounds. Selecting solution curves between bounds is too restrictive for path planning. Clearly, the minimum-norm solution should also to be considered.

1.2. Self-motion-bounds Prediction

We propose to compute off-line the minimum-norm solution together with the upper and lower bounds of each manipulator self-motion for a grid of positions [7]. A fuzzy logic system will be trained off-line in order to predict on-line the bounds of the self-motions. Then a user or a decision-making system will be able to selection a feasible trajectory on the surface of the manifold that satisfies the required trajectory and other criteria.

2. Workspace Discretization

The manipulator workspace is divided into three zones of different density, as shown in Fig. 4. An analytical method is used off-line at each point to compute the upper and lower bounds of the self-motion together with the minimum-norm solution. The dataset contains 3356 points for training and 100 undisclosed points for testing purposes with the manipulator geometry and trajectory shown in Fig. 1.

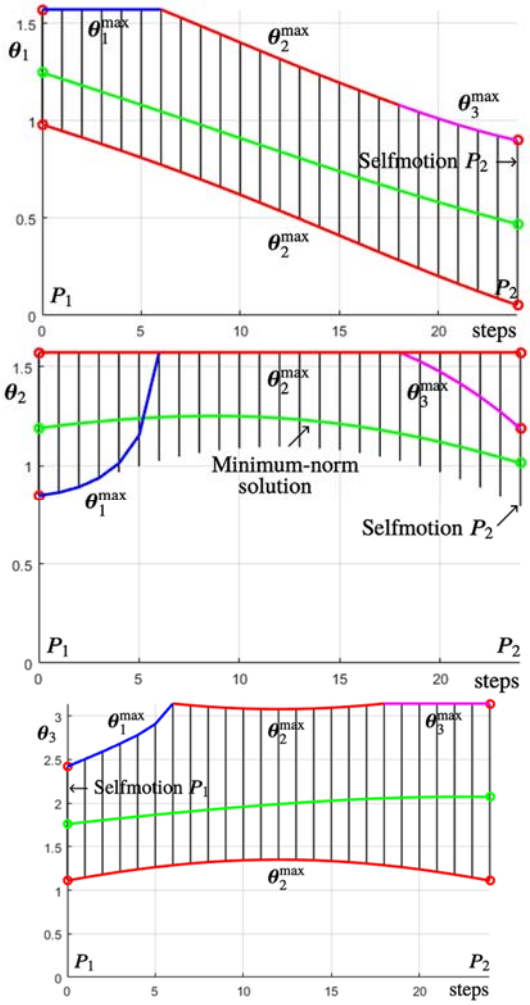


Fig. 3. Projection of the self-motion of Fig. 2 on each individual joint: the minimum-norm solution is in green, while the bounds limited by θ_1 in blue, limited by θ_2 in red and limited by θ_3 in purple.

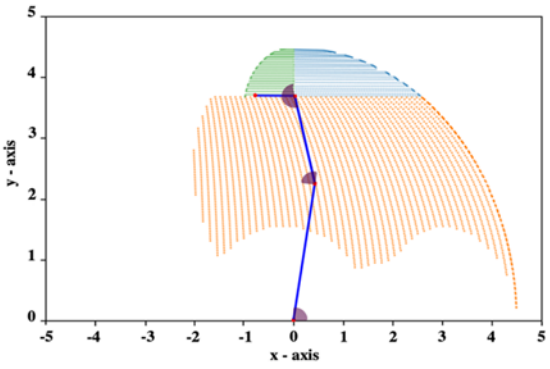


Fig. 4. Discretization of the workspace.

3. Bounds Prediction with ANFIS

The Adaptive Neuro-Fuzzy Inference System (ANFIS) is an artificial neural network with hybrid learning procedure that combines the principle of fuzzy set theory and fuzzy inference systems with the learning ability of the neural networks. ANFIS from the MATLAB's Fuzzy Logic Toolkit is used in this work.

Three different ANFIS systems, such as the one shown in Fig. 5, are trained for the prediction of θ_1-upper-bound, θ_1-lower-bound and θ_1-minimum-norm-solution. The input layer consists of x and y variables with 8 Gaussian membership functions. Therefore, a maximum of 64 rules using *AND operator* (minimum and product) and *OR operator* (maximum and probor) is used on the fuzzification layer with consequent as Sugeno-type linear relationship on the output. The centroid is used for defuzzification method.

After 150 training epochs, we compare the upper and lower bounds of self-motion as predicted by ANFIS with the test dataset. As shown in Fig. 6, the bounds prediction is very close to the test dataset.

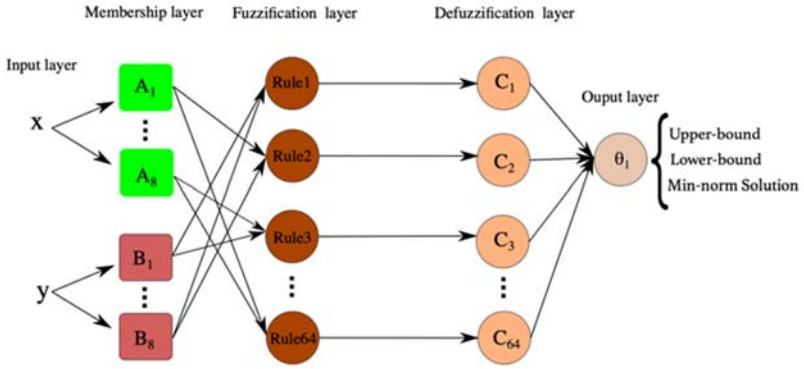


Fig. 5. ANFIS bounds prediction structure.

4. Numerical Example

The prediction of bounds with fuzzy logic is validated on a triangular trajectory C from $P_0 = [-1\ 3]^T$ to $P_1 = [1\ 3]^T$ to $P_2 = [0\ 4]^T$ and back to P_0. Along the trajectory, we use ANFIS to predict the bounds of the self-motion, such as shown in Fig. 7.

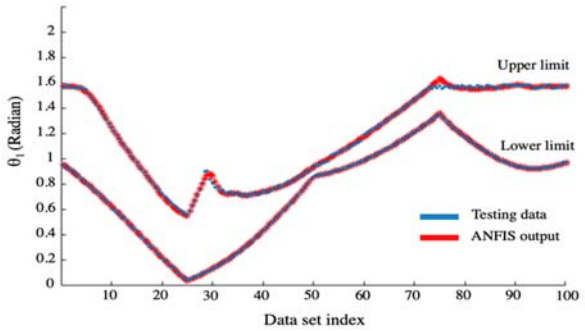


Fig. 6. Bounds prediction and testing dataset.

The computation is performed by ANFIS in just a few milliseconds rather than minutes by analytical methods based on the projection of a gradient vector onto the null-space of the instantaneous Jacobian matrix. Once a solution of a joint position is selected, we use a deterministic geometrical method to compute the other joint positions. Obviously, multiple joint solutions still exist, but their number is greatly reduced as well as the complexity of the equations to be solved.

For trajectory C, if we plot the self-motion bounds along θ_2 (instead of θ_1 as in Fig. 6), we obtain bounds that cross each other for which the minimum-norm solution becomes useful to the user.

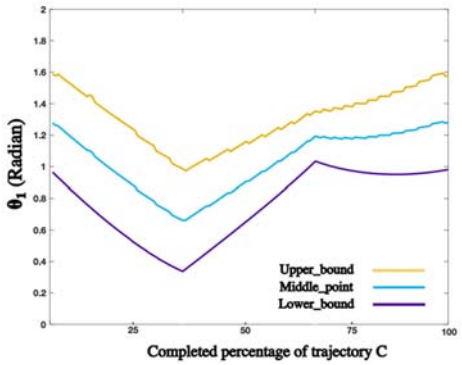


Fig. 7. Self-motion-bounds along trajectory C.

5. Conclusions

ANFIS has been successfully used to predict the self-motion bounds of a planar 3-DOF redundant serial manipulator. The resolution of the x and y input grid must to be adapted in order to achieve the desired prediction accuracy on the self-motion bounds. Their prediction by ANFIS takes only a few milliseconds with the computation of other joint positions. The methodology should to be extended to manipulators with more joints and having both positioning and orientation mobility. This ANFIS bounds prediction tool facilitates users' trajectory planning.

Acknowledgements

We acknowledge the financial support of the Natural Sciences and Engineering Research Council of Canada under grant RGPIN-203618.

References

[1]. S. Chiaverini, G. Oriolo, I. D. Walker, Kinematically redundant manipulators, in Handbook of Robotics (B. Siciliano, O. Khatib, Eds.), *Springer*, 2008, pp. 245-268.

[2]. L. Huo, L. Baron, The self-adaptation of weights for joint-limits and singularity avoidances of functionally redundant robotic-task, *Robotics and Computer Integrated Manufacturing*, Vol. 27, Issue 2, April 2011, pp. 367-376.

[3]. L. Das, J. Nanda, S. Mahapatra, A comparative study of prediction of inverse kinematics solution of 2-dof, 3-dof and 5-dof redundant manipulators by ANFIS, *International Journal of Computer Science and Network*, Vol. 3, Issue 5, October 2014, pp. 304-308.

[4]. D. Deshmukh, D. K. Pratihar, A. K. Deb, H. Ray, A. Ghosh, ANFIS-based inverse kinematics and forward dynamics of 3 DOF serial manipulator, *in Proceedings of the International Conference on Hybrid Intelligent Systems (HIS'20)*, 2020, pp. 144-156.

[5]. S. F. M. Assal, K. Watanabe, K. Izumi, Neural network-based kinematic inversion of industrial redundant robots using cooperative fuzzy hint for the joint limits avoidance, *IEEE/ASME Transactions on Mechatronics*, Vol. 11, Issue 5, October 2006, pp. 593-603.

[6]. F. Flacco, A. De Luca, O. Khatib, Control of redundant robots under hard joint constraints: saturation in the null space, *IEEE Transactions on Robotics*, Vol. 31, Issue 3, June 2015, pp. 637-654.

[7]. N. Hassantabar, Automating kinematic resolution of redundant planar manipulators using Adaptive Neuro-Fuzzy Inference System (ANFIS) modeling, MD Thesis, *Polytechnique Montréal*, December 2021.

(2497)

Saliency Map from Subitizing Processing: First Approximation

Carola Figueroa-Flores [1] and Ismael Miranda [2]
[1] Department of Computer Science and Information Technology, Universidad del Bío-Bio, Chile
[2] School of Computer and Information Engineering, Universidad del Bío-Bío, Chile
E-mail: cfigueroa@ubiobio.cl

Summary: Most of the saliency methods are evaluated on their ability to generate saliency maps, and not on their functionality in a complete vision pipeline, like for instance, image classification or salient object subitizing.

In this paper, we study the problem of salient object subitizing, i.e., predicting the number of salient objects in a synthetic image (SID4VAM and Toronto). This task is inspired by the ability of people to quickly and accurately identify the number of items within the subitizing range (1-4). This means that the subitized information will tell us the number of featured objects in a given image, and will thus subsequently obtain the location or appearance information of the featured objects, and everything will be done within a weakly supervised configuration.

Keywords: Saliency prediction, Subitizing, Object recognition, Deep learning and convolutional neural network.

1. Introduction

For humans, object recognition is a nearly instantaneous, precise, and extremely adaptable process. Furthermore, we have the innate ability to learn new classes of objects from a few examples [1, 2]. The human brain reduces the complexity of incoming data by filtering out some of the information and processes only those things that grab our attention. This, combined with our biological predisposition to respond to certain shapes or colors, allows us to recognize at a glance the most important or outstanding regions of an image. This mechanism can be observed by analyzing which parts of the image's humans pay more attention to; for example, where they fix their eyes when they are shown an image [3]. The most accurate way to record this behavior is by tracking eye movements while the subject in question is presented with a set of images to evaluate.

Computational estimation of salience (or salient or salient regions) aims to identify to what extent regions or objects stand out from their surroundings or background to human observers. Saliency maps can be used in a wide range of applications, including object detection, image and video understanding, and lastly eye tracking. On the other hand, we know that the human visual system can effortlessly identify the number of objects in the range 1 to 4 by doing just one glance [4]. Since then, this phenomenon, coined later by [5] as Subitization, has been studied and tested in various experimental settings [6].

Therefore, inspired by subitization and the results obtained in [7, 8], the main objective of this project is to incorporate the subitization of salient objects (SOS), in order to improve the results obtained in the thesis. This means that the subitized information will tell us the number of outgoing objects in a given image, and thus subsequently provide us with the location or appearance information of the outgoing objects explicitly, and everything will be done within a weakly supervised configuration. It should be noted that when we train our network with the subitization supervisions, the network will learn to focus on the regions related to the outgoing objects. Therefore, we will design a saliency subitization process (SSP) architecture that is responsible for extracting attention regions as salience map.

Finally, in this work we propose to design and build a convolutional neural network (CNN), which will basically consist of a process that will be in charge of SSP. The general model of our proposal is shown in Fig. 1.

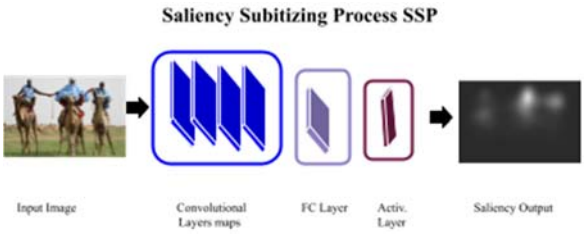


Fig. 1. Overview of our method.

2. Related Work

Saliency is generally known as local contrast [9], which generally originates from contrasts between objects and their surroundings, such as differences in color, texture, shape, etc. This mechanism measures intrinsically salient stimuli to the vision system that primarily attract the attention of humans, in the initial stage of visual exposure to an input image [10].

To quickly extract the most relevant information from a scene, the human visual system pays more attention to highlighted regions, as seen in Fig. 1. Research on computational saliency focuses on the design of algorithms that, like human vision, predict which regions of a scene stand out [11, 12].

Initial efforts to model saliency involved multi-scale representations of color, orientation, and intensity contrast. These were often biologically

inspired, such as the well-known work of [9, 13]. From that model, a large number of models were based on the manual elaboration of these features to obtain an accurate saliency map [14, 15], either maximizing [16] or learning statistics from natural images [10, 17]. Relevancy research was further driven by the availability of large datasets that enabled the use of machine learning algorithms, primarily pre-trained on existing human fixation data.

The question of whether saliency is important for object recognition and tracking has been raised in [18]. More recent methods take advantage of end-to-end convolutional architectures by fine-graining on fixation prediction [19, 20]. But the main goal of these works was to estimate a saliency map, not how saliency might contribute to object recognition.

Several works have shown that having the saliency map of an image can be useful for object recognition. For example, in [21, 7]. Since the saliency map can help focus attention on the relevant parts of the image to improve recognition. Additionally, it can help guide training by focusing backpropagation on relevant image regions. In previous work, we have shown that saliency modulated image classification (SMIC) is especially efficient for training on datasets with few labeled data [7]. The main drawback of these methods is that they require a trained saliency method. Also, in [8] we show that this restriction can be removed and that we can hallucinate the saliency image from the RGB image. By training the network for image classification on the ImageNet dataset [22], we can obtain the saliency branch without using human reference images.

Recently, the progress in the detection of salient objects has grown substantially, mainly benefiting from the development of Deep Neural Networks (CNN). In this work [23] a CNN based on the use of super pixels for saliency detection was proposed. Instead [24] used multi-scale features extracted from a deep CNN. In [25] proposed a multi-context deep learning framework to detect salient objects with two different CNNs, which were useful for learning local and global information. In [26], they proposed a saliency detection framework, which extracted the correlations between object contours and RGB features of the image. On the other hand, [27] defined a pyramid-shaped structure to expand the receptive field in visual attention. In [28], they introduced short connections for edge or contour detection. Zhu, on the other hand, in [29] proposed a visual attention architecture called DenseASPP, to extract information.

Chen in [30] proposed a spatial attenuation context network, which recursively translated and aggregated the context features in different layers.

Tu En [31] introduced an edge-guided block to embed boundary information in saliency maps. Zhou in [32] proposed a multi-type self-attention network to learn more semantic details from degraded images. However, these methods rely heavily on pixel-based monitoring. Overcoming the scarcity of pixel-based data, we focus on the saliency detection task.

Weakly supervised saliency detection. There are many works using weak supervisions for the saliency detection task.

For example, Li in [33] used the image-level labels to train the classification network and applied coarse activation maps as saliency maps. Wang in [34] proposed a weakly supervised two-stage method by designing an inference network to predict foreground regions and Global Smooth Pooling (GSP) to aggregate responses from those predicted objects. On the other hand, Zeng in [35] designed a unified network, which is capable of weak monitoring of multiple sources, including image labels, captions, and pseudo-labels. Furthermore, they designed a loss of attention transfer to transmit signals between subnetworks with different supervisions.

Different from the previous methods, we propose to use subitization information as weak supervision in the saliency detection task. Where we will first study the problem of subitization of the outgoing object and the relationships between subitization and saliency detection.

3. Proposed Method

3.1. Subitization of Salience Process (SSP)

It should be noted that the information provided by the subitization process will indicate the number of outgoing objects in a given image [36]. Therefore, it will not explicitly provide the location or information related to the appearance of the output objects. However, when the network is being trained with subtizing (simulating supervised learning), the network will learn to focus on the regions related to the most salient (or salient) objects. For the same reason, we will design the SSP to extract these regions as if it were a salience mask.

In addition, we apply the Grad-CAM technique to extract salient regions such as initial salience maps, containing the gradient information flowing into the last convolutional layers during the recoil phase. Gradient information represents the importance of each neuron during inference of the network. We assume that the features produced from the last convolutional layer has a channel size of K. And, finally we add an activation map with ReLU function layer, that this function filters negative gradient values, since only the positive ones contribute to the class decision, while the negative values contribute to other categories. The size of the saliency map is the same as the size of the last convolutional feature maps (1/8 of the original resolution).

3.2. Experimental Setup

Datasets. We have computed the saliency maps for images from distinct eye-tracking dataset, corresponding to 120 real scenes (Toronto), and

230 synthetic images with specific feature contrast (SID4VAM) see Table 1. We have computed these images class dataset with our approach, supervised artificial model that specifically compute high-level features (DeepGazeII, ML-Net, SAM, salGAN), and models biological inspiration (IKN, AIM, SDLF and GBVS).

Table 1. Characteristics of eye-tracking dataset.

Dataset	Type	# Images	Resolution
TORONTO	Indoors & Outdoors	120	681×511
SID4VAM	Synthetic pop-out	230	1280×1024

Networks architectures. We evaluate our approach using two network architectures: Alexnet [44] and Resnet-152 [43] and modify it to meet our requirement. In both cases, the weights were pretrained on Imagenet and then finetuned on each of the datasets mentioned above. The networks were trained for 70 epochs with a learning rate of 0.0001 and a weight decay of 0.005. The top classification layer was initialized from scratch using Xavier method. The SSP consists of four convolutional layers for Alexnet and four residual blocks for Resnet-152.

Comparison. In this work, we are going to compare our proposal with other models, for instance, DeepGazeII summed the center baseline whereas ML-Net and SAM the learned priors are used for modulating the result of the network.

4. Conclusions

In this paper, we proposed a method for the saliency estimation with subitizing supervision. We design a model with the Saliency Subitizing process SSP. Our method could outperform other weakly supervised methods and even perform comparable to some fully supervised methods.

Acknowledgements

We thank for the support from FOVI21001 Fomento a la Vinculación Internacional para Instituciones de Investigación Regionales (ANID, Chile) Agencia Nacional de Investigación y Desarrollo, ALBA Research Group (Algorithms and Database) 2130591 GI/VC, Ayudantes para el Fortalecimiento de Investigación FACE 2022 and Proyecto de Reinserción: DIUBB 2230508 IF/RS of the University of Bío-Bío.

References

[1]. X. Sun, H. Yao, R. Ji, X. Liu, Toward statistical modeling of saccadic eye-movement and visual saliency, *IEEE Transactions on Image Processing*, Vol. 23, Issue 11, 2014, pp. 4649-4662.

[2]. B. T. Vincent, B. W. Tatler, Systematic tendencies in scene viewing, *Journal of Eye Movement Research*, Vol. 2, Issue 2, 2008, 5.

[3]. R. Achanta, S. Hemami, F. Estrada, S. Susstrunk, Frequency-tuned salient region detection, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'09)*, 2009, pp. 1597-1604.

[4]. J. Stanley, The power of numerical discrimination, *Nature*, Vol. 3, Issue 3, 1871, pp. 367-367.

[5]. E. L. Kaufman, M. W. Lord, T. W. Reese, J. Volkmann, The discrimination of visual number, *The American Journal of Psychology*, Vol. 62, 1949, pp. 498-525.

[6]. J. Whalen, C. Gallistel, R. Gelman, Nonverbal counting in humans: The psychophysics of number representation, *Psychological Science*, Vol. 10, 1999, pp. 130-137.

[7]. C. Figueroa-Flores, B. C. Raducanu, D. Berga, J. v. d. Weijer, Hallucinating saliency maps for fine-grained image classification for limited data domains, in *Proceedings of the 16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP'21)*, Vol. 4, 2021, pp. 163-171.

[8]. C. Figueroa-Flores, B. D. J. van deWeijer, B. Raducanu, Saliency for free: Saliency prediction as a side-effect of object recognition, *Pattern Recognit. Letter*, Vol. 150, 2021, pp. 1-7.

[9]. I. Itti, C. Koch, Computational modeling of visual attention. Nature review, *Neuroscience*, Vol. 2, 2001, pp. 194-203.

[10]. J. Harel, C. Koch, P. Perona, Graph-based visual saliency, *Advances in Neural Information Processing Systems*, Vol. 19, 2007, pp. 545-552.

[11]. Y. Li, X. Hou, C. Koch, J. Rehg, A. Yuille, The secrets of salient object segmentation, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'14)*, 2014, pp. 280-287.

[12]. K. Xu, J. Ba, R. Cho, A. Courville, R. Salakhudinov, Y. Bengio, Show, attend and tell: Neural image caption generation with visual attention, in *Proceedings of the International Conference on Machine Learning (ICML'15)*, 2015, pp. 2048-2057.

[13]. L. Itti, C. Koch, E. Niebur, A model of saliency-based visual attention for rapid scene analysis, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 20, Issue 11, 1998, pp. 1254-1259.

[14]. A. Borji, L. Itti, State-of-the-art in visual attention modeling, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 35, Issue 1, 2013, pp. 185-207.

[15]. Z. Bylinskii, E. DeGennaro, R. Rajalingham, H. Ruda, J. Zhang, J. Tsotsos, Towards the quantitative evaluation of visual attention models, *Vision Research*, Vol. 116, 2015, pp. 258-268.

[16]. N. Bruce, J. Tsotsos, Saliency based on information maximization, in *Proceedings of the 18th International Conference on Neural Information Processing Systems (NeurIPS'05)*, 2005, pp. 155-162.

[17]. A. Torralba, A. Oliva, M. Castelhano, J. Henderson, Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search, *Psychological Review*, Vol. 113, Issue 4, 2006, pp. 766-786.

[18]. S. Han, N. Vasconcelos, Biologically plausible saliency mechanisms improve feedforward object recognition, *Vision Research*, Vol. 50, 2010, pp. 2295-2307.

[19]. M. Cornia, L. Baraldi, G. Serra, R. Cucchiara, Predicting human eye fixations via an LSTM-based saliency attentive model, *IEEE Transactions on Image Processing*, Vol. 27, Issue 10, 2018, pp. 5142-5154.

[20]. J. Pan, C. Canton, K. McGuinness, N. O'Connor, J. Torres, E. Sayrol, X. Giro-i-Nieto, SalGAN: Visual saliency prediction with generative adversarial networks, *arXiv Preprint*, 2018, arXiv:1701.01081.

[21]. C. Figueroa-Flores, A. Gonzalez-Garcia, J. van de Weijer, B. Raducanu, Saliency for fine-grained object recognition in domains with scarce training data, *Pattern Recognition*, Vol. 94, 2019, pp. 62-73.

[22]. O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Imagenet large scale visual recognition challenge, *International Journal of Computer Vision*, Vol. 115, Issue 3, 2015, pp. 211-252.

[23]. Y. Qin, H. Lu, Y. Xu, H. Wang, Saliency detection via cellular automata, in Proceedings of the *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'15)*, 2015, pp. 110-119.

[24]. C. Li, Y. Yuan, W. Cai, Y. Xia, D. Feng, Robust saliency detection via regularized random walks ranking, in Proceedings of the *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'15)*, 2015, pp. 2710-2717.

[25]. J. Zhao, S. Sun, X. Liu, A. Yang, A novel biologically inspired visual saliency model, *Cognitive Computation*, Vol. 6, Issue 4, 2014, pp. 841-848.

[26]. Y. Yuan, C. Li, J. Kim, W. Cai, D. Feng, Reversion correction and regularized random walk ranking for saliency detection, *IEEE Transactions on Image Processing*, Vol. 27, Issue 3, 2017, pp. 1311-1322.

[27]. W. Wang, J. Shen, Deep visual attention prediction, *IEEE Transactions on Image Processing*, Vol. 27, Issue 5, 2018, pp. 2368-2378.

[28]. X. Hou, L. Zhang, Saliency detection: A spectral residual approach, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'07)*, 2007, pp. 1-8.

[29]. W. Zhu, S. Liang, S. Wei, J. Sun, Saliency optimization from robust background detection, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'14)*, 2014, pp. 2814-2821.

[30]. C. Chen, H. Tang, Z. Lyu, H. Liang, J. Shang, M. Serem, Saliency modeling via outlier detection, *Journal of Electronic Imaging*, Vol. 23, Issue 5, 2014, 053023.

[31]. Z. Tu, Y. Ma, C. Li, J. Tang, B. Luo, Edge-guided Non-local Fully Convolutional Network for Salient Object Detection, http://arxiv.org/abs/1908.02460

[32]. Z. Zhou, Z. Wang, H. Lu, S. Wang, M. Sun, Multi-type self-attention guided degraded saliency detection, in Proceedings of the *AAAI Conference on Artificial Intelligence*, Vol. 34, 2020, pp. 13082-13089.

[33]. G. Li, Y. Yu, EEP contrast learning for salient object detection, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'16),* 2016, pp. 478-487.

[34]. L. Wang, H. Lu, Y. Wang, M. Feng, D. Wang, B. Yin, X. Ruan, Learning to detect salient objects with image-level supervision, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'17)*, 2017, pp. 3796-3805.

[35]. Y. Zeng, Y. Zhuge, H. Lu, L. Zhang, M. Qian, Y. Yu, Multi-source Weak Supervision for Saliency Detection, http://arxiv.org/abs/1904.00566

[36]. S. He, J. Jiao, X. Zhang, G. Han, R. Lau, Delving into Salient Object Subitizing and Detection, in *Proceedings of the IEEE International Conference on Computer Vision (ICCV'17)*, 2017, pp. 1059-1067.

[37]. F. Murabito, C. Spampinato, S. Palazzo, D. Giordano, K. Pogerelov, M. Riegler, Top-down saliency detection driven by visual classification, *Computer Vision and Image Understanding*, Vol. 172, July 2018, 2018, pp. 67-76.

(2517)

Inspection Approach for Automated In-line Defect Detection on Decorated Foil Plates

P. J. Krassnig, M. Haselmann and D. P. Gruber
[1] Polymer Competence Center Leoben GmbH, Roseggerstraße 12, 8700 Leoben, Austria
Tel.: + 433842429627 +749/+44
E-mails: paul.krassnig@pccl.at, matthias.haselmann@pccl.at

Summary: Optical quality inspection plays a major role in the manufacturing industry, and elaborate manual visual inspection processes are increasingly being replaced and automated by means of ML-based visual inspection systems. The visualization of product-specific defect types at sufficient contrasts can usually only be realized with customized optical modalities, further building the data basis for defect detection algorithms. This extended abstract presents a novel inspection approach for visualizing production-related defect types of decorated foil plates, enabling automated in-line full-surface defect detection. The inspection approach consists of three different optical modalities embedded in a dynamic sequential defect detection procedure, applicable for the integration into the production line. The developed optical modalities enable the visualization of the demanded production-related defect classes with sufficient contrast, and can thus be applied for automated defect detection.

Keywords: Optical quality control, Automated defect detection, Defect visualization, ML-based visual inspection system, Decorated foil plates.

1. Introduction

In the manufacturing industry, optical quality inspection is of great importance, as even the smallest defects in high-quality components lead to customer complaints. In order to meet the quality demands of premium manufacturers, every single part produced is visually inspected over its entire surface. Manual inspections require an enormous number of human resources and result in quality fluctuations and unnecessary rejects due to the monotonous work and the subjective assessment of the operator [1]. In order to improve competitiveness, it is necessary to automate these processes with the help of ML-based visual inspection systems [2, 3], increasing the efficiency of production lines and subsequently reducing expenditures for quality control and preserving production resources.

Particularly in the production of decorative foil plates, optical quality control is an enormous challenge due to the diverse product portfolio. The variety of product designs as well as the sequence of production steps lead to a wide range of different defect causes and defect classes. In order to visualize the variety of different defect classes with optimal contrast, customized optical modalities are required [4]. An inspection system for automated full-surface defect detection of decorated foil plates, under consideration of the allowed design and product variability and adaptability to new designs, is not available so far. In the following, an inspection approach consisting of three different optical modalities is presented, which allows automated full-surface defect detection under the demanded criteria, like applicability in the production line and visualization ability of production-related defect classes.

1.1. Abbreviations and Acronyms

LSM-1...Line Scan Modality 1, LSM-2...Line Scan Modality 2, ASM...Area Scan Modality.

2. Methods

2.1. Decorated Foil Plate and Defect Classes

Decorated foil plates consist of multiple printed layers, deposited on plastic-carrier foils that assemble to complex designs. Due to different production steps, a variety of defect classes emerge, as briefly listed below: print errors, inclusions, mechanical deformations, scratches, smears, pinholes, dust and other product-specific defects. Generally, the defects appear small in relation to the product dimensions, which presents a major challenge for the design of inspection modalities and data processing.

2.2. Inspection Approach

As shown in Fig. 1, the inspection approach consists of three different optical modalities: line scan camera in dark field illumination (LSM-1), line scan camera in backlight illumination for transmission measurements (LSM-2) and area scan camera in bright field illumination (ASM). Line scan cameras in combination with high intensity line bars are generally the appropriate choice for the dynamic inspection of flat surfaces, as they are capable of capturing high resolution images at high measurement speeds, regardless of the sample size in the transport direction. However, the experiments conducted with the line

scan camera revealed that the detection of defects on the transparent top layer was not satisfactory. Therefore, an optical modality, consisting of an area scan camera and a light bar aligned in direct reflection, was designed. A complete inspection is performed by a sequential run of the sample through all three optical modalities via a transport system, considering the demanded measurement cycle. The lower limit of the measurement cycle is thereby determined by the required exposure times, which ensure the highest possible contrast of relevant defect classes. A trigger signal (Acquisition Control Block) is mandatory to enable undistorted image acquisition, especially in line scan applications. Subsequent to triggered image acquisition, overlapping image patches of size 256×256 px are extracted and processed by means of a CNN (trained in a supervised manner), enabling patch-wise defect detection (Inference / Evaluation Block). Finally, the inference results are merged to full image size for enhanced defect evaluation interpretability, and shared with the production line (Production Line Interface Block). The subimages in Fig. 2 were captured with commercial 16 K line, 2.2 MP area scan cameras (RGB, monochrome) and high intensity LED light bars (white, 5624 K) at contrast optimized exposure times (100 µs to 500 µs) using above-described modalities.

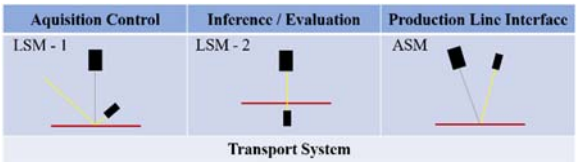


Fig. 1. Procedure of the inspection approach.

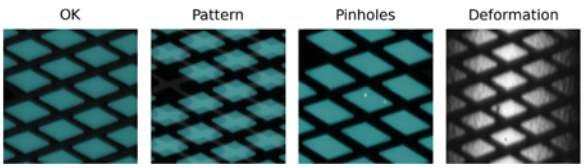


Fig. 2. Selected defect classes of optical modalities: LSM-1, LSM-2, ASM.

3. Results

3.1. Defect Visibility

Fig. 2 shows patches of preprocessed (256×256 px) images of selected defects. Subimage “OK” displays a defect-free patch of the sample and, like subimage “Pattern”, was acquired with LSM-1. The production-related misalignment of one or more layers leads to a large-area displacement of the original pattern. Bright field images of the ASM differ significantly from dark field images (OK, Pattern) and reveal defects in the transparent top layer, such as the depicted punctual mechanical deformations in subimage “Deformation”. The majority of defects in the transparent top layer only appear in the ASM. An example of pinhole defects captured with LSM-2 is shown in the subimage “Pinholes”. In addition, with the help of this setup, it is possible to detect inhomogeneities in semitransparent layers. Utilizing all 3 modalities it was feasible to visualize the required production-related defect classes with sufficient contrast within the range of a few pixels in extension and an object pixel size ranging from approx. 70 µm to 150 µm.

4. Conclusions

The developed inspection modalities enable the visualization of the demanded production-related defect classes with sufficient contrast, and can thus be applied for automated defect detection. The possible integration of the inspection approach into the production line and the assignment of detected defect classes to individual production steps also provides the ability to identify production deviations at an early stage and to intervene in a corrective manner.

Acknowledgements

The research work was performed within the COMET-project „Deep on-line learning for highly adaptable polymer surface inspection systems” (project-no.: 879785) at the Polymer Competence Center Leoben GmbH (PCCL, Austria) within the framework of the COMET-program of the Federal Ministry for Climate Action, Environment, Energy, Mobility, Innovation and Technology and the Federal Ministry for Digital and Economic Affairs. The PCCL is funded by the Austrian Government and the State Governments of Styria, Lower Austria and Upper Austria

References

[1]. J. E. See, C. G. Drury, A. Speed, et. al., The role of visual inspection in the 21st century, *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 61, Issue 1, September 2017, pp. 262-266.

[2]. Z. Ren, F. Fang, N. Yan, et al., State of the art in defect detection based on machine vision, *International Journal of Precision Engineering and Manufacturing-Green Technology,* Vol. 9, Issue 2, 2022, pp. 661-691.

[3]. A. A. R. M. A. Ebayyeh, A. Mousavi, A review and analysis of automatic optical inspection and quality monitoring methods in electronics industry, *IEEE Access*, Vol. 8, 2020, pp. 183192-183271.

[4]. B. Yoon, H. Lee, J. Jeong, Improved U-Net++ with patch split for micro-defect inspection in silk screen printing, *Appl. Sci.*, Vol. 12, 2022, 4679.

(2914)

What a Wasted Municipal Solid Waste's Data an Innovative Data Analytics Driven Strategic Framework for Future MSW Decision Making

R. Hussein
Syracuse, NY, USA 13135
Tel.: + 1-315-695-2340
E-mail: ezpscg@gmail.com

Summary: For the first time, this paper coins the concept of data-driven municipal solid waste decision making (DDMSWD). We aim to achieve goals such as reduction, reuse, recycling, recovery, and disposal of waste for the preservation of natural and built environments, including energy, climate, and water and soil resources. Our original concept draws the line between DDMSWD and traditional sustainable materials management. The latter term refers to matters that are destined for landfill or municipal combustors.

The municipal solid waste (MSW) industry (MSWI) collects massive amounts of complex and unpredictable waste-related data to analyze and, accordingly, make management and operational decisions. This paper debates some strategic issues of the MSWI's data-related practices, such as the following:

1. The modern world is rapidly changing, with an unpredictable trajectory. Who could have imagined the recent international waste-related new policies, epidemics, conflicts, and the state of the economy, just to name some examples? At the national [1], state [2], and county [3] levels, the US Centers for Disease Control and Prevention (CDC) collected the COVID-19 data and aggregated it on a weekly basis because of the unpredictability issue. The logical question is thus: can the massive amounts of seasonably collected MSW data from the modern world be relied upon for making long-term strategies? Evidently, the approach of the MSWI to data mining needs to be insightfully revisited starting from its roots, i.e., data, to where it is nowadays.

2. The previously mentioned collected MSW data and generated quantitative judgments suggest the MSWI is paying attention to "understanding" various facets of waste such as composition, characterization, rates, generation, and differences. On the one hand, the invaluable treasure of outcomes obtained from the MSW collected data [4, 7 to 13] clearly shows the extraordinarily complex and diverse inherited characteristics of the MSW data. On the other hand, the same treasure shows the use of simple numerical conventional tools to analyze that kind of data. The logical question is: by putting the two hands together, what is the significance of the statistics disseminated for the industrial SMW practices? In addition, in today's economically stressed MSWI [4] and a world that runs on live streaming technologies, can the MSWI afford to continue doing business-as-usual?

3. The same treasure also suggests the MSWI is paying attention to trends in waste-related goals, e.g., zero waste, efficient waste collection, waste reduction, recycling, and resource recovery. The logical question is: how can the goals of the US's Federal Resource Conservation Act (RCRA) [5] and New York State Municipal Solid Waste (MSW) [6, 7, 8] frameworks, plans, acts, and other schemes be achieved using simple conventional numerical data manipulation?

There are other relevant questions than the above to ask and observations to note. However, Albert Einstein once said, "In the middle of difficulty lies an opportunity." The MSW can not predict the unpredictable, but it must always be prepared. Whether the collected data and generated statistics indicate a favorable trend [7 to 12] or not, the MSWI must adapt its analytic system to today's world and technologies rather than a world that no longer exists. This vision ensures that viable decisions are made in response to the unprecedented challenges as they occur.

A key objective of this paper is to introduce an overall review of the futuristic concept of data-analytics-driven robust DDMSWD to the MSWI, in general, to benefit from the corporations that have been on the data-driven technologies path and to revamp its practices. It is worth noting that this paper interchangeably used terms such as management and operation; MSW industry and MSW organizations; waste and municipal waste; and strategy, plan, and framework. Also, this paper refers to the State of New York's MSWI as just an illustrative example, though it has every right to be proud of its leadership.

Keywords: Data analytics, Data science, Municipal waste, Waste management, Resource recovery, Recycling, STEM.

1. Introduction

The New York State Department of Environmental Conservation's (NYSDEC) "Beyond Waste" plan [8] instituted seventeen strategic goals for "sustainable materials management." The Onondaga County Resource Recovery Agent (OCRRA) recently disseminated a treasure of valuable statistics based on the collected data but apparently using basic conventional analytic tools [9 to 13]. Because at the time of developing those statistics, no one could even imagine the state of modern data analytics (DA). One may wonder about the undetected facts in those meritorious reports. This observation calls for urgent corrective actions to be taken to adjoin data-related waste practices to almost 60% of the top corporations in the USA that run on DA for viability and sustainability. The list of names is long and includes [14, 15]. Google, IBM, Facebook, Instagram, Apple, Amazon, Starbucks, Tesla, Nintendo, Netflix, Bristol Myers-Squibb, McDonald, just to name a few selected examples. Because municipalities strive to optimize extracting values from waste, identifying areas for improvement in current frameworks will help the

MSWI achieve its goals. The MSWI wouldn't have lagged other corporations if it weren't for its use of simple conventional analytic methods. Hasn't the time come for the MSWI to revamp its data-related tools?

The MSWI presumably recognizes the benefits of data-driven decision procedures but seemingly does not currently have a known strategy, or framework, or plan to implement the most recognized modern technologies, i.e., DA. The available literature [9 to 13] suggests sufficiently waste-related large data sets in quantity, variety, and complexity, yet their quantification methods are detached from modern practices. The MSW authority needs to focus on the first step any organization takes, i.e., analysis of data, before considering further steps such as minimization or maximization, increasing or decreasing, creating or advancing, engaging, or fostering the identified outcomes. As previously stated, this paper brings a fresh set of eyes to the MSWI in general. The following sections explore various facets of DA, challenges, and opportunities for implementing the technologies in the MSWI, which has a long way to catch up.

2. What is Data Analytics?

Data analytics encompasses the qualitative and quantitative process of analyzing complex data sets for the extraction of insights. DA is an all-in-one tool rather than a one-size-fits-all tool. Without doubt, it has revolutionized almost all other industrial fields as well. Volume, variety, and velocity are the key features that distinguish data sets. Data with a high velocity indicates near real-time collection and integration of information, and data with a high variety implies heterogeneity in the structure and type of data stored. So, it is no surprise that these features render basic conventional data analysis techniques inappropriate for modern challenges. As previously mentioned, the viability of the top corporations evidently depends on DA's wading through the data tsunamis. To this end, the following two common models are convincingly adequate for DDMSWD:

1). Predictive analytics to identify trends and predictions about waste generation.

2). Performance analytics to assess detailed performance metrics.

The DA techniques rest on mathematics and statistics, computer and information science, machine learning, and other areas. It is worth noting that the applications of DA go much further and deeper than just DDMSWD.

3. Innovative Data Analytics Driven Integrated Framework for MSWI

All the available MSW data-related statistics suggest using the massive amounts of collected data and basic conventional methods for analysis of the overly complex pervasive data. Now that the world has explored and validated DA's powerful capabilities, the time has come to deviate from the business-as-usual approach to modernize, if not revolutionize, industrial data-related practices. Our revamping vision is simple: don't reinvent the wheel, but reforge it. Today's world and challenges are different from those that existed even a few years ago. For example, data analytics technologies were not as common in everyday industrial applications as they are nowadays. Via his demonstrated national and international academic and industrial research and development, including advanced computer development [16 to 22], this author suggests reforging the existing MSWI starting with capacity building, including the following

1). The development and dissemination of guidelines on how the MSWI can benefit from DA.

2). Training sessions for those already in the workforce to revamp their skills, such as coding.

3). Encourage and assist the schoolers who are the seeds of the next generation of industrial leaders to consider DA for acquiring life-long industrial credentials. The New York Times, the Washington Post, the Guardian, and other sources attested to the importance of data science and analytics for tackling modern challenges. In this regard, data science has massive data in common with DA. Data analytics is essentially a special application of data science in which the datasets are massive and require overcoming logistical challenges to deal with them. STEM stands for Science, Technology, Engineering, and Mathematics, which constitute the data analytics foundation. We envisioned [22] a trifold DA-STEM-MSW incubator for developing workforce capable of converting massive MSW data into practical benefits. Because of its widespread application in DA, Python is an appropriate coding language for the trifold approach. Python has many advantages, including being open-source and free, being easy to intuitively learn, having an excellent online community, integrating well with other packages, and being faster than similar tools such as R and Matlab. In addition, other initiatives should consider internships, tuition reimbursement, and collaborative efforts with schools on hands-on learning.

4). Outreach to the forward-thinking stakeholders who are concerned about the impact of waste on the future of the MSWI.

4. Conclusions

Humans live in a material world, and there will be waste as long as they live. How humans use materials is vital to the sustainability of our planet. Thus, sustainable use of materials and municipal waste are inseparable. Data and waste are two faces of the same coin. We need waste to generate data and need data to manage the waste. In a world that is rapidly changing with an unpredictable trajectory, it is evidently clear that the MSWI needs modern data-driven decision systems. This paper argued the case of why modern data analytics technologies for the MSWI are the

central pole for any DDMSWD future framework. The paper innovatively conceptualized a viable modernization agenda for DDMSWD in the context of MSWI. Whether or not the massive amounts of collected and analyzed waste data indicate improvement, the methods used in today' MSWI for data analysis are basic and need to be revamped. As Albert Einstein once said, you cannot keep doing the same thing and expect a different result. Data analytics provides guaranteed success in revamping the MSWI, as it did for the most successful corporations in the world. To this end, this paper presented an innovative framework to explore uncharted terrain in the MSWI.

References

[1]. Centers for Disease Control and Prevention, Trends in Number of COVID-19 Cases and Deaths in the US, (https://covid.cdc.gov/covid-data-tracker/#trends_dailycases)

[2]. Centers for Disease Control and Prevention, Trends in Number of COVID-19 Cases and Deaths in NYS, (https://covid.cdc.gov/covid-data-tracker/#county-view?list_select_state=New+York&data-type=CommunityLevels)

[3]. Centers for Disease Control and Prevention, Trends in Number of COVID-19 Cases and Deaths in NYS, (https://www.cdc.gov/coronavirus/2019-ncov/index.html?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fcoronavirus%2Findex.html)

[4]. Onondaga County Resource Recovery Agency (OCRRA), 2019 Annual Report on Recyclables Recovered, (https://ocrra.org/wp-content/uploads/1/2020/03/recycling_2019.pdf)

[5]. US Environmental Protection Agency (EPA), Resource Conservation and Recovery Act Laws and Regulations, (https://www.epa.gov/rcra), 1990.

[6]. New York Department of Environmental Conservation, Municipal Solid Waste (MSW) Landfills in New York State, (https://www.dec.ny.gov/chemical/23682.html).

[7]. Onondaga County Resource Recovery Agency (OCRRA), Onondaga County Comprehensive Solid Waste Management Plan Update for 2021-2022, (https://ocrra.org/wp-content/uploads/1/2022/03/CSWMP-Update-2021-COMPILED-FINAL-Rev1-reduced.pdf).

[8]. New York State Department of Environmental Conservation (NYSDEC), Beyond Waste - A Sustainable Materials Management Strategy for New York State, (https://www.dec.ny.gov/docs/materials_minerals_pdf/frptbeyondwaste.pdf)

[9]. Onondaga County Resource Recovery Agency (OCRRA), 2019 Waste Characterization Study, (https://ocrra.org/wp-content/uploads/1/2020/05/WasteQC_2019_Final_KRL_050720.pdf), 2020.

[10]. Onondaga County Resource Recovery Agency (OCRRA), 2020 Annual Report on Recyclables Recovered, (https://ocrra.org/wp-content/uploads/1/2021/04/recycling_2020.pdf,) 2021.

[11]. Onondaga County Resource Recovery Agency, Comprehensive Recycling and Market Analysis, (https://ocrra.org/wp-content/uploads/1/2020/08/CRA_2019_082420.pdf)

[12]. Onondaga County Resource Recovery Agency, Compliance Report for 2019 – 2020 Plan Update for 2021 – 2022, (https://ocrra.org/wp-content/uploads/1/2022/03/CSWMP-Update-2021-COMPILED-FINAL-Rev1-reduced.pdf), 2021.

[13]. Onondaga County Resource Recovery Agency (OCRRA), 2020 Annual report, (https://ocrra.org/wp-content/uploads/1/2021/03/AnnualReport_2020_KRL_031921-1.pdf,) 2021.

[14]. Will Hillier, 7 Interesting Companies Using Data Analytics, (https://careerfoundry.com/en/blog/data-analytics/data-analytics-companies/), 2021.

[15]. Riya KumariMar, 10 Companies that Uses Big Data,(https://www.analyticssteps.com/blogs/companies-uses-big-data), 2021.

[16]. Hussein, R., An Analytic Solution for the Mechanics Performance of Packaging Corrugated Cardboards Containers with Non-rigid Bonding and Under Edgewise Loads, *Journal of Packaging Technology and Research,* Vol. 4, 2020, pp. 85-94.

[17]. Hussein, R, Be Kind and Keep no Waste Left Behind in Mind, *Presentation to Onondaga County Resource Recovery Agent' leadership*, 2022.

[18]. Hussein, R., Sustainable Municipal Solid Waste Management Data Analytics Driven Perspective, *Acta Scientific Computer Sciences,* Vol. 4, Issue 10, 2022, https://actascientific.com/ASCS/ASCS-04-0335.php

[19]. Hussein, R., AI-based Tools for Performance and Monitoring of Sustainable Built and Natural Environments, and the Climate, in Advances in Artificial Intelligence, Vol. 1 (Ed. Yurish, S.), *IFSA Publishing*, Chapter 7, 2019, pp. 185- 203.

[20]. Hussein, R., Knowledge-based Tools for Monitoring and Management, and Design of the Engineered Infrastructure Construction Systems, in Advances in Computers and Software Engineering, Vol. 2 (Ed. Yurish, S.), *IFSA Publishing*, Chapter 5, 2019, pp. 199- 251.

[21]. Hussein, R., Treatise on Sustainable Infrastructure Construction: Green Composites, Cross Laminated/Mass Timber, Wood Truss Connectors, Nondestructive Technologies, Health Assessment and Monitoring: Utility Poles and Geofoam, in Advances and Technologies in Building Construction and Structural Analysis, (Ed. Alireza Kaboli), Chapter 1, 2021, pp. 122-135.

[22]. Hussein, R., A novel python-based threefold data science, stem, climate change framework, in *Proceedings of the 3rd International Conference on Advances in Signal Processing and Artificial Intelligence,* Portugal, pp. 28-30.

(3003)

Attention-based Stacked Long-short Term Memory Network for Displacement Time History Prediction of Reinforced Concrete Bridge Piers Using Experimental Real-time Hybrid Simulation

O. Yazdanpanah, M. Chang and E. Ali Bakhshi
Department of Civil and Environmental Engineering, Myongji University, Yongin-si, Republic of Korea
Tel.: +82-31-330-6406, fax: +82-31-330-9705
E-mails: omidyazdanpanah@mju.ac.kr, cmw321@mju.ac.kr, nc.alibakhshi@gmail.com

Summary: The recursive long short-term memory (LSTM) network is utilized as a powerful deep learning algorithm to predict the displacement time histories of reinforced concrete bridge piers using real-time hybrid simulation. The proposed stacked LSTM network includes multiple uncertain input variables, horizontal and vertical ground motions, horizontal and vertical actuator loads, the effective height of the bridge pier, the moment of inertia, and the mass assumed for hybrid simulation. The functional application programming interface in the Keras Python library is used to develop a complex model with the above multiple inputs. Bidirectional LSTM and attention layers are also hired in the proposed model to accurately obtain the peak displacement of the piers. The database includes 12 experimental hybrid simulation tests for training, validation, and test dataset. The results reveal that there is a reasonable correlation coefficient, more than 80 %, between the predicted and the reference values of the displacement time series for all the datasets.

Keywords: Stacked long short-term memory network, Bidirectional layer, Attention layer, Reinforced concrete bridge piers, Experimental real-time hybrid simulation, Estimated displacement time histories.

1. Introduction

Real-time hybrid simulation (RTHS) is an effective economic method to experimentally evaluate the seismic responses of the critical load-rate-dependent components of complex structures subjected to earthquake excitations. One of the most important pros of the RTHS is that the physical test setup is required only for the critical substructure and the rests are developed numerically, which enables to achieve a significant cost reduction compared to full-scale experiments. Finally, the experimental result, the restoring force, is combined with the computer simulation in real-time to solve the equation of motion and obtain the global seismic response of the original complex structure [1-5].

To get much more insight into the substructure's behavior through RTHS under earthquake excitations, a lot of experimental specimens should be conducted and tested considering various uncertain parameters. This results in an expensive task in both the experimental setups and time. An effective solution to mitigate this issue can be the accurate prediction of the time history responses using surrogate models.

Numerous investigations have been assigned on the dynamic response prediction of civil infrastructures subjected to ground motions using autoregressive integrated moving average (ARIMA) with or without exogenous (X) input [6], symbolic and Bayesian regressions [7], multi-layer perceptron (MLP) network as the most famous artificial neural network (ANN) [8], and 1D convolutional neural network (CNN) [9]. The drawbacks addressed to ARIMA are its linearity and stationary dynamics. In addition, MLP, due to simple architecture, and CNN result in poor accuracy, especially for the structures with high nonlinearity and large plastic deformations.

A recurrent neural network (RNN) solves the problem associated with large input features and has been verified using numerous applications. Nevertheless, a major shortcoming still remains regarding future forecasting in a very deep network, such that after a few hidden layers, the information disappears through the network due to a vanishing gradient. Therefore, this loss results in low accuracy when predicting time series data [10-12]. Long-short term memory network (LSTM) is a new development of RNN that saves the information of each hidden layer for future purposes. It is done by adding a cell state to store the information of the inputs [13].

LSTM network can be addressed as a reliable and robust deep learning algorithm to predict time histories in both the linear and nonlinear regions. Recently, Zhang et al. [14] exploited the LSTM network through a sequential application programming interface (API) to predict the buildings' story-based displacement time history subjected to various ground motions. They introduced a stacked scheme that reduces the length of ground motion time history to estimate the displacement time history. The advantage of the proposed approach is to save running time and more accurate prediction as well. More recently, Kundu et al. [15] implemented a stacked LSTM network using the functional API to predict the displacement and drift time histories of a bridge column and a 6-story building frame considering more uncertain input variables.

The objective of this paper is to hire a stacked LSTM network to accurately predict the displacement time history, especially the peak value, of bridge piers

tested physically using RTHS. Fig. 1 presents a flowchart of the step-by-step process of the proposed approach.

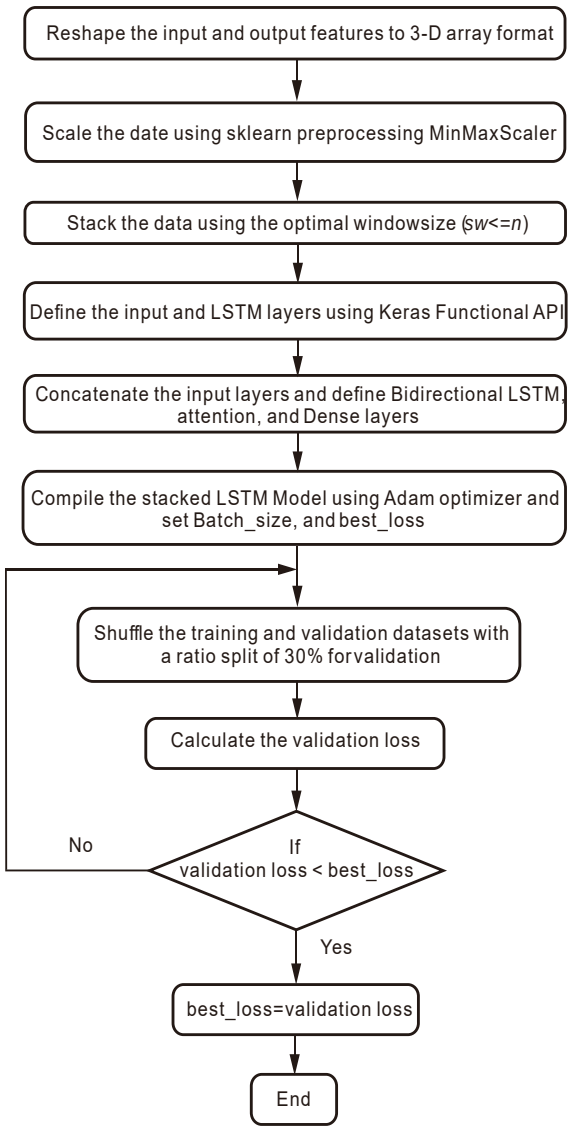
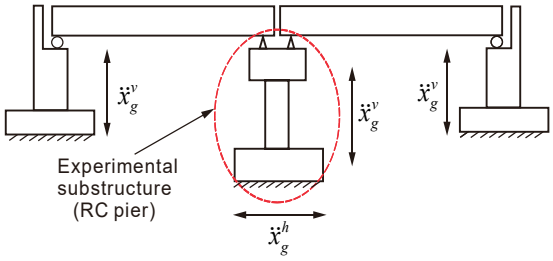


Fig. 1. Flowchart of the proposed approach used in this study.

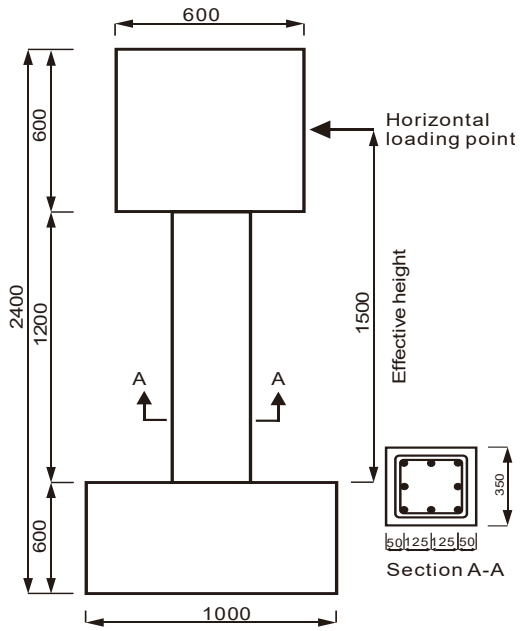
2. Experimental Setup

The experimental database includes a total of 12 reinforced concrete (RC) bridge piers conducted in detail, one test in 2017, six tests in 2018, and five tests in 2022 for RTHS. The tests include different horizontal and vertical ground motions with different seismic intensities, different horizontal and vertical actuator forces, different cross-sections, and mass assumed for RTHS. These characteristics are outlined in Table 1.

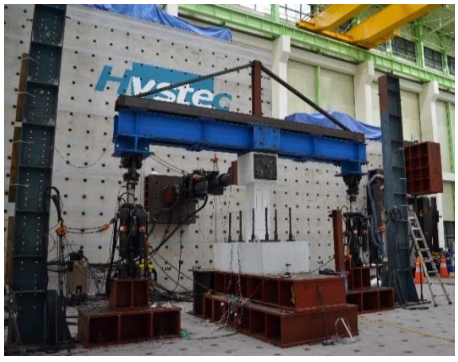
Fig. 2 shows a sketch of the experimental test setup of the RC pier test specimen of a 2-span bridge structure for RTHS. It is worth mentioning that the bridge deck and girders are analytically modeled.



(a) A two-span bridge structure for RTHS;



(b) Square section with vertical load (unit: mm);



(c) Experimental test setup for RTHS with vertical load (2022).

Fig. 2. Test setups constructed at the Hybrid Structural Testing Center (Hystec), Myongji University.

3. Long Short-term Memory (LSTM) Network

LSTM is a really big step for RNNs that makes the possibility of solving the struggles associated with long-term dependency. Therefore, it performs very well for time series data such as nonlinear displacement time history prediction under seismic excitations. A detailed description of how LSTM works can be found in [13] and [16].

Table 1. The characteristics of the test setups.

Specimen #	Test Method	Horizontal Ground Acc. (m/s^2)	Vertical Load (N)	Vertical Ground Acc. (m/s^2)	Effective Height (m)	Moment of Inertia (m^4)	Mass (kg)	Year
1	RTHS	GA_{h1} (0.1)	14.3×10^4	GA_v (0.1)	2	1.917×10^{-4}	1019.7	2018
2	RTHS	GA_{h1} (0.35)	14.3×10^4	GA_v (0.35)	2	1.917×10^{-4}	1019.7	2018
3	RTHS	GA_{h1} (0.5)	14.3×10^4	GA_v (0.5)	2	1.917×10^{-4}	1019.7	2018
4	RTHS	GA_{h1} (0.1)	0	0	2	1.917×10^{-4}	1019.7	2018
5	RTHS	GA_{h1} (0.35)	0	0	2	1.917×10^{-4}	1019.7	2018
6	RTHS	GA_{h1} (0.5)	0	0	2	1.917×10^{-4}	1019.7	2018
7	RTHS	GA_{h2} (1.0)	0	0	1.55	6.75×10^{-4}	6614	2017
8	RTHS	GA_{h1} (0.30)	39.2×10^4	0	1.50	1.25×10^{-3}	8×10^4	2022
9	RTHS	GA_{h1} (0.30)	39.2×10^4	GA_v (1.0)	1.50	1.25×10^{-3}	8×10^4	2022
10	RTHS	GA_{h1} (0.30)	0	0	1.50	1.25×10^{-3}	8×10^4	2022
11	RTHS	GA_{h1} (0.30)	39.2×10^4	GA_v (0.3)	1.50	1.25×10^{-3}	8×10^4	2022
12	RTHS	GA_{h1} (0.30)	39.2×10^4	0	1.50	1.25×10^{-3}	8×10^4	2022

3.1. Bidirectional LSTM

Despite remarkable results reported for LSTM, it may face an overfitting problem in some tasks. To avoid this shortcoming, there are some solutions. One of them is to go through bidirectional LSTM, training the input sequence in both forward and backward directions and concatenating both interpretations [17-19].

3.2. Attention Layer

The model can ignore a part of important data during the training process, particularly, when the dataset is huge and has long sequences, in sequence-to-sequence modeling. In such a case, adding an attention layer can improve the model performance and pay attention more to the momentous data [20-22].

The proposed stacked LSTM network presented in Fig. 3 includes two LSTM layers, a concatenation layer, a bidirectional LSTM layer, an attention layer, and two fully-connected dense layers, where GA, F_a, F_{v1}, F_{v2}, He, I, and M denote ground motion, horizontal actuator force, the right and left vertical actuators forces presented in Fig. 2(c), effective height (Fig. 2(b)), the moment of inertia, and the mass assumed for RTHS, respectively.

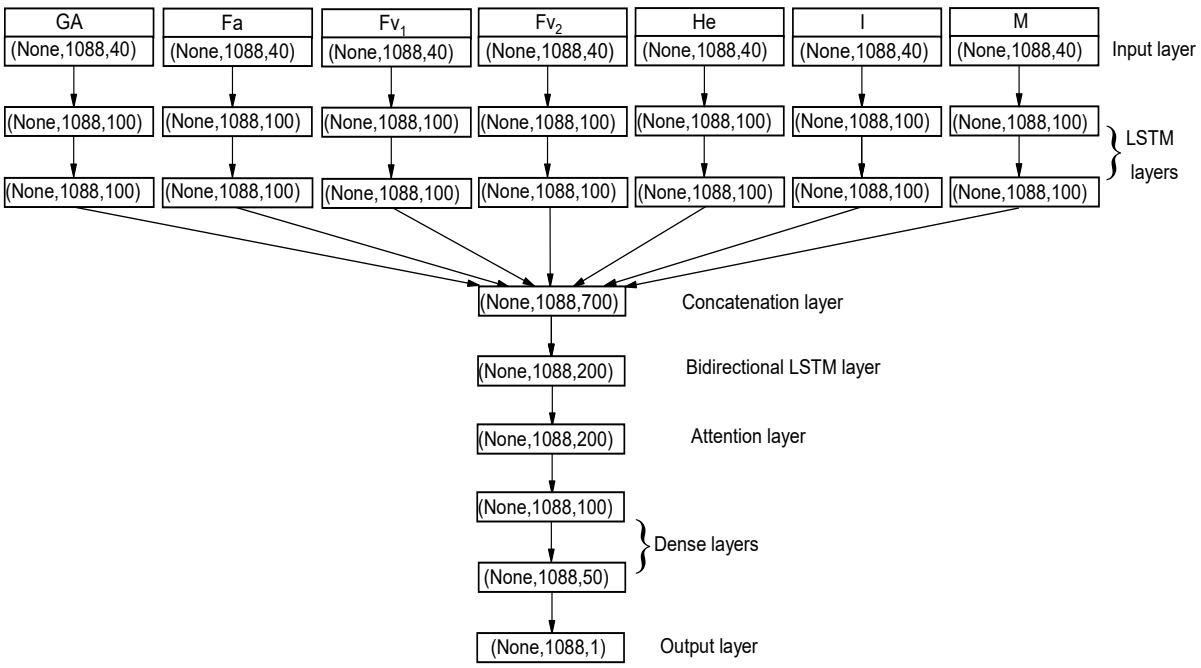


Fig. 3. The architecture of the proposed stacked LSTM with multiple inputs.

4. Results and Discussions

As can be observed in Fig. 4, both the training and validation losses are monotonically decreased and after 5000 epochs they reach zero approximately. It is worth mentioning that the datasets for the training and validation process are shuffled randomly with a percent of 30 for validation. Such that, the quota for the training is seven tests and three tests are considered for validation. All hyperparameter tuning with the LSTM network is based on GridSearchCV [23] implemented in Python package. Therefore, the following settings should be applied to the default parameters implemented in Python Keras LSTM library [24] as: units = 100, dropout = 0.2. The window size, in the stacked method, and batch size are set to be 40 and 10, respectively. Adam (Adaptive Momentum Estimation) is selected as the optimizer with learning_rate = 0.001, and decay = 0.005. Unites for Dense layers are set to be 100, and 50, respectively. Other parameters are assumed to be set as default. It is noteworthy that sklearn.preprocessing.MinMaxScaler of Python library is hired here to scale all the data in the range of [-1, 1] [25].

Performance comparison of the predicted and experimental displacement time histories is presented in Fig. 5 for all the training (5(a)-5(g)), validation (5(h)-5(j)), and test (5(k)-5(l)) datasets. The results show that there is good compatibility between the predicted and referenced displacement time histories. In addition, the correlation coefficients outlined in Table 2 confirmed this result.

Figs. 5(k) and 5(l) demonstrate that the error between the peak observed and predicted displacements extracted from the test dataset (unseen data) is less than 6 %.

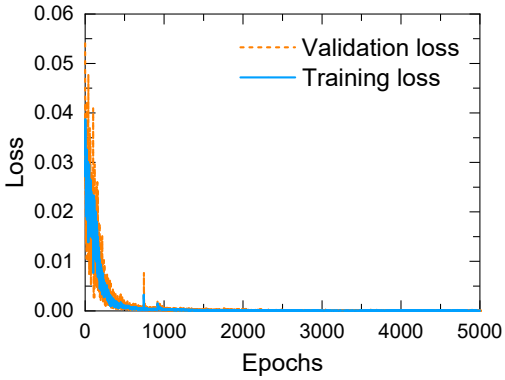
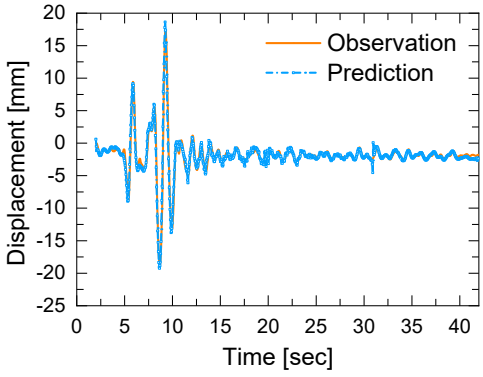
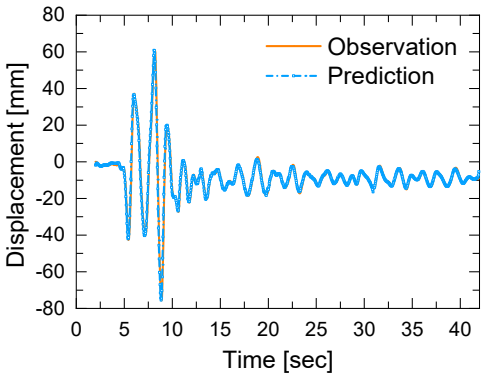


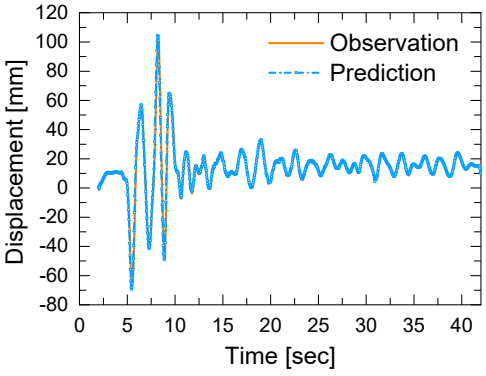
Fig. 4. Variation of training and validation losses.



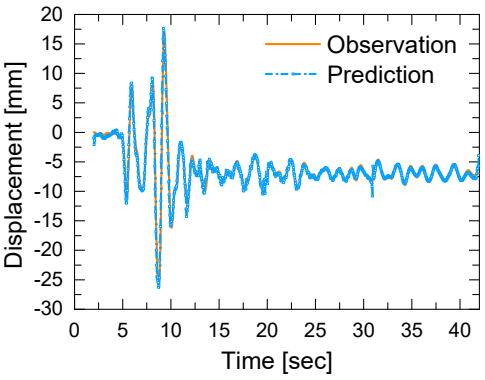
(a) Displacement time histories (Specimen 1)



(b) Displacement time histories (Specimen 2)

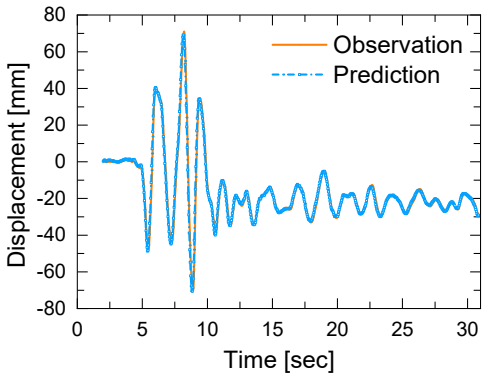


(c) Displacement time histories (Specimen 3)

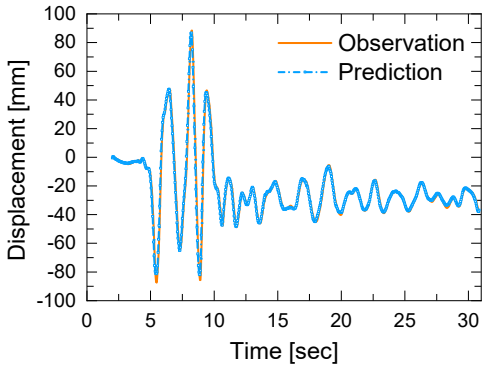


(d) Displacement time histories (Specimen 4)

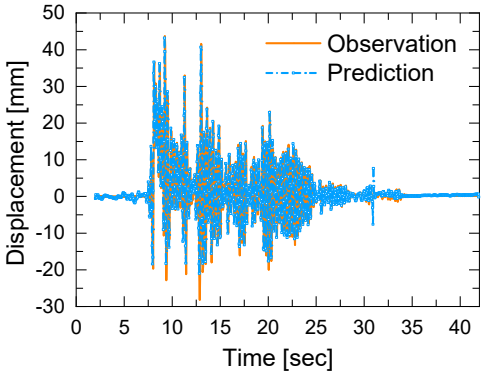
Fig. 5 (a-d). Performance comparison of the predicted and experimental displacement time histories.



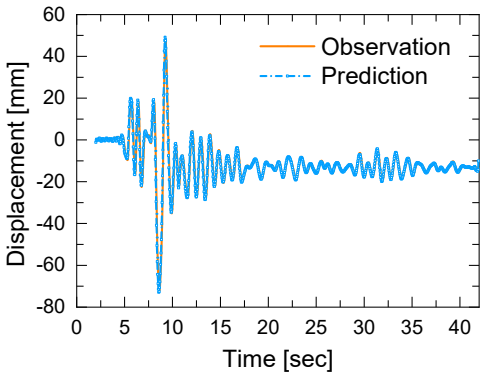
(e) Displacement time histories (Specimen 5)



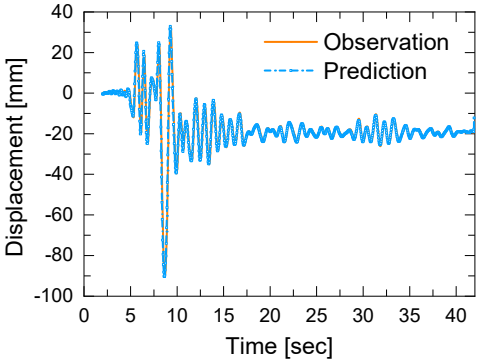
(f) Displacement time histories (Specimen 6)



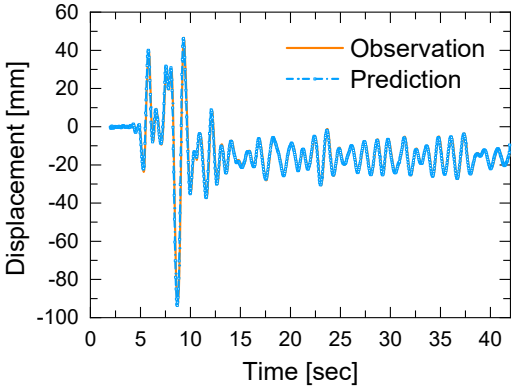
(g) Displacement time histories (Specimen 7)



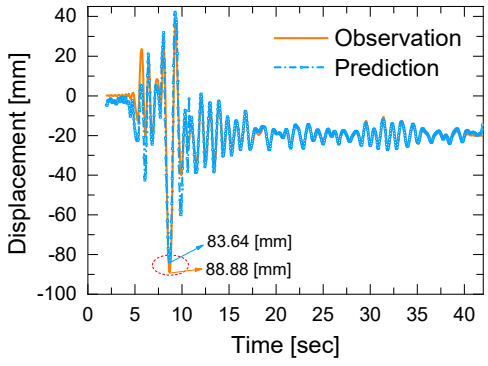
(h) Displacement time histories (Specimen 8)



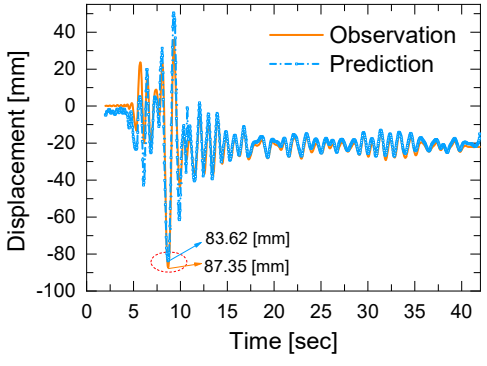
(i) Displacement time histories (Specimen 9)



(j) Displacement time histories (Specimen 10)



(k) Displacement time histories (Specimen 11)



(l) Displacement time histories (Specimen 12)

Fig. 5 (e-l). Performance comparison of the predicted and experimental displacement time histories.

5. Conclusions

The linear, nonlinear, and also peak values of displacement time histories of different bridge piers tested using RTHS are accurately predicted through a deep LSTM recurrent neural network considering multiple uncertain parameters. The proposed approach exploits a stacked LSTM network together with a bidirectional LSTM and two fully-connected dense layers. Keras functional API is used to model multiple inputs. 12 RTHS tests conducted in 2017, 2018, and 2022 are collected here to make a database for training, validation, and testing. The results confirm that the proposed approach is accurate, reliable, and straightforward and can be used in different engineering fields. It can be used to reduce the experimental cost due to manufacturing new specimens and test setups and saving time as well.

Table 2. Correlation coefficient for training, validation, and test datasets.

Specimen #	ρ
1	0.975
2	0.987
3	0.987
4	0.984
5	0.990
6	0.989
7	0.804
8	0.982
9	0.985
10	0.987
11	0.932
12	0.922

Acknowledgments

The authors gratefully acknowledge the financial support provided by Brain Pool program funded by the Ministry of Science and ICT through the National Research Foundation of Korea (2021H1D3A2A01095957).

References

[1]. Y. Chae, M. Park, C. Y. Kim, Y. S. Park, Experimental study on the rate-dependency of reinforced concrete structures using slow and real-time hybrid simulations, *Engineering Structures*, Vol. 132, 2017, pp. 648-658.

[2]. Y. Chae, J. Lee, M. Park, C. Y. Kim, Real-time hybrid simulation for an RC bridge pier subjected to both horizontal and vertical ground motions, *Earthquake Engineering & Structural Dynamics*, Vol. 47, Issue 7, 2018, pp. 1673-1679.

[3]. T. L. Karavasilis, J. M. Ricles, R. Sause, C. Chen, Experimental evaluation of the seismic performance of steel MRFs with compressed elastomer dampers using large-scale real-time hybrid simulation, *Engineering Structures*, Vol. 33, Issue 6, 2011, pp. 1859-1869.

[4]. Y. Chae, J. M. Ricles, R. Sause, Large-scale real-time hybrid simulation of a three-story steel frame building with magneto-rheological dampers, *Earthquake engineering & structural dynamics*, Vol. 43, Issue 13, 2014, pp. 1915-1933.

[5]. V. Saouma, G. Haussmann, D. H. Kang, W. Ghannoum. Real-time hybrid simulation of a nonductile reinforced concrete frame, *Journal of Structural Engineering*, Vol. 140, Issue 2, 2014, 04013059.

[6]. V. Ş. Ediger, S. Akar, ARIMA forecasting of primary energy demand by fuel in Turkey, *Energy Policy*, Vol. 35, Issue 3, 2007, pp. 1701-1708.

[7]. O. Yazdanpanah, K. M. Dolatshahi, O. Moammer, Earthquake-induced economic loss estimation of eccentrically braced frames through roof acceleration-based nonmodel approach, *Journal of Constructional Steel Research*, Vol. 187, 2021, 106888.

[8]. W. Ying, W. Chong, L. Hui, Z. Renda, Artificial neural network prediction for seismic response of bridge structure, in *Proceedings of the International Conference on Artificial Intelligence and Computational Intelligence* (*AICI'09*), Vol. 2, 2009, pp. 503-506.

[9]. P. Huang, Z. Chen, Deep learning for nonlinear seismic responses prediction of subway station, *Engineering Structures*, Vol. 244, 2021, 112735.

[10]. S. Hochreiter, Untersuchungen zu dynamischen neuronalen Netzen, *Diploma*, *Technische Universität München*, Vol. 91, Issue 1, 1991.

[11]. Y. Bengio., P. Simard, P. Frasconi, Learning long-term dependencies with gradient descent is difficult, *IEEE Transactions on Neural Networks*, Vol. 5, Issue 2, 1994, pp. 157-166.

[12]. S. Hochreiter, The vanishing gradient problem during learning recurrent neural nets and problem solutions, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, Vol. 6, Issue 2, 1998, pp. 107-116.

[13]. S. Hochreiter, J. Schmidhuber, Long short-term memory, *Neural Computation*, Vol. 9, Issue 8, 1997, pp. 1735-1780.

[14]. R. Zhang, Z. Chen, S. Chen, J. Zheng, O. Büyüköztürk, H. Sun, Deep long short-term memory networks for nonlinear structural seismic response prediction, *Computers & Structures*, Vol. 220, 2019, pp. 55-68.

[15]. A. Kundu, S. Ghosh, S. Chakraborty, A long short-term memory based deep learning algorithm for seismic response uncertainty quantification, *Probabilistic Engineering Mechanics*, Vol. 67, 2022, 103189.

[16]. Understanding LSTM Networks, http://colah.github.io/posts/2015-08-Understanding-LSTMs/

[17]. https://machinelearningmastery.com/develop-bidirectional-lstm-sequence-classification-python-keras/

[18]. How to Develop LSTM Models for Time Series Forecasting, https://machinelearningmastery.com/how-to-develop-lstm-models-for-time-series-forecasting/

[19]. Complete Guide To Bidirectional LSTM, https://analyticsindiamag.com/complete-guide-to-bidirectional-lstm-with-python-codes/

[20]. D. Bahdanau, K. Cho, Y. Bengio, Neural machine translation by jointly learning to align and translate, *arXiv Preprint*, 2014, arXiv 1409.0473.

[21]. A Comprehensive Guide to Attention Mechanism in Deep Learning for Everyone,

https://www.analyticsvidhya.com/blog/2019/11/comprehensive-guide-attention-mechanism-deep-learning/
[22]. https://machinelearningmastery.com/the-attention-mechanism-from-scratch/
[23]. How to Grid Search Hyperparameters for Deep Learning Models in Python with Keras, https://machinelearningmastery.com/grid-search-hyperparameters-deep-learning-models-python-keras/
[24]. LSTM Layer, https://keras.io/api/layers/recurrent_layers/lstm/
[25]. sklearn.preprocessing.MinMaxScaler, https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html

(3030)

Unscented Particle Filters for Estimation of Gene Expression Time Series Data

M. T. Noman, N. Amor and M. Petru
[1] Department of Machinery Construction, Institute for Nanomaterials,
Advanced Technologies and Innovation (CXI) Technical University of Liberec,
461 17, Studentska 1402/2 Liberec 1, Czech Republic.
E-mail: muhammad.tayyab.noman@tul.cz

Summary: This work addresses the problem of estimating gene expression using the unscented particle filter. The response of Gene Regulatory Networks (GRN) to functional requirements in the cell and environmental conditions evolve over time. Dynamic biological processes such as cancer progression and treatment recovery depend on the collected genetic profiles. These processes are behind genetic interactions that rewire over the course of time. The GRN is formulated as a nonlinear/ non-Gaussian state-space model. The particle filters (PF) represent the state-of-the-art in the optimal estimation for nonlinear/non-Gaussian dynamical systems. However, the PF is ineffective for high dimensional state-space systems. The unscented particle filters (PF-UKF) able to overcome the problem known as the 'curse of dimensionality'. Therefore, this work presents the PF-UKF to estimate the evolution of gene expression over time from series data considering that the GRN has a high dimensional space. Simulation results on real-world gene expression datasets, shows that the PF-UKF offers effective results compared to three other algorithms.

Keywords: Unscented Kalman filter, Particle filter, Gene regulatory network, Prediction, Bayesian estimation.

1. Introduction

The biological mechanisms that govern our development are complex and crucial to understand the cellular system. The gene regulatory network (GRN) controls the expression of thousands of proteins and genes in any specific cellular function. However, the biological processes are dynamic and evolve over time in response to extrinsic factors and various intrinsic, such as cellular development, targeted therapy disease progression and environmental conditions [1]. Understanding these gene regulatory networks can help us significantly enrich our knowledge of health and disease. In GRN, the state estimation is based on time series data representing by genome expressions [2], and it can be represented in multiple methods varying in their degree of sophistication [2]. In addition, the prediction of the evolution of gene expression is defined by a dynamic nonlinear state transition where it has a high dimensional state space model. Different approaches have been proposed to estimate the gene expression time series including, extended Kalman filter [1] and particle filtering [2].

Particle filters (PF) provides powerful estimation for nonlinear and non-Gaussian state-space scenario [3]. The main idea of PF is to estimate the state optimally by employing a group of random weighted particles. However, it is ineffective in high dimensional spaces where the number of particles needed increases super-exponentially with the dimension of the state [3].

Recently, various approaches have been emerged to ameliorate the performance of PF, where unscented Kalman filter (UKF) or extended Kalman filter (EKF) are used to draw the importance distribution [4]. These approaches called respectively, the Extended Particle Filter (PF-EKF) and Unscented Particle Filter (PF-UKF). It is having been confirmed in many applications that PF-UKF provide robust estimation comparing with PF-EKF [4]. Therefore, the main purpose of this work is to use the unscented particle filter to estimate the evolution of gene expression.

2. Unscented Kalman Filter

Consider a discrete state-space model given by:

$$x_k = f(x_{k-1}) + w_k, \tag{1}$$

$$y_k = h(x_{k-1}) + v_k, \tag{2}$$

where x_k is the state vector and y_k is the measurement vector. f_k and h_k are nonlinear state and measurement functions, respectively. w_k and v_k are respectively the system and measurement noises. From the Bayesian framework, the posterior distribution can be computed using the following prediction and update steps:

$$P(x_k \mid y_{1:k-1}) = \int P(x_k \mid x_{k-1}) P(x_{k-1} \mid y_{1:k-1})\, dx_{k-1}, \tag{3}$$

$$P(x_k \mid y_{1:k}) = \frac{P(y_k \mid x_k)\, P(x_k \mid y_{1:k-1})}{P(y_k \mid y_{1:k-1})} \tag{4}$$

For the nonlinear case, it is impossible to use Eqs. (3)-(4) because the introduced integrals are intractable. PF estimate the state using an ensemble of

weighted particles $\{x_k^{(i)}, w_k^{(i)}\}_{i=1}^N$ [3]. The approximation of the posterior distribution defined as:

$$P(x_k \mid y_{1:k}) = \sum_{i=1}^{N} w_k^{(i)} \delta\left(x_k - x_k^{(i)}\right), \quad (5)$$

where δ is the Dirac delta function.

Ideally, the particles need to be sampled from the true posterior, which is not available. Therefore, another distribution, referred to as the proposal distribution, $q(x_k / x_{k-1}, y_k)$ is used. The importance weights are given by:

$$w_k^{(i)} = w_{k-1}^{(i)} \frac{p(y_k \mid x_k^{(i)})\ p(x_k^{(i)} \mid x_{k-1}^{(i)})}{q(x_k^{(i)} \mid x_{k-1}^{(i)}, y_{1:k})} \quad (6)$$

Given the discrete approximation to the posterior distribution in Eq. (5), the mean of the state at time *k* is:

$$\hat{x}_k = \sum_{i=1}^{N} w_k^{(i)}\, x_k^{(i)} \quad (7)$$

After few iterations, the weights of particles in PF may are zero or close to zero. This phenomenon is known as 'degeneracy problem'. To address this problem, many approaches have been emerged to overcome this problem such as PF-EKF, PF-UKF and Markov Chain Monte Carlo particle filter (PMCMC). All these approaches based mainly on generating the particle from proposal distribution using Metropolis hasting (MH) or gips sampling (GS) in PMCMC, EKF in PF-EKF, and UKF in PF-UKF. It has been shown that the UKF provides a better proposal distribution than the prior transition in that it contains the most recent measurements, that often contain a lot of worthy information for estimating the states spaces [4]. Therefore, we used in this work the PF-UKF to estimate the evolution of gene series data.

3. Problem Formulation

The tracking of genes expressions using a measurement data is represented as:

$$x_k = A\, g_{k-1} + w_k, \quad (8)$$

where x_k represents the evolution of gene expression at time k. f represents a non-linear function that represents the regulatory relationship between several genes. h represents a nonlinear measurement function. y_k represents the microarray data. A represents the matrix coefficient values of the regulatory relationship between various genes. g_{k-1} represents a sigmoid squash function of the latent state estimate.

The main idea is to track the evolution of the gene expressions $x_{1:T}$ given an observations data $y_{1:T}$ over time. In this context, we proposed to use each PF-UKF to estimate the evolution of every gene over time. Furthermore, we have reduced the dimensions of the state space model and overcome the problem of 'curse of dimensionality'.

4. Simulation on Real Biological Data

We applied the unscented particle filter to predict the evolution of gene expression from real data. In this work, we used the real data for worm time series, which was presented in [5]. We considered also eight genes for this section. We tracked the eight genes expression using 1000 particles for PF, PMCMC, PF-UKF, and we implemented 100 Monte Carlo Runs. The coefficients of regulatory relationship that is matrix A are computed using normal distribution.

We computed the root mean square error (RMSE) between the real and estimated state of eight worm time series data. Then, we averaged the RMSE for eight genes: UKF = 0.0189, PF = 0.4521, PMCMC = 0.4799 and PF-UKF = 0.0019. Furthermore, PF-UKF offers the best results in terms RMSE while PF, and PMCMC result in large estimation error.

5. Conclusions

This work introduced the unscented particle filter for estimation of gene expression. The gene regulatory networks are modelled as a high dimensional problem. Simulation results on real-world data showed that the PF-UKF provided a robust estimation for the evolution of gene expression with less RMSE compared to other filters, i.e., UKF, PF and PMCMC. Furthermore, PF-UKF provides an alternative solution to the problem of estimation of gene expression time series data.

Acknowledgements

Financially supported by the Ministry of Education, Youth and Sports of the Czech Republic and the European Union (European Structural and Investment Funds – Operational Programme Research, Development, and Education) in the frames of the project "International Research Laboratories," Reg. No. CZ.02.2.69/0.0/0.0/18_054/0014685

References

[1]. A. Noor, E. Serpedin, M. Nounou, H. N. Nounou, Inferring gene regulatory networks via nonlinear state-space models and exploiting sparsity, *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, Vol. 9, Issue 4, 2012, pp. 1203-1211.

[2]. F. Bugallo, C. Tasdemir, P. M. Djuric, Estimation of gene expression by a bank of particle filters, in *Proceedings of the 23rd European Signal Processing Conference (EUSIPCO'15),* Nice, France, 31 August 2015 – 04 September 2015, pp. 494-498.

[3]. A. Doucet, A. M. Johansen, A tutorial on particle filtering and smoothing: Fifteen years later, *Handbook of Nonlinear Filtering*, Vol. 12, 2009, pp. 656-704.

[4]. S. J. Julier, J. K. Uhlmann, Unscented filtering and nonlinear estimation, *Proceedings of the IEEE*, Vol. 92, Issue 3, 2004, pp. 401-422.

[5]. M. Maduro, J. Rothman, Making worm guts: The gene regulatory network of the Caenorhabditis elegans endoderm, *Developmental Biology*, Vol. 246, 2002, pp. 68-85.

(3109)

Collective Obstacle Avoidance Strategy – An Agent-based Simulation Approach

J. Grosset [1,2], A.-J. Fougères [2], M. Djoko-Kouam [2,3] and J.-M. Bonnin [1]
[1] IMT Atantique, IRISA, Rennes, France
[2] ECAM Rennes, Louis de Broglie, Campus de Ker Lann, Bruz, Rennes 35091, France
[3] IETR, CentraleSupélec, Rennes, France
E-mail: juliette.grosset@ecam-rennes.fr

Summary: The context of factory 4.0 leads more and more to decentralised solutions, as centralisation shows its limits. One of the research areas of Industry 4.0 is the use of autonomous guided vehicles (*AGVs*), autonomous industrial vehicles (*AIVs*). We want to show that cooperation is useful and necessary to increase their autonomy. We propose in this paper an agent model to test scenarios in Industry 4.0 environments with a fleet of *AIVs*. In addition, we are interested in the resolution of global obstacle avoidance by *AIVs* with a collective strategy. The results of the simulation will be evaluated by performance indicators such as distance and time in order to compare the different proposed approaches.

Keywords: Agent modelling, Cooperative strategy, Simulation, Industry 4.0, Autonomous industrial vehicles, Obstacle avoidance.

1. Introduction

One of the fundamental axes of Industry 4.0 is the search for decentralised solutions in the decision-making process because centralisation quickly shows its limits [1]. Indeed, if an industry is centrally controlled, then if a technical failure, e.g., a network failure occurs, everything will come to a standstill. Decentralised solutions allow for more flexible control, which is preferable.

The use of autonomous guided vehicles (*AGVs*) or autonomous industrial vehicles (*AIVs*) is a strategic research area in the context of Factory 4.0 and Logistic 4.0. Currently, vehicles have autonomy over their navigation linked to rails, physical or virtual markers that allow them to get to their goal. They follow predetermined trajectories and they are autonomous in their navigation after having received an order from a central, a supervisor. If they detect a problem, or an obstacle they can report the information to the central, but it is the control centre that will decide for the other vehicles.

Our point of view to improve the autonomy of *AIVs* is to move towards decentralised solutions with collective strategies taken by the fleet of vehicles. To increase their intelligence and their individual and collective decision-making, they need to become more cooperative [2-4]. Thus, the AIVs will be more efficient in performing the five core tasks they have to control (task allocation, localisation, path planning, motion planning and vehicle management) [5].

Thus, they could carry out their missions with a collective strategy. This translates for example into a fleet of *AIVs* that communicate and share information. In the first level of strategy, there is the shared perception of their vision of the environment. A higher level that would achieve a collective global strategy for the realisation of missions would be the allocation, scheduling and distribution of tasks between them in real-time. The aim is for robots to be more efficient, waste less time and use less energy to accomplish all their tasks.

Our objective is to improve the *AIV* autonomy integrated into a fleet based on collective intelligent strategies. Among the problems to be solved to make *AIVs* more autonomous, we can particularly identify 1) the avoidance of collisions between vehicles or with static or dynamic obstacles [6, 7], and 2) the path planning or path finding when an obstacle hinders the passage of *AIVs* or obstructs a predefined path [8-10].

The problem of avoidance is an individual problem that is addressed in a more global and collective way. There are three phases: perception of the obstacle (detection), rerouting or trajectory planification (avoidance) and the overall strategy. In the following, we will deal with avoidance through the choice of the route, the path but not the trajectory. This chosen path can be costly when an aisle is completely obstructed. This can be solved by enhanced cooperation [11, 12], i.e., the ability of autonomous industrial vehicles (*AIVs*) to exchange relevant information with each other on their traffic conditions.

This is how we augmented the Bahnes' algorithm [11], in a previous work [13], to take into account obstacle avoidance, in addition to communication for avoidance at intersections.

Solutions are needed for the second problem, especially in the case where an obstacle completely obstructs a lane. This prevents a mission from not being completed because the main route defined by an *AIV* is compromised by an obstruction. Path planning or path-finding algorithms [9, 14] make it possible to re-plan their route, depending on whether the target of the mission has been reached or not.

In this paper, we will propose a global obstacle avoidance with shared perception. Indeed, we will not

use a local avoidance solution to improve autonomy. But rather a collaboration between the different vehicles of the fleet to move towards the basis of a global collective strategy in our future work. The proposed approach is adaptable and flexible for centralised or decentralised solutions. In other words, we are dealing with a distributed system where each robot decides by itself, augmented with communication between vehicles to share information about the environment. The communication may also go through a central supervisor before being redistributed.

The paper is organised as follows: Section 2 describes the implementation of a collective strategy consequent to the detection with agent modelling, the exchange of messages to cooperate and the different approaches we will compare. The description of the simulation environment, the implementation of the experiments and the results are analysed in Section 3. A discussion on the cooperative infrastructure and how fuzzy logic can help to qualify the obstacles are also made in Section 3. Finally, conclusions and future work are presented in Section 8.

2. Collective Strategy Consequent to the Detection

2.1. Modelling Agent

In order to validate the use of collective strategy algorithms to increase the autonomy of *AIVs*, we rely on a methodological framework consisting of 3 phases: (1) agent modelling, (2) verify in simulations the correct performance of the algorithm, and (3) experiment with TurtleBot3 Burger robots in perspective.

We propose an agent model in Fig. 1 that adapts to different situations and scenarios that are tested in the context of Industry 4.0.

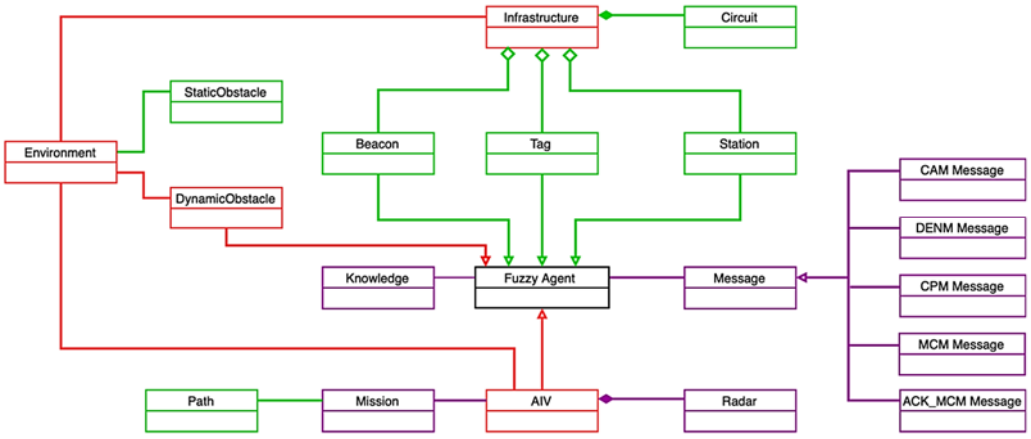


Fig. 1. Modelling agents adaptable to different scenarios for Industry 4.0.

Infrastructure is deployed in the environment and is composed of a traffic plan, and active elements such as beacons, tags and stations. Since they are active, these elements of infrastructure are modelled as software agents. These agents are defined as fuzzy agents because they can take fuzzy decisions from uncertain knowledge or perception, but these possibilities are not illustrated in this paper.

Industrial vehicles, which are also modelled as autonomous agents, carry out missions defined by routes and sent by a supervisor of the environment. They are equipped with radar and thanks to their knowledge of the environment they can avoid collisions with static obstacles (goods in the aisles), or dynamic obstacles (other vehicles or operators). *AIVs* can also cooperate by communicating together. We choose to use different types of standardised messages transposed from ETSI for ITS.

2.2. Exchanging Messages to Cooperate

The solutions proposed in the literature make it possible to solve an avoidance problem locally. In order to be able to carry out all the missions assigned to the vehicles, it is necessary that they cooperate and share information on their perception of the environment. Indeed, if an obstacle is perceived by a vehicle, the sending of a *Cooperation Perception Message* (CPM) (ETSI TR 103 562 standard [15]) makes it possible to warn and help the other vehicles to make decisions about their route. This means that if an unavoidable obstacle is perceived on its route, it can recalculate and plan a new route to be able to complete its task.

2.3. Obstacle Avoidance Approaches

We propose to study three approaches to obstacle avoidance by *AIVs* that will be tested in the simulation. Each robot has knowledge of the environment, i.e., the position of aisles, intersections, *AIVs* stations, and mission points. These various important traffic points of the circuit will be called nodes in the following, and are identified by respective numbers shown in Fig. 3.

In the first approach, agents do not have the ability to re-plan a route and change their route when faced with an obstacle.

In the second approach, vehicles are able to change routes and plan a new route when an obstacle is detected, but they do not cooperate with other vehicles. This means that vehicles will enter routes where obstacles are present and will have to change their route, even though another vehicle has already perceived this. They are able to calculate the path to their destination.

The path planning algorithm chosen for these experiments is Dijkstra. The objective is not to optimise a path planning algorithm (each *AIV* agent having to execute this algorithm), but to use an efficient algorithm to measure the performance of a collective strategy compared to an individual approach in carrying out missions in an industrial environment.

The last approach will be a collective strategy upon detection of an obstacle. This strategy is established according to two points of view:

- The vehicle detects an obstacle, which determines the level of obstruction. Then, it communicates the description (position, level of obstruction, type – static or dynamic) of the obstacle to the other vehicles. It can react according to its static or dynamic type to avoid it if possible, or it can replan another path to reach its objective and accomplish its mission. This algorithm is described in Fig. 2: Algorithm 1.
- The vehicle receives information about an obstacle in the environment. It first tries to determine whether the obstacle is on its path to accomplishing its mission. If this is the case and the obstacle is obstructing its planned path, it can replan its route. If the obstacle can be avoided, it can act and anticipate this static or dynamic obstacle by arriving in the vicinity of the received position of the obstacle. This improves the safety of the area and prevents accidents. This algorithm is described in Fig. 3: Algorithm 2.

Algorithm 1: AIV detects an obstacle ahead

```
Input : AIV attributes, Radar Function
        Communication to other AIVs
if AIV detects an obstacle then
    Determine level of obstruction of the obstacle
    Send message to other AIVs -
    (presence of an obstacle, position and level of obstruction)
    if Obstacle is static then
        if Possibility to bypass the static obstacle then
            Bypass the static obstacle
        else
            Dijkstra algorithm -
            calculate a new path to continue its mission
        end
    else
        Slowing down -
        increasing caution (dynamic obstacle: pedestrian, etc.)
        if Possibility to bypass the dynamic obstacle then
            Bypass the dynamic obstacle
        else
            Dijkstra algorithm -
            calculate a new path to continue its mission
        end
    end
end
```

Algorithm 2: AIV receives an information about an obstacle

```
Input : AIV attributes, Radar Function,
        Communication to other AIVs
if Obstacle is on its path then
    if Obstacle is static then
        if Possibility to bypass the static obstacle then
            Bypass the static obstacle
        else
            Dijkstra algorithm -
            calculate a new path to continue its mission.
        end
    else
        Slowing down - increasing caution
        (dynamic obstacle: pedestrian, etc.)
        if Possibility to bypass the dynamic obstacle then
            Bypass the dynamic obstacle
        else
            Dijkstra algorithm -
            calculate a new path to continue its mission
        end
    end
else
    Continue its path and its mission
end
```

Fig. 2. Collective intelligent strategy algorithm.

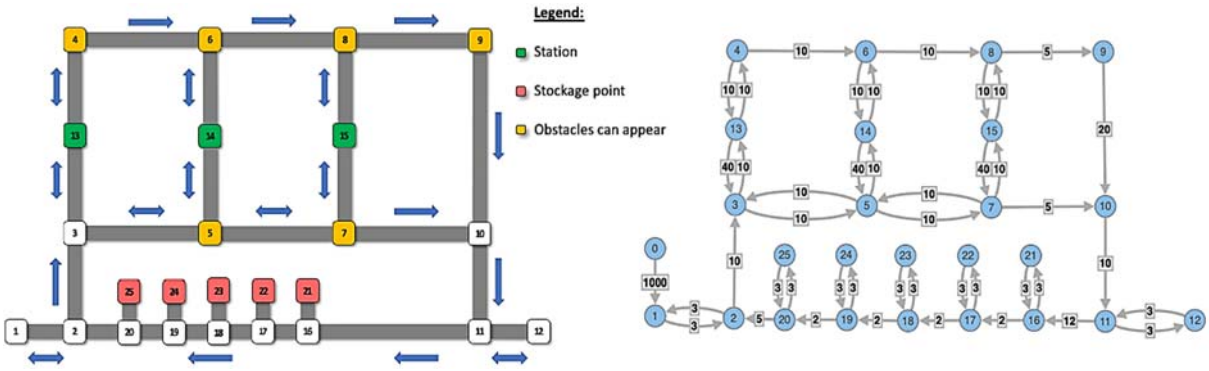


Fig. 3. Warehouse environment [16] with a) adapted direction of traffic, and b) it's oriented graph with costs.

3. Simulation Results and Discussion

3.1. Environment Simulation

The environment chosen to test the different approaches is the warehouse presented by Bechtsis *et al.* [16]. The directions of circulation have been adapted for our experiments and are identified in Fig. 3. We work with five *AIVs* corresponding to the five parking spaces available in the environment.

AIVs must perform missions in the environment shown in Fig. 3. A mission consists of travelling from its car park to a storage point and back to its car park. Thus, the *AIVs* applying the collective intelligent strategy have two route calculations to perform with the Dijkstra algorithm. The first one is to calculate the shortest path to the storage point. The second one is to calculate the shortest route back to their assigned parking space.

The costs in distance between the different nodes of the circuit have been chosen and applied to favour certain directions of traffic. For example, vehicles should go from node 14 to node 6, which has a cost of 10, rather than to node 5, which has a cost of 40. These costs between the different nodes are shown in the directed graph in Fig. 3. These costs are used by Dijkstra's algorithm to find the path that costs the least distance, and therefore the fastest.

To simulate these different experiments, we implemented a graphical application with different functionalities. The different classes presented in Fig. 1 have been implemented in Python from scratch. We use and develop this simulator for various experiments in the laboratory, and for teaching to engineering students. Its interface is shown in Fig. 4. In the central part, we have the representation of the warehouse presented by Bechtsis *et al.* [16], which we have adapted in Fig. 3. The different nodes present in Figs. 3 and 4 are represented by white squares in the interface. The vehicles are visualised by small coloured circles, and the obstacles are coloured black with different sizes corresponding to the obstruction levels in the aisle.

In the following, we will denominate the scenarios *sc1*, *sc2* and *sc3*. The missions will be named *mi* and the nodes *ni*.

The functionalities proposed by our application (Fig. 4) are to:

- Stop the simulation;
- Place obstacles randomly in terms of size T1 to T4 (T4 corresponding to a size that obstructs an entire aisle – see Fig. 4) and position at the six possible nodes shown in yellow in Fig. 3;
- Launch a scenario: the agents must then carry out missions given by the supervisor. That is, go from their parking place to one of the three green nodes in Fig. 3, before returning to their parking place. On the other hand, pressing the scenario 1, 2 or 3 button allows them to choose which strategy (*sc1*, *sc2* or *sc3*) they can perform.

For instance, in Fig. 4, an avoidable obstacle with a size *T2* in *n6* is present, and an obstacle obstructs the whole passage with a size *t4* in *n8*. The size of the obstacle in the interface corresponds to its obstruction level.

These different approaches were detailed in Subsection 3.3. *Sc1*, therefore, does not allow the agents to use Dijkstra's algorithm. *Sc2* does, and *sc3* allows agents to communicate with each other about the description of a perceived obstacle.

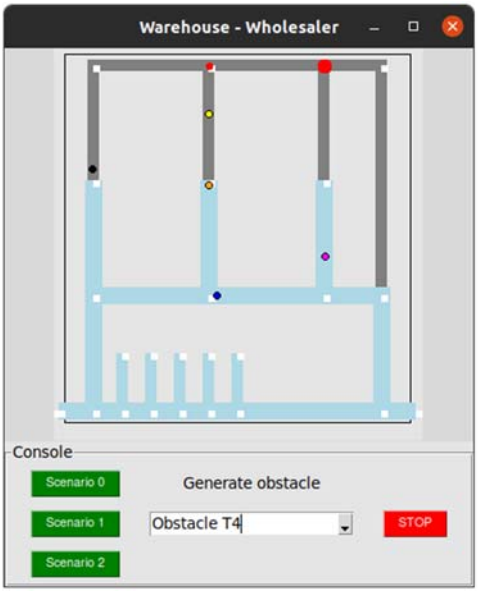


Fig. 4. Application interface to simulate missions in the warehouse environment.

3.2. Setting up the Experiments

The three approaches were tested in six different experiments. Each experiment corresponded to an obstacle completely obstructing an aisle on one of the six nodes: *n4*, *n5*, *n6*, *n7, n8* and *n9* shown in orange in Fig. 3. During these six experiments, the agents had to perform 1 of 3 missions consisting in going from the car park to one of the nodes *n13*, *n14* and *n15* randomly chosen by the supervisor, and then returning to their parking place. They were equipped with one of the three approach capabilities described in Subsection 3.3. That is, three experiments were performed with *sc1*, *sc2* and *sc3* with an obstruction at the same node *ni* to compare the approaches.

The performance indicators that were used to define the efficiency and performance of these approaches are:

- The validity of the set of missions performed by the five agents;
- The distance travelled to carry out the different missions;
- The overall time is taken to complete the various tasks.

Distance and time are performance indicators that were chosen because they are strongly related to the energy used by a robot, an *AIV*. Thus, this indicates the energy impact of the different approaches.

3.3. Experiments: Comparisons with Performance Indicators

Table 1 summarises the different results of the agents in the six experiments according to *sc1*, *sc2* or *sc3*. *Sc1* does not allow the agents to complete all the missions, which is explained by the fact that they are not able to recalculate their routes. Thus, if any agent was blocked by an obstruction on a node, it remained without the ability to react and the agents' overall mission failed.

Conversely, *sc2* and *sc3* applied by the agents allowed in both cases to fulfil all the missions thanks to their ability to adapt and to plan a new route (with the Dijkstra algorithm) when encountering an obstacle obstructing a path.

The distance and overall execution time to complete all the missions have therefore been quantified in Table 2 and Table 3 respectively.

Table 1. Completed missions in relation to the strategy number of agents.

Completed missions						
Obstacle node	N4	N5	N6	N7	N8	N9
Strategy number						
1	False	False	False	False	False	False
2	True	True	True	True	True	True
3	True	True	True	True	True	True

Table 2. Total distance travelled by agents in relation to the strategy number.

Total distance						
Obstacle node	N4	N5	N6	N7	N8	N9
Strategy number						
2	600	740	660	660	740	755
3	600	700	640	660	710	715

Table 3. Total execution time to carry out the missions in relation to the strategy number.

Total execution time						
Obstacle node	N4	N5	N6	N7	N8	N9
Strategy number						
2	489.75	677.56	521.23	559.59	582.97	567.52
3	482.14	662.82	456.60	542.04	466.35	470.40

The analysis that emerges is that *sc3* allows the agents to travel less distance in 4 out of 6 experiments and allows them to complete the missions more quickly in all six experiments.

The cooperative perception allows the agents to not go directly to the obstacle if it is on their way, but to redefine with the help of Dijkstra another path to reach their goal. Thus, logically, this collaboration takes precedence over the individual strategy, and communication in a fleet of *AIVs* thus increases their autonomy and their ability to adapt to new situations. Over the six experiments, there is a 3 % gain in overall distance and a 9 % gain in time for the whole fleet of *AIVs*. This cooperative perception is an essential basis for a future collective global strategy for the sharing and organisation of distributed missions within a fleet.

3.4. Cooperative Infrastructure

After, the different results mentioned in the previous section, we will discuss the possible cooperation of the infrastructure. Indeed, it could help the collective *AIVs* strategy with local information. This knowledge of the environment could be shared by active elements of the infrastructure such as cameras. This could help to qualify an obstacle or even detect an obstacle before a vehicle enters an area with an unavoidable obstacle.

3.5. Fuzzy Logic to Qualify Obstacles

In perspective, the *AIV* agents of our model presented in Fig. 1 may become fuzzy. Indeed, the agents will have incomplete, fragmented, fuzzy and uncertain knowledge in certain situations. The fuzzy logic could allow the agents to be able to recalculate or not their itinerary according to whether an obstacle seems avoidable or not, depending on the qualification of the disturbance caused by the obstacle.

4. Conclusions and Perspectives

In the context of Industry 4.0, mobile robots must become increasingly autonomous in order to avoid obstacles in an intelligent way. We are interested in the possible impact of cooperative perception between the vehicles of a fleet during a mission. Our aim is to move to a higher level of collaboration with the possibility of mission sharing and reorganisation in a distributed way.

The comparison of the three approaches evaluated using distance, execution time and whether or not a mission has been successfully achieved, showed that it is essential that vehicles have the ability to re-plan their routes while necessary. In addition, the last approach, which allowed agents to communicate the description of perceived obstacles, allowed for gains in distance and execution time of overall missions. Indeed, it allows *AIVs* to avoid traffic around an area where a static or dynamic obstacle (human operator, for instance) is present in an aisle. It also allows operators or other robots to be safe in that area: for example, to remove a static obstacle.

This very preliminary study already shows the interest of collaboration to increase the collective and individual efficiency of the vehicles in a fleet. It opens the door to more advanced global collective strategies

with the possibility of the allocation, scheduling and distribution of tasks between them in real-time after the perception of an obstacle. Furthermore, the tracking of obstacles by the robots, or assisted by the infrastructure, i.e., a cooperative perception, would also optimise the efficiency of the fleet in carrying out these missions.

The perspectives that emerge from this work are to increase the work of the analysis with other performance indicators, but especially more experiments. Indeed, we can define several missions in a row by robots with several obstacles obstructing the lanes at different places. Another strong perspective to our work is to experiment the collective strategy with the TurtleBot3 Burger robots.

In addition, the notions of cooperative infrastructure and the use of fuzzy logic to qualify obstacles are also planned to be further developed in future work.

Acknowledgements

The authors would like to thank the region Bretagne- France for funding the *VIASIC* project as part of the ARED-2021-2024 call for projects concerning the strategic innovation area: Economics for industry for intelligent production.

References

[1]. A. R. Boccella, P. Centobelli, R. Cerchione, T. Murino, R. Riedel, evaluating centralized and heterarchical control of smart manufacturing systems in the era of Industry 4.0, *Applied Sciences*, Vol. 10, Issue 3, 2020, 755.

[2]. U. Cao, A. S. Fukunaga, A. B. Kahng, Cooperative mobile robotics: Antecedents and directions, *Autonomous Robots*, Vol. 4, 1995, pp. 7-27.

[3]. Z. H. Ismail, N. Sariff, E. G. Hurtado, A survey and analysis of cooperative multi-agent robot systems, challenges and directions, in Applications of Mobile Robots, *IntechOpen*, 2018, pp. 8-14.

[4]. Y. Rizk, M. Awad, E. W. Tunstel, Cooperative heterogeneous multi-robot systems, A survey. *ACM Computing Surveys*, Vol. 52, Issue 2, 2019, pp. 1-31.

[5]. M. De Ryck, M. Versteyhe, K. Shariatmadar, Resource management in decentralised industrial automated guided vehicle systems, *Journal of Manufacturing System*, Vol. 54, 2020, pp. 152-173.

[6]. S. Haddadin, A. De Luca, A. Albu-Schäffer, Robot collisions, A survey on detection, isolation, and identification. *IEEE Transactions on Robotics*, Vol. 33, Issue 6, 2017, pp. 1292-1312.

[7]. A. M. Nascimento, L. F. Vismari, et al., A systematic literature review about the impact of artificial intelligence on autonomous vehicle safety, *IEEE Transactions on Intelligent Transportation Systems*, Vol. 21, Issue 12, 2019, pp. 4928-4946.

[8]. V. Kunchev, L. Jain, V. Ivancevic, A. Finn, Path planning and obstacle avoidance for autonomous mobile robots, a review, in *Proceedings of the 10th International Conference on Knowledge-Based & Intelligent Information & Engineering Systems (KES'06)*, 2006, pp. 537-544.

[9]. H. Ma, S. Koenig, Optimal target assignment and path finding for teams of agents, in *Proceedings of the Int. Conference On Autonomous Agents and Multi Agents Systems (AAMAS'06)*, 2016, pp. 1144-1152.

[10]. D. H. Kim, N. T. Hai, W. Y. Joe, A guide to selecting path planning algorithm for Automated Guided Vehicle (AGV), *Lect. Notes Electr. Eng.*, Vol. 465, 2018, pp. 587-96.

[11]. N. Bahnes, B. Kechar, H. Haffaf, Cooperation intelligent autonomous vehicles to enhance container terminal operations, *Journal of Innovation in Digital Ecosystems,* Vol. 3, Issue 1, 2016, pp. 22-29.

[12]. M. Hafner, D. Cunningham, L. Caminiti, D. Del Vecchio, (2013). Cooperative collision avoidance at intersections, Algorithms and experiments, *IEEE Transactions on Intelligent Transportation Systems*, Vol. 14, Issue 3, 2013, pp. 1162-1175.

[13]. J. Grosset, A.-J. Fougères, M. Djoko-Kouam, C. Couturier, J.-M. Bonnin, Simulation of a collision and obstacle avoidance algorithm for cooperative industrial autonomous vehicles. in *Proceedings of the 2nd IFSA Winter Conference on Automation, Robotics & Communications for Industry 4.0 (ARCI'22)*, Andorra la Vella, Andorra, February 2-4, 2022, pp. 88-91.

[14]. K. Karur, N. Sharma, C. Dharmatti, J. E. Siegel, A survey of path planning algorithms for mobile robots, *Vehicles*, Vol. 23, 2021, pp. 448-468.

[15]. Intelligent Transport Systems (ITS). Vehicular Communications; Basic set of Applications; Analysis of the Collective Perception Service (CPS), Release 2, TR 103 562 V2.1.1, *ETSI,* 2019.

[16]. D. Bechtsis, N. Tsolakis, D. Vlachos, J. S. Srai, Intelligent autonomous vehicles in digital supply chains, a framework for integrating innovations towards sustainable value networks, *Journal of Cleaner Production*, Vo. 181, 2018, pp. 60-71.

(3610)

Autonomous Robotic Sorting of Recyclable Waste

V. Tzounakos and M. Maniadakis
Foundation for Research and Technology Hellas,
N. Plastira 100, Vasilika Vouton, Herakleion, Crete, Greece
E-mail: mmaniada@ics.forth.gr

Summary: The use of robots in Material Recovery Facilities (MRFs) can significantly facilitate the treatment of recyclables. Such robots need sophisticated visual and manipulation skills to be able to undertake the extremely heterogeneous, complex and unpredictable waste sorting task in industrial environments. We examine the industrial application of an autonomous robotic system for the categorization and physical sorting of recyclables, according to their material type. In particular, we focus on the development of (i) a low-cost computer vision module that exploits the deep learning technology to identify and categorize recyclables and (ii) a robotic controller for fast and accurate pick-and-place of the recovered materials. The composite system is deployed in a waste processing plant, where it is successfully assessed in recyclable separation in difficult and demanding industrial conditions.

Keywords: Machine learning, Robotics, Recyclable material recovery application.

1. Introduction

The current trend of environment-friendly 'circular economy' changes the way products are made and consumed. Circular economy is based on the assumption that materials will not end up at landfills but they will be recovered and reused. Traditionally, the recovery of materials from waste streams has been made by human's workers, a solution that suffers from low productivity and increased health risks. The last two decades, the solution of optical sorters has been made popular in Material Recovery Facilities (MRFs). It uses a combination of lights and sensors to illuminate and capture images of the objects. This technology faces difficulties in dealing with heavy items such as semi-full container packages, which are not rare in waste streams [1].

As a means to overcome the limitations of optical sorters, a new trend has appeared the last years, which considers the use of robotic technology for waste recovery and sorting [2]. This approach assumes low integration cost and easy installation to existing MRFs. Robotic waste sorting systems rely on high-cost imaging technology that exploits recent AI advancements to identify and categorize waste [3]. The present work supports the development of low-cost computer vision-based waste categorization modules that can be directly applied in the industry. To this end, we exploit the industrial research setup implemented in the MRF unit of the island of Crete, Greece, to collect images of different waste types. The images are further processed in the lab, to develop a rich and well documented recyclable waste dataset that is used to train an AI-powered module for recyclable detection and categorization, applied in the industry.

The computer vision modules provide a robust solution to a three-fold problem that of object identification (bounding box specification), localization (masking) and material type attribution (classification). The simultaneous accomplishment of this triplet makes our solution applicable in industrial conditions where several and potentially overlapping recyclables shown in the very same image, need to be identified, localized and classified. The computer vision module is integrated with a delta robot which perform the picking of recovered materials and their placement into bins dedicated to specific materials. The composite system is tested in demanding industrial conditions with very promising results in material sorting (average success rate 91:8 %).

2. Industrial Research Setup

We have implemented an industrial research setup to facilitate data collection, experimentation and system assessment in realistic conditions. The key components deployed to facilitate investigation and experimentation are presented in the following.

Conveyor Belt. We developed an industrial installation of 22:5 m long, 1:0 m wide (usable width: 8 m) belt, that moves with speed up to 0:25 m = sec and an encoder system that reports its speed in real-time.

Waste Feeder. A dual waste feeder is used to shed waste in a controlled rate onto the conveyor belt; one with uncontrolled urban waste, and another with a controlled waste-mix, for experimentation purposes.

Camera. A standard stereo full HD ZED camera is used to enable the automated categorization of the recyclable materials. The camera is placed 148 cm before the robot (in the conveyor's movement direction), at a height of 75 cm above the conveyor belt, looking downwards.

Robot. An ABB IRB360 delta robot has been installed above the conveyor belt, to enable the pick and transfer of the waste to the bins. The robot employed in the present work has a payload of 6 Kg being highly appropriate for applications assuming repetitive and fast-executed manipulation tasks.

Vacuum Gripper. To enable the automated pick-n-place of recyclables, a vacuum gripping system is attached to the endeffector of the robot, which consists of a vacuum blower that provides high volume suction to pick and hold the materials. Vacuum technology provides a robust and low-cost solution for material transfer [4].

3. Material Detection and Manipulation

We have implemented an integrated robotic system for recyclable sorting that is composed of two parts, a robotic manipulator for the physical separation of waste to different bins, depending on their material type and a vision-based material detection and categorization module.

Pick-n-Place of Recyclables. An essential part of the autonomous sorting process regards the physical separation of recyclables. We follow a typical pick-n-place approach consisting of the following steps: (i) horizontal translation from the current position to the target-waste position, (ii) vertical translation towards the waste with the pathetic involvement of the shock absorber which ensures good contact of the suction cup with the material surface, and activation of the vacuum to implement gripping, (iii) follow the material transferred on the conveyor belt for a short period of time to improve sealing, (iv) vertical translation upwards, moving away from the conveyor belt, with the requested material grasped by the suction cup, (v) horizontal translation towards the bin that corresponds to the waste material, where vacuum is deactivated for the disposal of the waste.

Vision-based Material Categorization. For the material classification, we adapt the well-known Mask R-CNN to the current classification task, by developing and testing four customized Mask R-CNN implementations that differ in (i) the dataset size, (ii) the number of the learning steps per epoch, (iii) the use or not of additional data augmentation strategies to improve generalization. The synthetic training datasets consists of aluminum cans in flat, crumpled and cylindrical shape, papers and nylons in flat, crumpled and ball shape, PET bottles in flat, arbitrarily crumpled and cylindrical shape with various contents (i.e., with or without liquids). After extensive experimentation we show that: (i) vision-based approaches can handle the demanding task of waste-material separation with overlapping, and (ii) the performance of the network Mask-RCNN is sufficiently effective even in multiple-object images that include difficult cases with partially overlapping recyclables, reaching an overall 91.8 % accuracy.

4. Industrial Application

The implemented system has been gradually integrated into the operation of the MRF unit to support the treatment of the increased volume of waste that is processed on a daily basis. *The performance of the robotic waste sorter is demonstrated in the following link: https://youtu.be/6dV4vg3EeqU.*

To further ensure the robustness of the system, the robot treats only the objects that have been recognized with a very high level of confidence (i.e., more than 98 %). Moreover, the pump-based vacuum has been very effective in grabbing and transporting objects with a weight of up to 500 g. Overall, the operation of the robot can effectively complement the manual sorting of recyclables, thus having a clearly positive impact on the productivity of the MRF. This is because, the fast, precise and tireless operation of the robot reduces the amount of waste to be treated by workers, thus facilitating the processing of the large volume of waste that arrives daily at the plant.

4. Conclusions

Current solid waste management practices are no longer sustainable. With the exploitation of state-of-the-art computer vision and robotics technology, the promise of finding new innovative ways to attaining sustainable waste management becomes more achievable and realistic. The proposed industrial research setup has crucially supported experimentation towards the development of a robotic waste sorter that is applicable in the real world.

Acknowledgements

This research has been co-financed by the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH—CREATE—INNOVATE (project name: ANASA, code: T1EDK-03110).

References

[1]. D. Bonello, M. Saliba, K. Camilleri, An exploratory study on the automated sorting of commingled recyclable domestic waste, *Procedia Manufacturing*, Vol. 11, 2017, pp. 686–694.

[2]. ZenRobotics Recycler – Robotic Sorting Using Machine Learning, https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.465.94&rep=rep1&type=pdf

[3]. Z. Zhang, H. Wang, H. Song, S. Zhang, J. Zhang, Industrial robot sorting system for municipal solid waste, in Intelligent Robotics and Applications (H. Yu, J. Liu, L. Liu, Z. Ju, Y. Liu, D. Zhou, Eds.), *Springer International Publishing*, 2019, pp. 342-353.

[4]. M. Koskinopoulou, F. Raptopoulos, G. Papadopoulos, N. Mavrakis, M. Maniadakis, Robotic waste sorting technology: Toward a vision-based categorization system for the industrial robotic separation of recyclable waste, *IEEE Robotics & Automation Magazine*, Vol. 28, Issue: 2, June 2021, 2021., pp. 50-60.

(3651)

Deep Learning Driven One-class Detection YOLO Model in RGB Data for Edge-computing Applications

P. Dolezel [1], D. Stursa [1], V. Rozsivalova [1] and P. Skrabanek [2]
[1] University of Pardubice, Faculty of Electrical Engineering and Informatics,
Studentska 95., 532 10 Pardubice, Czech Republic
[2] Brno University of Technology, Institute of Automation and Computer Science,
Technická 2896/2, 616 69, Brno, Czech Republic
Tel.: + 420 466 036 551
E-mail: petr.dolezel@upce.cz

Summary: With rapidly growing demand for various security and surveillance applications, the number of advanced sensors, especially smart cameras, capable of automated scene analysis is constantly increasing. To fulfill this demand, the researchers continuously utilize numerous computer vision, machine learning, and deep learning approaches to achieve as precise and efficient scene analysis as possible. Here, a comprehensive evaluation of various types of YOLO architecture, together with ASP U-Net, are tested for use in a one-class detection problem. Specifically, seven models are applied to detect people heads in several scenes with different difficulties. These difficulties include low light, crowded scene, overlap of heads, etc. As a result, an overall evaluation of each architecture in terms of accuracy for one-class detection, as well in terms of computational costs for edge-computing, is presented.

Keywords: Object detection, Surveillance, YOLO, ASP U-Net, Head detection.

1. Introduction

With rapidly growing demand for various security and surveillance applications, the number of advanced sensors, especially smart cameras, capable of automated scene analysis is constantly increasing. In order to fulfill this demand, the researchers and developers continuously utilize numerous computer vision, machine learning and deep learning approaches to achieve as precise and efficient scene analysis as possible.

Nevertheless, the approaches, which provide the best accuracy in scene analysis, also require high resolution input data and are computationally intensive. Moreover, the training stages of these approaches are often demanding in terms of memory and computation time, since millions of parameters, or even more, are required to be optimized. These features naturally increase the price of smart sensors based on mentioned approaches. Therefore, the development of edge-computing is highly encouraging in scene analysis tasks, since it efficiently reduces the hardware requirements for data processing.

Edge-computing serves as a middle layer between the sensor and the central processing device or cloud computing. It efficiently provides real-time small-scale computing and storage capabilities to ensure low latency and initial preprocessing of the data acquired from the sensor. Evidently, more advanced edge-computing solutions reduce the demands on the performance of the central processing device or cloud computing.

Edge-computing is a very important phenomenon in surveillance applications, since it directly (without subsequent processing in cloud) enables person detection, person counting, pose estimation, fall detection, face recognition, and other essential image processing tasks. In addition, it provides other benefits connected to privacy, i.e., no personal or sensitive data needs to be sent to the cloud.

One-class object detection is the simplest type of object detection task, but it is one of the most commonly applied detectors in surveillance systems. As a successful example of such detector, Sivachandiran et al. presented automated person detection model on surveillance videos [1], Skrabanek et al. proposed a head detector for orthogonally placed monocular cameras [2], Akter et al. offered deep learning-based surveillance system for unmanned aerial vehicle detection [3], etc.

Currently, methods for object detection generally fall into deep learning-based approaches [4]. Already classic neural network-based approaches include Fast R-CNN [5], Faster R-CNN [6], SSD [7], or SPPNet [8]. Detectors based on the YOLO family of architectures [9] are particularly successful and computationally efficient.

Unlike other object detection methods, YOLO-based architectures do not divide the object detection task into more processes, e.g., object region prediction and class prediction. The YOLO algorithm integrates both tasks into a single neural network model to achieve fast detection with high accuracy.

The authors of YOLO progressively published several generations of architectures – YOLOv2, YOLO9000, YOLOv3. In addition, the YOLO models were scaled for different ratios of accuracy to computational costs, which led to architectures such as tinyYOLO [10]. The original YOLO was further developed by other authors, leading to YOLOv4 [11],

YOLOv5 [12], YOLOv6 [13], YOLOv7 [14] and other successors. Although there exist several survey articles that compare the performance of different types of YOLO architectures, these articles generally use multi-class datasets, such as MS COCO or PASCAL VOC [15]. In this study, we aim at surveillance applications, where the goal is often to detect only one type of object or phenomenon in the monitored area. On the other hand, a given object is expected to be detected with high accuracy regardless of lighting conditions, scene clarity, background type, etc. Therefore, we decide to test and compare selected types of YOLO architectures for a one-class detection problem. We compose a robust dataset with various representations of one type of object taken under a variety of conditions – see Fig. 1. Using this dataset, we train and compare selected YOLO architectures and additionally evaluate them in terms of computational costs. In order to get closer to edge-computing applications, we use NVIDIA Jetson NANO for evaluation of experiments.

Fig. 1. Example of images in dataset.

2. Methodology

Here, the aim of this study is properly defined. Subsequently, the resources and algorithms used to retrieve the results are described.

2.1. Problem Formulation

The objective of our work is to design a one-class detector suitable for surveillance applications. Moreover, the detector should meet the requirements of edge-computing applications. Therefore, the detector is expected to fulfill several specific features. First of all, high accuracy of object centroid detection is required. On the contrary, there is no particular reason to specify whole bounding boxes or object shapes, as this information would be useless for surveillance applications. Furthermore, the detector operation shall require low memory consumption and be as computationally inexpensive as possible.

The spatial arrangement of the detector for human head detection, and the required detector output are shown in Fig. 2.

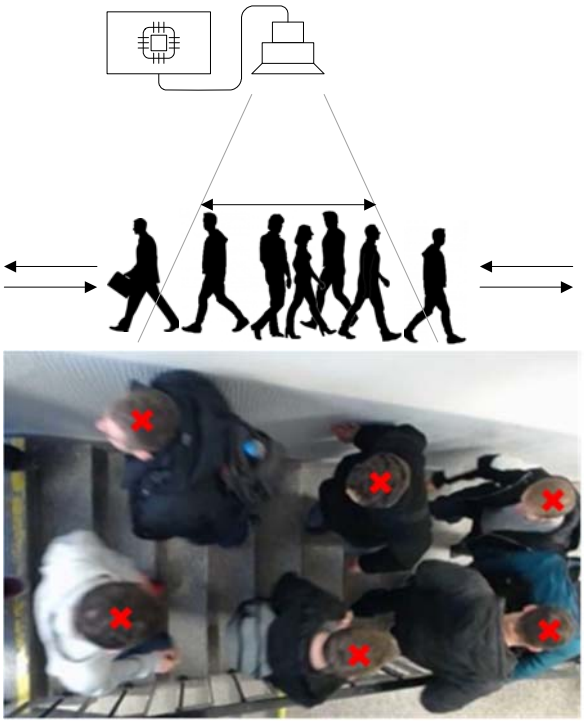


Fig. 2. The spatial arrangement of the detector for human heads detection (upper image), and the required detector output (red crosses represent centroids of the human heads).

2.2. Selected Architectures

As discussed above, many deep learning architectures, suitable for object detection, were introduced in the last decade. YOLO-based architectures are particularly suitable for edge-computing due to their single-step character. Based on some pilot studies and literature survey, we select 6 specific implementations of YOLO for testing. Moreover, we add a custom architecture ASP U-Net developed by the authors [16] for comparison. This architecture was originally used for grasping point detection, but it can be redesigned to detect any type of significant point in visual data. While YOLO architectures provide bounding boxes of detected objects, ASP U-Net directly delivers centroid coordinates in the form required in this study. The complete list of the selected architectures is summarized in Table 1.

Table 1. Selected architectures for evaluation.

Architecture	Source	Memory size
YOLOv5s	[12]	14 MB
YOLOv5n	[12]	4 MB
YOLOv5l	[12]	90 MB
YOLOv7tiny	[14]	11 MB
YOLOv7	[14]	73 MB
YOLOv7x	[14]	138 MB
ASP U-Net	[16]	39 MB

2.3. Dataset

The quality of a dataset predetermines performance of the detector. To guarantee robustness of the head detector, we collect data in diverse environments which include staircases, corridors, and entries into means of transport. We capture video streams above walking persons at eight different locations. The lighting conditions differ among the experiments. Additionally, we merge the captured video streams with publicly available datasets [17-19].

We extract frames from the captured videos to create a set of 25 301 8-bit RGB images. We resize images to 288×288 px. Finally, we randomly split the images into training set, testing set, and validation set in the ratio 74:13:13.

2.4. Evaluation Measures

Since the detector is expected to detect the positions of the centroids of the heads in the image, we select Total Localization Error as the primary metric for our tests. In simple words, Total Localization Error is a sum of the smallest relative distances between ground truth centroid coordinates and their closest predictions. Each such prediction is associated exactly with one ground truth label, and simultaneously, each ground truth label is associated exactly with one prediction. The lower the metric value provided, the better is the detector. This metric was originally introduced in [20].

Relative Inference Time is selected as the secondary metric. This metric expressed the ratio between the inference time of the actual detector and the inference time of the baseline detector. The inference times are statistically evaluated using the NVIDIA Jetson NANO.

2.5. Detector Training

We train the detectors using the Adam algorithm based on its generally acceptable performance with binary cross entropy as the cost function. Initial weights are set randomly with normal distribution (mean = 0, standard deviation = 0.05). The experiments are performed twenty times due to a stochastic character of training. Moreover, we use data augmentation to avoid overfitting by the training of the map generators. Specifically, we use random rotation ± 20 degrees, random horizontal and vertical flipping with probability 0.5, random horizontal and vertical translation (up to ± 20 % of image height and width), random rescaling (zoom range 0.2) and random horizontal and vertical shear (shear intensity 0.2). We set 200 as the maximal number of the epochs and we use the best performance over the validation set as the selection criterium for the detector model.

Some of the selected YOLO variants use various other parameters for training. In such cases, we keep these parameters as the default, as recommended in the respective sources.

3. Results and Discussion

We train the selected detectors with the dataset described in Section 2.3 according to the procedure addressed in Section 2.5. In order to show the performance capabilities of the trained detectors, we summarize the best resulting values of the evaluation measures described in Section 2.4 to Table 2.

Additionally, we demonstrate absolute frequencies of differences between numbers of ground truth labels and numbers of predictions of the detectors for the test dataset (Table 3).

Table 2. Measures of the selected detectors for the testing set.

Architecture	Total Localization Error	Relative Inference Time
YOLOv5s	0.5684	1.2267
YOLOv5n	0.5893	1.2301
YOLOv5l	**0.5510**	1.9847
YOLOv7tiny	0.5731	**1.0000**
YOLOv7	0.6871	1.5082
YOLOv7x	0.6481	1.8017
ASP U-Net	0.5572	2.9917

Table 3. Absolute frequencies of differences between numbers of ground truth labels and numbers of predictions for the testing set.

Architecture	<-2	-2	-1	0	1	2	>2
YOLOv5s	69	42	176	2656	243	22	0
YOLOv5n	55	38	159	2610	315	31	0
YOLOv5l	72	54	184	**2705**	184	9	0
YOLOv7tiny	10	32	145	2557	373	78	13
YOLOv7	76	72	237	2517	286	20	0
YOLOv7x	74	63	236	2566	253	16	0
ASP U-Net	4	21	237	2547	301	72	24

The best overall values of Total Localization Error are provided by the detector based on the YOLOv5l architecture. Moreover, this detector provides the least error in differences between numbers of ground truth labels and numbers of predictions (see Table 3). If memory consumption and inference time are critical for the intended application, it is possible to consider the YOLOv7tiny architecture with its outstanding relative inference time, and it is only 11 MB in size. ASP U-Net, which was selected because it directly provides centroids of detected objects, provides above average results in terms of accuracy. However, its inference time is by far the worst of all the selected architectures.

4. Conclusions

In this study, various types of YOLO architecture, together with ASP U-Net, were compared for use in a

one-class detection problem with a special attention to edge-computing surveillance applications. Head detection problem in various types of crowded scene was specifically addressed. In general, detectors based on the YOLO architecture are accurate and efficient. When choosing a particular architecture, a trade-off between overall accuracy and inference time should always be sought. Conveniently, within the YOLO family of architectures one can find both tiny models suitable for single-chip applications, and large models characterized by high accuracy and generalization capability.

Acknowledgement

The work has been supported by the SGS grant no. SGS_2022_014 at the University of Pardubice. This support is very gratefully acknowledged.

References

[1]. S. Sivachandiran, K. Jagan Mohan, G. Mohammed Nazer, Deep learning driven automated person detection and tracking model on surveillance videos, *Measurement: Sensors*, Vol. 24, 2022, 100422.

[2]. P. Skrabanek, P. Dolezel, Z. Nemec, D. Stursa, K. Xie, Person detection for an orthogonally placed monocular camera, *Journal of Advanced Transportation*, Vol. 2020, 2020, pp. 1-13.

[3]. R. Akter, V.-S. Doan, J.-M. Lee, D.-S. Kim, CNN-SSDI: Convolution neural network inspired surveillance system for UAVs detection and identification, *Computer Networks*, Vol. 201, 2021, 108519.

[4]. S. S. A. Zaidi, M. Samar Ansari, A. Aslam, N. Kanwal, M. Asghar, B. Lee, A survey of modern deep learning based object detection models, *Digital Signal Processing*, Vol. 126, 2022, 103514.

[5]. R. Girshick, Fast R-CNN, in *Proceedings of the IEEE International Conference on Computer Vision (ICCV'15)*, 2015, 2015, pp. 1440-1448.

[6]. S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: Towards real-time object detection with region proposal networks, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 39, Issue 6, 2017, pp. 1137-1149.

[7]. W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, A. C. Berg, SSD: single shot MultiBox detector, in *Proceedings of the European Conference on Computer Vision (ECCV'16),* 2016, pp. 21-37.

[8]. K. He, X. Zhang, S. Ren, J. Sun, Spatial pyramid pooling in deep convolutional networks for visual recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 37, Issue 9, 2015, pp. 1904-1916.

[9]. J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: Unified, real-time object detection, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'16)*, 2016, pp. 779-788.

[10]. J. Cui, Y. Zhang, Z. Wang, J. Liu, Light-weight object detection networks for embedded platform, *Acta Optica Sinica*, 2019, Vol. 39, Issue 4, 0415006.

[11]. A. Bochkovskiy, C.-Y. Wang, H.-Y. M. Liao, Yolov4: Optimal speed and accuracy of object detection, *arXiv Preprint*, 2020, arXiv:2004.10934.

[12]. YOLOv5, https://github.com/ultralytics/yolov5

[13]. C. Li, et al., YOLOv6: A single-stage object detection framework for industrial applications, *arXiv Preprint*, 2022, arXiv:2209.02976.

[14]. C.-Y. Wang, A. Bochkovskiy, H.-Y. M. Liao, YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors, *arXiv Preprint*, 2022, arXiv:2207.02696.

[15]. K. Gauen, R. Dailey, J. Laiman, et al., Comparison of visual datasets for machine learning, in *Proceedings of the IEEE International Conference on Information Reuse and Integration (IRI'17),* 2017, pp. 346-355.

[16]. P. Dolezel, D. Stursa, D. Kopecky, J. Jecha, Memory efficient grasping point detection of nontrivial objects, *IEEE Access*, Vol. 9, 2021, pp. 82130-82145.

[17]. S. A. Velastin, R. Fernández, J. E. Espinosa, A. Bay, Detecting, tracking and counting people getting on/off a metropolitan train using a standard video camera, *Sensors*, 2020, Vol. 20, Issue 21, 6251.

[18]. PHDF-Dataset, https://github.com/SMWSIELab/PHDF-Dataset

[19]. S. Sun, N. Akhtar, H. Song, C. Zhang, J. Li, A. Mian, Benchmark data and method for real-time people counting in cluttered scenes using depth sensors, *IEEE Transactions on Intelligent Transportation Systems*, Vol. 20, Issue 10, 2019, pp. 3599-3612.

[20]. P. Dolezel, P. Skrabanek, D. Stursa, B. Baruque Zanon, H. Cogollos Adrian, P. Kryda, Centroid based person detection using pixelwise prediction of the position, *Journal of Computational Science*, Vol. 63, 2022, 101760.

(3664)

Web Application to Support the Credibility of Social Web Content

K. Machová and L. Pončáková
Technical University of Košice, Department of Cybernetics and Artificial Intelligence,
Letná 9, 04200 Košice, Slovakia
Tel.: +421 55 6024142, fax: +421 55 6253574
E-mail: kristina.machova@tuke.sk

Summary: Social media has become part of our daily lives and is a source of information as well as a media channel for following and disseminating the latest news around the world. Users can in an anonym way express their opinions on social networks by sharing or commenting news. Today, social media and online communication play a crucial role in the influencing a society, which has its advantages and disadvantages. In this article, we will look at the issue of the trustworthiness of the social web. The paper presents a web application helpful in recognition of a credibility of the conversational content. The parts of this application are at first a model for recognition of hoaxes based on a dictionary approach, but also an interactive game. The task of the game is to point out what is important to notice and what to watch out for when drawing information from the Internet. Among other things, the web application also includes the possibility to verify the website and static parts containing videos about harmful asocial behaviour, explanations of terms etc.

Keywords: Social media, Incredible web content, Web application, Dictionary approach to text processing, Interactive game.

1. Introduction

Social media has seen increased use as a source of information and is mainly used to search for information on serious topics. There has also been great use by those who seek health information. People use social "tools" to gather information, share stories, but also discuss issues. Similarly, healthcare organizations see benefits of social media because they give them access to healthcare information.

As Sutton, Palen and Shklovski [1] suggest, social media comes to the fore as a source of information in times of disaster and risk situations, although the accuracy of the information that is shared through these channels is unclear. Therefore, it is essential to learn more about how people evaluate the information they receive on social media websites, especially in terms of their credibility.

1.1. WISDOM System

The credibility of information on the Internet is becoming increasingly important, and therefore the web analytical system WISDOM (Web Information Sensibly and Discreetly Ordered and Marshaled) has been developed, which is based on the criteria of sender, content, and appearance (page layout, number of advertisements, providing of contact address) analyses the website, thus promoting the credibility of the information. With this information, one assesses the credibility of a website without understanding its content. For example, if a user finds a contact address on a website, they will find that the site is more trustworthy than if it is not there. Also, if he finds more than one advertisement on a website, he thinks the site is less trustworthy [2]. The use of appearance of information in the web information analysis system is illustrates in Fig. 1.

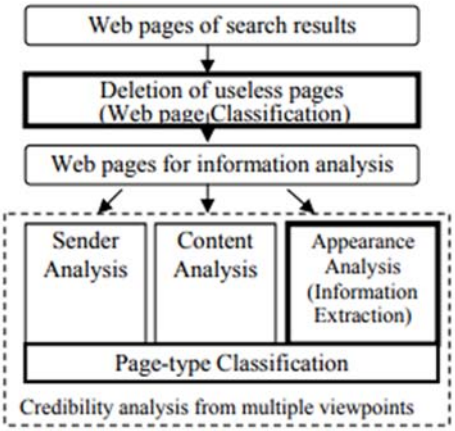


Fig. 1. Use of appearance information in Web information analysis system [1].

The evaluation of existing games showed that bad news improves the ability to resist misinformation after it is played [3].

2. The Model Based on a Dictionary Approach

We created a model that is based on a dictionary approach for hoaxes recognition [4] in the Python programming language. First, to be able to apply the given model, we need a dictionary for applying to the selected text. The model is illustrated in Fig. 2.

2.1. Dictionary Creation

Creating a dictionary was time consuming because the dictionary was created manually in the Slovak language, so it was necessary to translate English words and determine the value – weight of the words.

The weight of the words ranges from 1 to 5, where the higher the number, the more toxic the word. The dictionary is composed of combinations of different dictionaries, which contain negative and vulgar words, their various forms and their synonyms. In addition, commonly used abbreviations that have a negative meaning have been added to the dictionary. Furthermore, this dictionary was enriched with phrases such as: "it's just my opinion", "I was just thinking" and others. The final dictionary consists of 1240 words and phrases.

Fig. 2. A model based on a dictionary approach (Google Cloud Platform).

2.2. Text Pre-processing

Pre-processing is very important in a dictionary-based method, as overall pre-processing plays an important role in achieving the expected results. The following points describe the steps we needed to take to have the resulting text divided into words and phrases as we needed to: convert text to lowercase, remove punctuation, replace numbers with an empty string, remove diacritics of phrases, assign value to phrases, split text into words, delete word traces, remove word accents, assign value to words.

2.3. Evaluation of the Model Based on the Dictionary

When evaluating the model, we first needed to obtain toxic and non-toxic reviews for a testing set. Toxic reviews, the number of which was 35, were obtained from the Facebook page Hoaxes and Frauds – Police of Slovak Republic. Non-toxic reports, the number of which was also 35, were obtained from various websites and newspapers, such as https://www.aktuality.sk/, https://www.sme.sk/ and others.

If we have already assigned values of toxicity to phrases and words, then we summed the given values and calculated the arithmetic mean of toxicity. If the resulting value was greater than 3, the text was considered toxic. If the resulting value was less than 3 but also greater than 0, the text was considered non-toxic. If the resulting value was 0, we considered the text to be probably non-toxic, but if the resulting value is equal to 3, we considered the text to be likely toxic.

Based on these results, we therefore decided to consider a text with a value of 3 as "likely to be toxic", because if we consider it directly as "toxic", messages that are non-toxic (with a value of 3) would be directly labelled ‘toxic’. Thus, when we mark a text with a value of 3 as "probably toxic", we let the user know that the text can also be "non-toxic", so it still needs to be verified. The Table 1 represents results of our dictionary-based model in values of Precision, Recall and Accuracy.

Table 1. Results of testing the dictionary-based model.

	Value 3 as nontoxic	Value 3 as toxic
Precision	0.800	1.000
Recall	0.933	0.814
Accuracy	0.871	0.886

3. Application to a Web Portal

We proceeded in two ways to apply the model to the web portal. The first way was to apply the model to the Google Cloud Platform. This method was more complicated because we had to download and install the Google Cloud SDK for our operating system, create a project on the Google Cloud Platform, where we then uploaded our model based on dictionary approach. The second way was to create an EXE file and apply it to a web portal where it is available for download.

The dictionary-based model was applied to the Google Cloud Platform, and it is available on our web portal: https://oversi.onrender.com/index.html.

3.1. Creating a Game for Application

Application obtained also an interactive game. We used the free development tool Twine to create the interactive game. In the process of the game development the following tools were used – JavaScript, CSS and so-called "Blocks" that follow each other, or even branch out. We also created a GitHub account, which we used to store the music and images applied in the game.

When creating the game, we focused on the school environment and tried to make this interactive game interesting not only with the story, but also with its educational content. The content is mainly focused on the credibility of information on the Internet. The story of this game lies in the high school life of 17-year-old Oliver. The main task is to help Oliver, who is not one of the best students, to manage the day at school as best he can.

The game begins with entering your name and its ratification. Then, you are in the room at the beginning of the game. The story goes in the direction you decide, by choosing from the offered answers. The game includes a test that contains 5 questions. There are two

answering options for each question, of which only one is always correct. An example of the game test is illustrated in Fig. 3, particularly Question 1 with two answers in green colour.

After confirming the correct or incorrect answer, a window appears, which will evaluate user's answer and explain why the answer is correct or incorrect (Fig. 4).

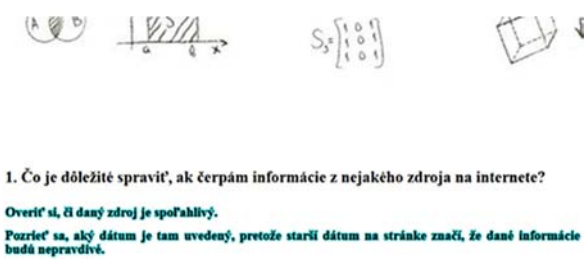


Fig. 3. An example of the game test (Question no.1).

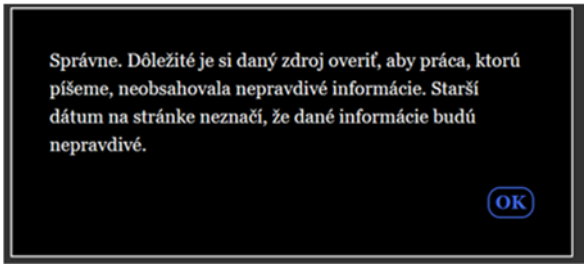


Fig. 4. Evaluation and explaining the correct answer on the question no.1.

Another part of the game is the selection (in green colour) of an untrustworthy article (Fig. 5) from a pair of provided articles A and B.

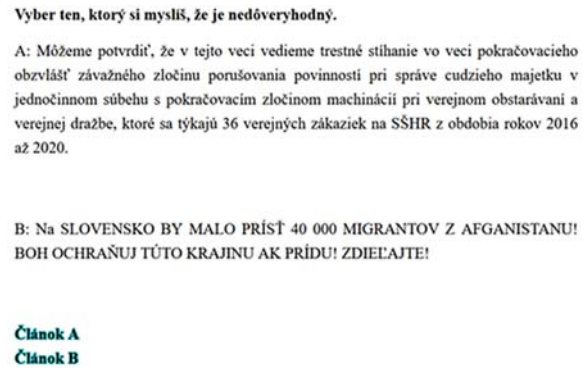


Fig. 5. A pair of two articles A and B, from which only one is trustworthy.

Game also includes a part where the player gets the opportunity to decide based on advertisement and thus the game will continue based on this decision. Once created, the game was also applied to our web portal (Fig. 6).

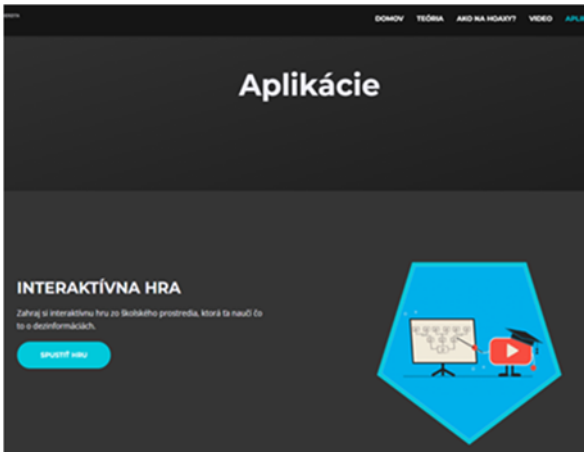


Fig. 6. An application of the interactive game on our web portal.

3.2. Evaluation of the Interactive Game

To evaluate our game, we decided to address the 48 students of the second and third year of high school in Košice. After the game, the students answered four questions:

1. Do you verify the information you find on the Internet?
2. After playing the game, did you gain new insights into the credibility of information on the Internet?
3. After playing the game, will you pay more attention to the sources of the news / posts you will find on the Internet?
4. Did the given web portal / presentation help you orient yourself in the issue of disinformation?

There were five possible answers on each question: Yes, Rather yes, I do not know, Rather no and No.

The results of the evaluation showed that the game was successful, and thus 69 % of students after playing the interactive game pay more attention to the sources of news / posts they find on the Internet, as only 23 % of students used to check information on the Internet before. Also, most students stated that they gained new knowledge after the game. As for the overall evaluation of the web portal, it was also a success, as up to 71 % of students stated that the web portal helped them to orientate themselves in the issue of misinformation what is illustrated in the Fig. 7.

The students also were familiar with our web portal applying three ways of learning:

1. Presentation from the content of the web portal;
2. Web portal searching;
3. Hybrid form – viewing some parts of the web portal enriched with a presentation "How not to get lost at the time of information".

During all these ways of learning, students had to fulfil the same questionary at the beginning and at the end of the learning activity. Then, the change in accuracy of an ability to detect a toxicity in online space was measured in all three cases.

Fig. 7. An application of the interactive game on our web portal. The frequency of five possible answers on the question "Did the given web portal / presentation help you orient in the issue of disinformation?".

During the first learning activity "*Presentation from the content of the web portal*", after completing the first questionnaire, the students became acquainted with the interactive game. Subsequently, the presentation of individual terms and advices what needs to be noted when verifying the information on the Internet and presentation from the content of the web portal were provided. Students had an opportunity to ask questions about the whole issue of misinformation through the "sli.do" platform or directly by the lifting hand. The answers to students' question were given continuously, but mainly at the end of the meeting. Before the end, they were provided with the second questionnaire with the same questions as at the beginning. On the basis of a comparison of results of these two questionnaires it was possible to evaluate the success of learning of the necessary information on this topic.

During the second learning activity "*Web portal searching*" after acquaintance with the interactive game, there was no presentation, but self-study and viewing the web portal separately. The meeting was terminated by a questionnaire with the same questions as at the beginning and the comparison was provided.

At the beginning of the third learning activity "*Hybrid form*", the students received a questionnaire from chosen terms and tips on how to do a critical analysis of the text on internet. Filling the questionnaire was followed by viewing the section "Theory" of our web portal along with trying out individual applications, namely interactive game, verification whether the site is or is not in the list of disinformation websites and verification of text, whether it has signs of toxicity. Later, the presentation was followed by the "How not to lose ourselves in the era of information", where the tips and tricks to properly verify the information on the Internet were presented.

At the end of each lesson of each learning activity, we had a questionnaire fulfilled by students, where more than 2 thirds answered that our web portal and its presentation helped them to orientate themselves in the issue of misinformation, what is also illustrated by Fig. 7. The results of students testing in all three learning activities are presented in the Table 2.

Table 2. Accuracy of an ability to detect toxicity in online space.

Number of learning activity	Test before activity [%]	Test after activity [%]
1	54.5	77.5
2	64.5	82.5
3	60.6	86.3

4. Conclusions

The results of the web application evaluation show that it was successful, and thus the desired goal of the work was met, as both – the model based on the dictionary approach and the interactive game achieved positive results. The Accuracy of our model for the toxic texts (hoaxes) recognition based on the dictionary approach was highest when we considered the text with a value of "3" to be toxic. On the other hand, we achieved a higher Recall when we considered the text to be non-toxic with a value of toxicity "3". The disadvantage of the dictionary approach is that the used dictionary may not contains all the necessary words, and therefore, even if the text is toxic, but none of the words in text is not in the dictionary, the approach cannot evaluate the text correctly. Also, the text can often contain a sarcasm, which is difficult to detect automatically and can skew the results of the model. Therefore, if training data are available, it is possible to use a machine-based approach.

Acknowledgements

This work was supported by the Slovak Grant Agency of the Ministry of Education and Academy of Science of the Slovak Republic under grant no. 1/0685/21.

References

[1]. I. Shklovski, L. Palen, J. Sutton, Finding community through information and communication technology in disaster response, in *Proceedings of the Conference on Social Semantic Web (CSSW'08),* San Diego, CA, USA, November 2008, pp. 1-10.

[2]. S. Akamine, et al., Using appearance information for web information credibility analysis, in *Proceedings of the 2nd International Symposium on Universal Communication (ISUC'08),* Dec. 2008, pp. 363-365.

[3]. S. Linden, et al., Inoculating against fake news about COVID-19, *Front Psychology*, Vol. 11, 2020, 566790.

[4]. S. Sucipto, et al., Hoax detection at social media with text mining classification system-based, *JurnalI lmiah Penelitiandan Pembelajaran Informatika*, Vol. 3, Issue 2, December 2018, pp. 94-100.

(3694)

Causal Effect Identification with Minimum Mean Estimation Error

J. Etesami, N. Kiyavash and M. Grossglausser
École polytechnique fédérale de Lausanne, Switzerland
Tel.: +41 21 693 14 09
E-mail: seyed.etesami@epfl.ch

Summary: Estimating the effect of an intervention (causal effect) from mere observational data is known as causal identification problem which appears in different applications. Existing a valid adjustment set for a causal effect makes that effect identifiable from observational data. In many real-world scenarios, there exist several adjustments sets for a causal effect. In this work, we study the problem of selecting an adjustment set for a causal effect that has the minimum mean estimation error. We introduce joint information-theoretic and graphical selection criteria that allows us to choose the best adjustment set. Such criteria are not limited to selection of the best valid adjustment set but could also be extended to selecting the best equivalent observational expression obtained by Pearl's do-calculus.

Keywords: Causal effect identification, do-calculus, Adjustment set, Sample complexity.

1. Introduction

The canonical inference problem in causal reasoning is to estimate the effect of so-called cause variables on a set of outcome variables. Performing interventions is the gold standard of causal reasoning. An intervention on a variable varies its conditional distribution given its direct causes. The information obtained from an intervention depends on the type of the intervention, as well as the size of the intervention, and the location of the targets of the intervention in the underlying causal graph. For instance, the changes in a causal model as the result of hard interventions can be modeled by procedures that Pearl called graph surgery [1]. This result in a new graph, the "interventional graph" in which the causal modules involving variables that have been intervened on are changed. While conditional distributions are based on the original causal graph, interventional predictions are based on the new manipulated graph. Fig. 1 depicts the process of graph surgery in a simple causal graph.

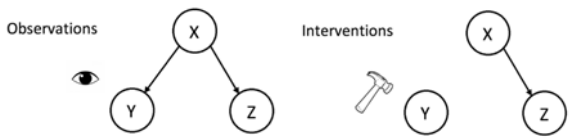


Fig. 1. The process of graph surgery in a simple 3 variable graph.

More precisely, in causal effect identification problem, we are given two disjoint subsets of variables, say X and Y and faced with inferring the causal effect of X on Y. Several measures have been developed in the literature to capture such effects. The most commonly accepted such measure is the post-interventional distribution $P\big(Y\big|do(X)\big)$. In many real-world applications, direct intervention is not possible. This might be due to the nature of the system that makes it impossible to intervene on X and observe its effect on Y or might be due to legal or ethical issues. For instance, it is both illegal and unethical to force a group of people to smoke in order to measure the effect of smoking on lung cancer. In such scenarios, the causal effects $P\big(Y\big|do(X)\big)$ should be estimated using only observational data. If this is possible, we say the causal effect is identifiable. For instance, when valid adjustment sets or instrumental variables exist, they allow us to find an equivalent expression for $P\big(Y\big|do(X)\big)$ from only observational joint distribution [1]. Proving that such equivalent expression exists is based on a set of rules known as do-calculus [1]. It is shown in [2], that the set of rules in do-calculus are complete meaning that if it is not possible to find an equivalent expression for $P\big(Y\big|do(X)\big)$ in a causal structure using the do-calculus, then it is impossible to identify the effect. The problem of finding equivalent observational expression (i.e., selecting the valid adjustment sets) for a given post-interventional distributions is profusely studied [2, 4].

1.1. Problem Setting

Valid adjustment sets are not necessarily unique. That is, several valid adjustment sets might exist that lead to different equivalent observational expression for a given post-interventional distribution. Different expressions lead to different estimation errors.

Formally, consider a set of variables V with the its corresponding causal directed acyclic graph (DAG) G. Let X, Y denotes two disjoint subsets of V. We are interested in estimating $P\big(Y\big|do(X)\big)$ from observational

data. Valid adjustment sets are one way to do so. Let $Z \subset V$ that contains no descendant of X and blocks (d-separates) all paths from X to Y with arrows pointing into X, then Z satisfies the so-called backdoor criteria [1] and hence it is an adjustment set, i.e.,

$$P\big(Y\big|do(X)\big) = \sum_Z P(Y|X,Z)P(Z). \qquad (1)$$

The right-hand side of the above expression is an expression that can be estimated using observational data. Different choices of Z will lead to different estimators for the left-hand side with different estimation errors. We consider the following choice of estimation error metric:

$$\mathbb{E}\Big[\ \Big|\Big|P\big(Y\big|do(X)\big) - \hat{P}_{N,Z}\big(Y\big|do(X)\big)\Big|\Big|\ \Big],$$

where $\hat{P}_{N,Z}\big(Y\big|do(X)\big)$ denotes the estimator obtained from selecting Z as the adjustment set with sample size N and $||.||$ is the l_1-norm. Estimators with lower sample complexities are more desirable. Hence, the selection inference problem can be formalized as

$$\arg\min_Z \mathbb{E}\Big[\ \Big|\Big|P\big(Y\big|do(X)\big) - \hat{P}_{N,Z}\big(Y\big|do(X)\big)\Big|\Big|\ \Big].$$

Note that the optimization problem fixes the sample size and asks for the best Z for a given sample size. Authors in [3] study this problem for a special setting in which all relations between variables are linear. We generalize this result to non-linear systems. Next, we study this problem in two special DAGs.

2. Special DAGs

Consider the causal DAG in Fig. 2. Suppose, we are interested in estimating the causal effect of X on Y. Using the back-door criterion, it is easy to see that $\{Z_1\}$, $\{Z_2\}$, and $\{Z_1, Z_2\}$ are all possible adjustment sets for $P\big(Y\big|do(X)\big)$.

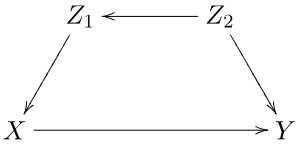


Fig. 2. A causal DAG.

We are interested in selecting an adjustment set that results in the smallest mean estimation error for a fixed number of observations.

As we show next the choice of the adjustment set is not only dependent on the causal graph but also on the marginal distributions of the variables. To do this end, from the structure of the DAG in Fig. 2, we obtain

$$H(Y|X,Z_2) \leq H(Y|X,Z_1), \qquad (2)$$

where $H(.)$ denotes the entropy function. This implies that the first term on the right side of Equation (1) can be estimated with lower sample complexity for Z_2 compared to Z_1. On the other hand, if Z_1 and Z_2 are related such that $H(Z_1) \geq H(Z_2)$, then Z_2 will be a better adjustment set compared to Z_1 in terms of mean estimation error. As an example, consider the next model for the causal DAG in Fig. 2 in which all variables are binary.

Model I: $P(Z_1 = 0|Z_2 = [0,1]) = [0.4,0.6]$,
$P(Z_2 = 0) = 1/7, P(X = 0|Z_1 = [0,1]) = [0.4,0.6]$,
$P(Y = 0|[X,Z_2] = [00,01,10,11]) = [0.5,0.501,0.02,0.01]$.

Fig. 3 depicts the mean estimation error with respect to the sample size N. As it is depicted in Fig. 3. (left), Z_2 is always a better choice than Z_1.

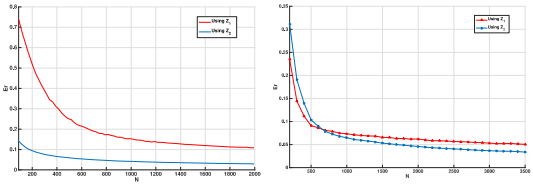


Fig. 3. Mean estimation error as a function of sample size for model I (left) and model II (right), respectively.

If $H(Z_2) \geq H(Z_1)$, the above statement does not always hold. To see this, consider the following model for the same causal DAG in Fig. 2.

Model II: $P(Z_1 = 0|Z_2 = [0,1]) = [0.998,0.99]$,
$P(Z_2 = 0) = 0.5, P(X = 0|Z_1 = [0,1]) = [0.99,0.02]$,
$P(Y = 0|[X,Z_2] = [00,01,10,11]) = [0.5,0.501,0.02,0.01]$.

In this case, as it is depicted in Fig. 3 (right) for small N, Z_1 is a better adjustment set than Z_2. Now, consider a slightly different DAG illustrated in Fig. 4.

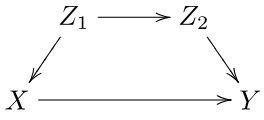


Fig. 4. A causal DAG.

It is easy to see that similar to the previous DAG both Z_1 and Z_2 are adjustment sets for $P\big(Y\big|do(X)\big)$ and $P(Y,X,Z_1,Z_2) = P(Z_1)P(Z_2|Z_1)P(X|Z_1)P(Y|X,Z_2)$. This implies inequality (2). Hence, the relationship between the marginal distributions of Z_1 and Z_2 will determine which one is a better adjustment set.

3. General DAGs

Consider a general DAG in which there exists two disjoint valid adjustment sets Z_1 and Z_2 for

$P(Y|do(X))$ such that both satisfy the back-door criterio. Moreover, assume that the directed path from Z_1 to Y contains Z_2. Thus, Z_2 d-separates Y from X and Z_1 or equivalently, inequality (2) holds. Therefore, deciding on which adjustment sets has lower mean estimation error comes down to the relationship between marginal distributions of Z_1 and Z_2.

4. Conclusions

We studied the problem of selecting an adjustment set with the minimum mean estimation error for identifying the effect of an intervention. We showed that this selection is based on joint information-theoretic and graphical criteria.

References

[1]. J. Pearl, Causality, *Cambridge University Press*, 2009.

[2]. I. Shpitser and J. Pearl, Identification of joint interventional distributions in recursive semi-markovian causal models, *in Proceedings of the National Conference on Artificial Intelligence AAAI*, Vol. 21, 2006, p. 1219.

[3]. I. Maathuis, M. H., Perkovic, E., and Henckel, L., Graphical Criteria for Efficient Total Effect Estimation via Adjustment in Causal Linear Models, *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, 84, 2, 2021, pp. 579-599.

[4]. M. Kuroki and Z. Cai, Selection of identifiability criteria for total effects by using path diagrams, in *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence*, 2004, pp. 333– 340.

(3700)

Crest Factor Minimization for Wide-bandwidth Multi-sine Signals with Non-flat Amplitude Spectra using Partial Modulation

H. Althoff, S. Geinitz and C. Linder
Fraunhofer Institute for Casting, Composite and Processing Technology IGCV,
86159 Augsburg, Germany
Tel.: + 4982190678222, fax: + 498219067840222
E-mail: steffen.geinitz@igcv.fraunhofer.de

Summary: Monitoring the behavior of adhesives using dielectric analysis is challenging due to additives and fast curing reactions. Based on previous work, using multi-sine signals, partial signal components are introduced to increase the number of monitored frequencies while at the same time reducing the crest factor. A small crest factor indicates a high-quality multi-sine signal, which stands for high information density. Previous problems include that many similar frequencies increase the crest factor. Integrating partial frequency components counteracts this. Further, the error resulting from DFT analysis decreases when using this modulation method. The construction of multi-sine waves designed for dielectric analysis usually leads to a significant error in DFT, which is why this result is particularly beneficial. Thus, more information can be obtained at the same time with higher accuracy. Due to the specific signal composition and the DFT algorithm, an upper limit in frequencies is observed.

Keywords: Dielectric analysis, Crest factor optimization, Multi-sine signal, FFT, DFT.

1. Introduction

The production monitoring of adhesives, fiber reinforced or neat plastics becomes more challenging as additives or new compounds are used to increase the performance and the reaction rates as well as reduce the process times. [1-5].

Dielectric Analysis (DEA) is a measurement principle suitable for online monitoring of curing or phase shift processes of materials, derived from the more general electrical impedance spectroscopy (EIS) [6]. Besides monitoring material behavior, EIS is becoming increasingly popular in bioimpedance monitoring [6-8].

However, with increasing reaction rates and complex material compositions, the classic DEA systems based on a single frequency have limited applicability. This paper addresses the issue using a multi-sine signal approach that provides high-density frequency information over high bandwidth. Based upon previous works [9, 10] an improved method using partial sinus signals within a multi-sine signal is provided, achieving much higher frequency density with at the same time low crest factor numbers.

A quasi-spectroscopic monitoring system is provided by monitoring as many frequencies parallel to the base frequency. Thus, instead of using few frequencies as supporting points for models to predict the material behavior resulting in inaccuracies due to the model assumptions [11], the reaction kinetics that occur can be directly observed.

2. State of the Art

In the following a short introduction to the signal optimization criteria, the Crest-Factor, and also a brief overview of the optimization of multi-sine signals is given.

2.1. Introducing the Crest-Factor

A common metric used to evaluate and compare multi-sine signals in the time domain is the Crest-Factor (CF). The metric shows how much amplitude is consumed by a signal to introduce a certain amount of energy into a system [12, 16]. Higher values indicate harmonics, while low values for multi-sine signals imply no or little interferences between the specific frequencies. For our application, it is desirable to set up a multi-sine signal with the smallest possible CF. The CF is calculated as the ratio between the peak value of a signal and its effective (root mean square) value:

$$CF = \frac{U_{peak}}{U_{RMS}} \tag{1}$$

For a signal in the time domain $s(t)$ measured over a time interval $[0;T]$, the $CF(s)$ is calculated according to the following formula:

$$CF(s) = \frac{\max_{t\epsilon[0,T]} |s(t)|}{\sqrt{\frac{1}{T}\int_0^T |s(t)|^2 dt}} \tag{2}$$

2.2. Current Optimization of Multi-Sine Signals

So far, the optimization of multi-sine signals has only been looked at from one perspective. By fitting

the phase angles of the signals, one can considerably minimize the CF of the multi-sine signal. Ojarand et al. have been developing several optimization algorithms for this topic yielding good results. Their approach to minimizing the CF is both analytical and iterative. They give various formulas for good starting phases that are adapted to the frequency density. Based on these formulas, he presented, among other things, an iterative algorithm for further optimization of the initially calculated phases [13, 14]. Recent work from Shibasaki [15] compares several algorithms for multi-tone signal generation, resulting in close outcomes for different approaches to manipulating the phase angles. Yang et al. [16] obtained very low CF using their improved algorithm combining Schroeder's formula with Van der Ouderaa's (VDO) numerical algorithm using an optimized clip function. The focus is on reducing the calculation time for arbitrary signals but is limited to small bandwidths. Alternative approaches are the use of CIRP or binary signals. An interesting idea is proposed by Yang et al. [17] with the use of tri-level excitation signals (normalized amplitude with +1,0,-1) and a Nuttal window function.

Nevertheless, the results are limited for large frequency ranges (starting from 5 decades). Frequencies in a similar range can compensate for each other's peaks by conveniently choosing their phase angles. This works within a smaller or bigger time interval dependent on the frequencies. When considering a multi-sine signal with frequencies spanning over a wide range, the higher frequencies won't be able to compensate for peaks of the lower frequency components. Instead of reducing the maximum amplitude value of the multi-sine signal, the higher frequency components will only increase it, resulting in a higher CF.

2.3. Advantages of Time-varying Frequencies

To overcome the drawback of many higher frequency components, time-varying frequencies could be an approach. The best-known representatives of this type of signals are chirp signals. The idea behind chirp signals is to start at a specific frequency and then in- or decrease it over time. Therefore, these kinds of signals could replace for example 5 or more high-frequency components. So far, much research regarding chirp signals has been done, and many ways of analyzing them with Short-Time-Fourier-Transform (STFT) or similar methods have been presented over the last years. [18-20]

3. Approach for Modulation of Multi-sine Signals with Partial Frequencies

The goal was to construct a multi-sine signal with a high-frequency bandwidth and a small CF. Regarding the difficulties explained in the previous section, the higher frequency components have to be handled differently. Replacing these components with chirp signals would be a good approach, but keeping all frequencies fixed is desired for the considered application in DEA with the primary goal of increasing the overall information within a single period of the lowest frequency segment. Using only modulated signal composition would lead to blurred results at higher frequencies limiting the ability to visualize fast chemical reactions and respective phase changes. Hence, the main idea was only to add these signals within a shorter time interval and separate from each other to the final multi-sine signal. This would solve the problem described above of having the high frequencies only increasing the overall amplitude of the multi-sine signal. The remaining issue to check is how this technique of assembling an arbitrary signal affects the CF. This technique would only be reasonable if it positively affects the CF, i.e., reduces it.

Given a certain number of signals with fixed frequencies, amplitude values and phase angles, the procedure of modulating a multi-sine signal with partially added signals is as follows:

The first thing to do is to choose how many frequencies should be added partially to the multi-sine signal. Here, each of these partial signals should be added for at least 30 periods. The reason behind this is that this causes a lower error in discrete FFT analysis in a later course due to less spectral leakage. Therefore, one can calculate the maximum number of signals that could be added without overlap in a given time interval by calculating how much time it would need to represent a certain frequency for 30 periods and adding those times up.

Once the number is fixed, choose as many of the signals with the highest frequency among all. This means if the given frequencies are $f_1, \dots, f_n$ and $m < n$ of them are chosen to be added partially, then this applies to the frequencies $f_{n-m+1}, \dots, f_n$.

After that the lower frequencies $f_1, \dots, f_{n-m}$ have to be modulated, therefore the meantime signal is then $S(x) = \sum_{i=1}^{n-m} a_i \cdot \sin(2\pi f_i x + p_i)$, with a_i the amplitude values and p_i the phase angles. The corresponding discrete signal is s with a length of $len(s) = number\_of\_samples$.

To make it possible that the final multi-sine signal can be run multiple times in a row, apply a window function to s to reduce spectral leakage further. This results in a continuous transition to zero at the beginning and end of the signal. For maintaining the signal in large part, we use a specially constructed window function here. The idea is to adjust the first and last 5 % of the signal. This means to leave the inner 90 % of the signal as it is and to only modify the ends of it. The procedure for this is as follows:

1. $num = int(0.05 \cdot num\_of\_samples)$
2. $x = linspace(0, 1, num)$
3. $for\ i\ in\ range(num)$:
   a. $s[i] *= 0.5 \sin\left(\pi \cdot x[i] - \frac{\pi}{2}\right) + 0.5$
4. $x = linspace(1, 2, num)$
5. $for\ i\ in\ range(num)$:

a. $s[19 \cdot start\_end\_num + i] * = 0.5 \cdot \sin\left(\pi \cdot x[i] - \frac{\pi}{2}\right) + 0.5$

After this assembling part, the remaining signals with higher frequencies can be added partially. Again, it is recommended to add those to the inner 90 % of the multi-sine signal. Otherwise, the part of applying the window function above would have no effect. So, if the time-interval is $T = [0, t_1]$, consider $T_{partial} = [0.05 \cdot t_1, 0.95 \cdot t_1]$ as the time-interval in which the remaining frequencies can be added. First, compute the time-intervals $T_i, i = 1, \dots, m$, in which the partial signals should be added. As already described above, the length of these intervals should be at least the time frequency i needed for 30 full periods and they should not overlap. Then compute for each of these intervals T_i the corresponding partial signal S_i (respectively s_i for the discrete case) with frequency $f_i, i = n - m + 1, \dots, n$. Then apply the window function described above to each of the signals $s_i, i = n - m + 1, \dots, n$. Finally, modulate the partial signals $s_i, i = n - m + 1, \dots, n$, onto the multi-sine signal s. Within this process, the partial signals have to be added at their associated time-interval.

4. Results and Discussion

With the signals created as described in Section 3, we tested the effect of the partial signals on the quality via calculation of the CF and the relative error in DFT.

4.1. Improvement of the Crest-Factor

It had to be tested to what extent this method of modulating a multi-sine signal improves its quality. The comparison was made here with Newman's analytical optimization method. He set up a formula for a good choice of phase angles of a multi-sine signal. For this test, the phase angles were calculated like in Newman's formula, the frequencies were chosen logarithmically distributed and the amplitude values descending from 1 to 0.2. Depending on the number of signals used in total, the frequency range has to be adapted. For less than 30 signals, we used $[10^2, 10^6]$ and $[10^2, 10^7]$ otherwise. This choice is because the error in FFT analysis increases when there are too many similar frequencies in a multi-sine signal. The fundamental frequency was set to 125 MHz due to the measurement platform the signal generation is developed for. It can be seen in Fig. 1 that partially added signals have a respectable impact on the CF. The CF of every tested signal decreased considerably.

This test confirmed the assumption made in Section 2.2. There we assumed that the high frequencies can no longer compensate for the peaks of the lower frequencies due to the wide frequency range. Ultimately, the high-frequency components only increase the overall amplitude of the multi-sine signal and thus also increase the CF. Due to the partial modulation, these high frequencies are only modulated for smaller time intervals, ensuring that they do not overlap in time. This leads to precisely the desired result, a smaller CF.

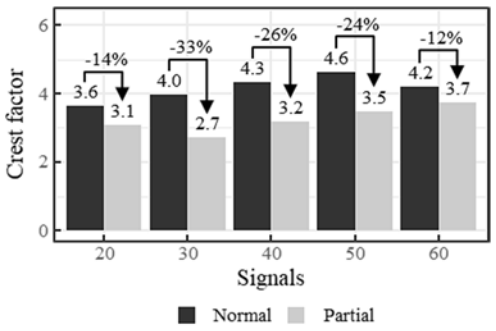


Fig. 1. CF minimization results.

4.2. Applying FFT to These Signals

In addition to improving the CF, we were interested to what extend this signal modulation affects the quality of the discrete FFT. To test this, the DFT results for a typically modulated multi-sine signal were compared to those of a partially modulated multi-sine signal.

For this test, multi-sine signals were first modulated without partial frequencies, and a DFT algorithm was applied. The amplitude values of the individual signals can be read off from the result of the DFT. Then, these were compared to the original amplitude values, which were used for the modulation of the multi-sine. The relative error between the read-off and original amplitude values was calculated for an optimal comparison independent of the quantity. As can be seen in Fig. 2, the relative error in such signals is very high. This is for two reasons. The first is that the DFT algorithm works suboptimally when there are many similar frequencies in the signal under consideration. It can be seen very clearly in Fig. 2 that the error increases as more frequencies are being used. The other reason is that the DFT algorithm only returns good results when there are enough data points for each frequency. This means that each frequency has to be modulated for a certain number of periods. In the best case, this should correspond to at least 100 periods. By design of our signals, the lower frequencies, in particular, are only modulated for 1-10 periods. Fig. 2 shows that the error is significantly higher at the lower frequencies.

For the second test, multi-sine signals with partially modulated frequencies were considered. After the higher frequency components have been added during a shorter time interval, the standard DFT can no longer be applied to the multi-sine signal to determine the correct amplitude values of those partial signals. Since the time interval in which those signals appear in the multi-sine signal is way shorter than T, the by DFT estimated values of their amplitude are accordingly smaller than initially.

To solve this problem, apply the standard DFT to the whole multi-sine signal to calculate the amplitude values of the lower frequency components. To estimate the amplitude values of the partially added higher frequency components, cut the discrete signal s into the parts where those signals have been added. Then apply the standard DFT algorithm to those shorter signals and filter the significant part. Filtering means going through the frequency array returned by the DFT algorithm and searching for the frequency that has been partially added to the cut signal. The corresponding amplitude value in the return of the DFT is the estimated amplitude value of this partial signal.

The second test could then be performed with this procedure. For this test, the multi-sine signals had the same configuration as described above. The same configuration means that the frequencies, amplitude- and phase- values have been chosen to be the same. We performed DFT on the signal as described above to calculate the relative error that appears in the amplitude values. From the result of the DFT, the calculated amplitude values can be read off and compared to the original value of the amplitude values. Again, this comparison was made based on the relative error.

A clear trend can be seen in Fig. 2 as to how the relative error behaves for typically modulated signals compared to signals with partial frequencies. In the lower frequency range, the modulation of both signals is the same. As a result, the errors in this area are also almost identical. Differences can only be seen in the higher frequency range. This is precisely the area in which frequencies were modulated partially. As a result, the relative error decreases by up to 75 % from normal to partially modulated multi-sine. When calculating an average value of the relative error in the DFT for the partially modulated signals, you get the same error values or even improvements compared to normally modulated multi-sine signals. The reason for the improved error values for multi-sine signals with partially modulated frequencies has already been explained in Section 4.2. In short, it is due to the way the DFT algorithm works. The more similar the frequencies in the signal to be examined, the more difficult it is for the DFT algorithm to correctly separate them within the multi-sine signal. If similar frequencies are modulated onto the multi-sine signal at different times, as tested here, the relative error in the result of the DFT is reduced.

Another trend that can be seen in Fig. 2 is that the relative error of the DFT in the low-frequency range increases the more frequencies are used in total for the multi-sine signal. The error in these tests rises to 60 %. For more than 60 signals, even higher errors can be expected. As a result, it can be taken that, for technical reasons, the DFT has an upper limit on the number of signals that can or should be modulated into a multi-sine signal. Of course, the larger the frequency range, the more signals can be modulated. However, from a practical point of view, a certain maximum frequency is specified.

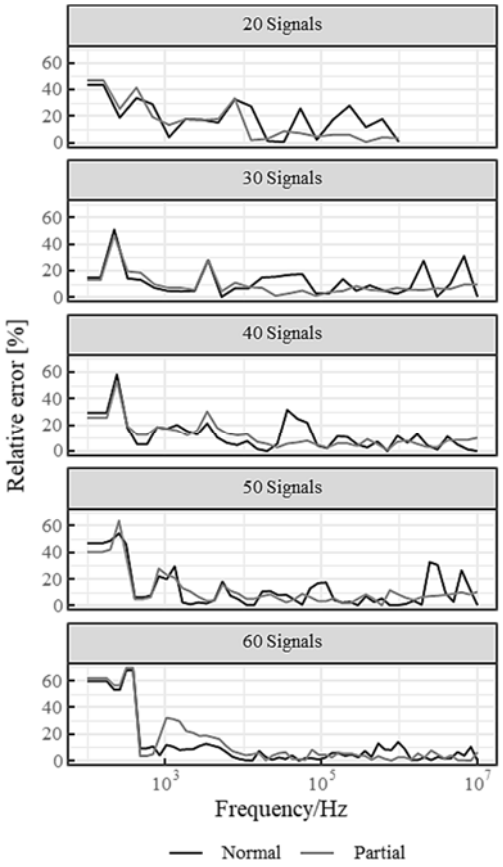


Fig. 2. Error for normally and partially modulated signals.

In addition, it was tested if the in Section 3 introduced window function works better than conventional windowing methods such as in [17]. These tests revealed significant improvements in the relative errors in DFT. This can be explained by the fact that the window function presented in Section 3 leaves a large part of the signal untouched. The signal is only adjusted in small areas at the beginning and end of the signal. This leads to smaller errors in the DFT since the frequencies in the signal are more clearly discernible than in signals on which a conventional window function has been applied.

5. Conclusion and Outlook

We could show that with the introduction of partially modulated signals our goals could be achieved. On the one hand it was possible to introduce more frequency blocks increasing the amount of information for time varying signals. On the other hand, the overall signal quality could be improved resulting in a much lower CF compared to non-partially modulated signals.

Our plans for future research are to analyze if other distributions of the partial signals have a big impact on the quality of the resulting multi-sine signal. Furthermore, it could be interesting to test how much more signal can be included in the multi-sine signal.

That means it could be tested to what extent the higher frequency components can be added multiple times to the final multi-sine signal without causing deterioration of the signal's quality. For this project, several constraints have to be specified and an algorithm to optimize the amount and distribution of the time intervals developed. Further improvements to reduce the relative error will focus on different analytical approaches using wavelet transformation.

Acknowledgements

This research was funded by the Bavarian Research Foundation, funding number: AZ-1466-20.

References

[1]. D. E. Kranbuehl, P. Kingsley, S. Hart, G. Hasko, B. Dexter, A. C. Loos, In situ sensor monitoring and intelligent control of the resin transfer moulding process, *Polymer Composites*, Vol. 15, Issue 4, 1994, pp. 299-305.

[2]. M. Eberhardt, S. Geinitz, R. Wendel, F. Henning, K. Drechsler, Simultaneous multi-frequency dielectric analysis of the polymerization process of anionic polyamide 6, in *Proceedings of the International Conference on Composite Materials (ICCM'19)*, Melbourne, Australia, 2019, pp. 2873-2884.

[3]. Y. Wang, W. Liu, Y. Qiu, Y. Wei, Single-component and fast-curing epoxy resin for liquid composite molding processes, in *Proceedings of the Conference on Composite Material, Polymer Science and Engineering (CPMSE'17)*, Toyama, Japan, 2017.

[4]. J. Xu, J. Yang, X. Liu, H. Wang, J. Zhang, S. Fu, Preparation and Characterization of Fast-Curing Powder Epoxy Adhesive at Middle Temperature, *Royal Society Open Science*, 2018.

[5]. R. Wendel, B. Thoma, F. Henning, Influence of water during manufacturing of aPA6 in the thermoplastic RTM process, in *Proceedings of the 33rd Polymer Processing Society Annual Meeting (PPS'17),* Cancun, Mexico, 2017.

[6]. V. Sirtoli, K. Morcelles, J. Gómes, P. Bertemes-Filho, Design and evaluation of an electrical bioimpedance device based on DIBS for myography during isotonic exercises, *Journal of Low Power Electronics and Applications*, Vol. 8, 2018, 50.

[7]. B. Sanchez, G. Vandersteen, I. Martin, D. Castillo, A. Torrego, P. J. Riu, J. Schoukens, R. Bragos, In vivo electrical bioimpedance characterization of human lung tissue during the bronchoscopy procedure. A feasibility study, *Medical Engineering & Physics*, Vol. 35, Issue 7, July 2013, pp. 949-957.

[8]. S. Abasi, J. R. Aggas, G. G. Garayer-Leyva, B. K. Walther, A. Guisseppi-Elie, Bioelectrical impedance spectroscopy for monitoring mammalian cells and tissues under different frequency domains: A review, *ACS Measurement Science*, 2022, to be published.

[9]. S. Geinitz, M. Eberhardt, D. Walser, D. Grund, B. Hollaus, A. Wedel, Simultaneous multi-frequency dielectric measurement technique for fast curing reactions, in *Proceedings of the 10th International Symposium on NDT in Aerospace (AeroNDT'18)*, Dresden, Germany, 24-26 Oct. 2018.

[10]. H. Althoff, M. Eberhardt, S. Geinitz, C. Linder, Advances in crest factor minimization for wide-bandwidth multi-sine signals with non-flat amplitude spectra, in *Proceedings of the 1st International Electronic Conference on Algorithms (IOCA'21)*, 2019, 27 Sep. – 10 Oct. 2021.

[11]. J. Ojarand, R. Land, M. Min, M. Rist, How many frequencies to use in electrical bioimpedance measurements, in Impedance Spectroscopy (O. Kanoun, Ed.), *De Gruyter*, 2019, pp. 161-168.

[12]. B. Sanchez, G. Vandersteen, R. Bragos, J. Schoukens, Basics of broadband impedance spectroscopy measurements using periodic excitations, *Meas. Sci. Technol.*, Vol. 23, 2012, 105501.

[13]. J. Ojarand, M. Min, Recent advances in crest factor minimization of multi-sine, *Electron. Electr. Eng.*, Vol. 23, 2017, pp. 59-62.

[14]. J. Ojarand, M. Min, P. Annus, Crest factor optimization of the multi-sine waveform for bioimpedance spectroscopy, *Physiol. Meas.*, Vol. 35, 2014, pp. 1019-1033.

[15]. Y. Shibasaki, K. Asami, R. Aoki, et al., Analysis and design of multi-tone signal generation algorithms for reducing crest factor, in *Proceedings of the IEEE 29th Asian Test Symposium (ATS'20)*, 2020, pp. 1-6.

[16]. Y. Yang, F. Zhang, K. Tao, B. Sanchez, H. Wen, Z. Teng, An improved crest factor minimization algorithm to synthesize multisines with arbitrary spectrum, *Physiol. Meas.*, Vol. 36, 2015, pp. 895-910.

[17]. Y. Yang, W. Zhang, F. Du, X. Tang, H. Wen, Z. Teng, Broadband bioimpedance spectroscopy based on a multifrequency mixed excitation and Nuttall windowed FFT algorithm, *Mathematical Problems in Engineering*, Vol. 2014, 2014, 539719.

[18]. C. K. Chui, Q. Jiang, et al., Analysis of an Adaptive Short-Time Fourier Transform-Based Multicomponent Signal Separation Method Derived from Linear Chirp Local Approximation, http://arxiv.org/pdf/ 2010.10760v1

[19]. J. Guo, H. Zou, X. Yang, G. Liu, Parameter estimation of multicomponent chirp signals via sparse representation, *IEEE Trans. Aerosp. Electron. Syst.*, Vol. 3, 2011, pp. 2261-2268.

[20]. M. Min, R. Land, T. Paavle, T. Parve, P. Annus, D. Trebbels, Broadband spectroscopy of dynamic impedances with short chirp pulses, *Physiological Measurement*, Vol. 7, 2011, pp. 945-958.

(4147)

Radio Map Reconstruction with Deep Neural Networks in a Weakly Labeled Learning Context with use of Heterogeneous Side Information

A. Malkova [1], M.-R. Amini [1], B. Denis [2] and C. Villien [2]
[1] Université Grenoble Alpes, LIG-APTIKAL, 700 Av. Centrale, 38401 Saint-Martin-d'Hères, France
[2] CEA-Leti, 17 Av. des Martyrs, 38054 Grenoble, France
E-mail: aleksandra.malkova@univ-grenoble-alpes.fr

Summary: Wireless localization is an essential component in Internet of Things applications. In this paper, we address the generalization problem of RSS map reconstruction for each city (usually, over gateways), relying on location-dependent radio measurements and taking advantage of side knowledge about the local region; like, city plan, terrain height, gateway position. We employ Neural Architecture Search to find an optimized Neural Network model with the best architecture for each of the supposed settings, depending on the amount of such prior side information. We show that using additional side information improves the final accuracy of the Received Signal Strength map reconstruction, especially in sub-areas close to the gateways where larger variations of the average received signal power are usually observed (ultimately, with a prominent beneficial impact onto positioning performance accordingly).

Keywords: Learning with weakly labeled data, Neural networks with optimized architecture, Radio map reconstruction.

1. Introduction

Retrieving the exact position of the connected objects has become an important feature of the Internet of Things (IoT). Such connected objects have indeed been widespread over the last few years thanks to the low cost of the radio integrated chips and sensors and their possibility of being embedded in plurality of the devices. For the location-dependent application and services these abilities to associate accurate location with physical data gives huge opportunities. One of the localization techniques is Global Positioning System (GPS) which has been widely used over the past decades, but it suffers from high energy consumption which is not suitable for IoT applications.

As an alternative, one can opportunistically measure location-dependent radio metrics, like Received Signal Strength (Indicator) (RSSI), Time (Difference) of Arrival (TDoA), Angle of Arrival (AoA), etc. Based on these metrics, there exist several methods to determine the node position: trilateration, triangulation, proximity detection, or fingerprinting. We will focus on fingerprinting [1] which requires (ideally) full map of mentioned above radio metrics, covering the zone of interest. However, collecting metrics in each point of the zone of interest is impractical and time costly in real-world scenarios, therefore most approaches rely on sparse and non-uniformly distributed measurements.

In this sense, classical map interpolation techniques such as Radial Basis Function (RBF) [2] or kriging [3] are used. Although these methods are relatively fast, they are quite weak in retrieving and predicting the complex and heterogeneous spatial patterns that are usually observed in real life signals. Another approach consists on deterministic simulation such as Ray-Tracing (RT) tools [4]. Given some real field measurements and then calibrated over them, these models predict the radio propagation while simulating electromagnetic interactions with the environment. These technologies, however, need a complete description of the environment (properties of the materials of the obstacles, shape, etc.); moreover, they are complex computationally. Recently, studies have employed machine learning for this task by considering radio maps as images and adapting neural network (NN) models that have been proposed for image completion. These models are based on the fully generated dataset by RT tools for predicting the signal propagation given the buildings mask and position of the transmitter [5]; or predicting the received power value for the LTE (Long Term Evolution) signal with use of additional information and NNs [6] with handcrafted structures.

In this work we will focus on the RSS map reconstruction, where only small amount of ground truth GPS-tagged measurements is available preventing to use existing NN models with handcrafted architectures and for which RT models could not be applied (due to the lack of information or high computational complexity). Our approach is based on *Neural Architecture Search (NAS)* [7] which aims to find an optimized NN model for this task – it is feasible to learn model parameters while exploring the architecture. In this paper, we provide:

A unified framework with the use of side information, for which we study the generalization ability of a neural network model which architecture is optimized over labeled and unlabeled data. This is an extension of the work of [8], where a NN was developed to exclusively find the RSSI measurements for a given unlabeled set without the use of side-information.

Empirical evaluation over two large-scale RSSI collections showing that the proposed approach is highly competitive compared to the SotA models in terms of quality metrics.

2. NAS with Genetic Algorithm for RSSI Map Reconstruction Using Side Information

Additional information could be represented in different manners, and they could be included into the algorithm in a variety of ways, such as independent channels, parallel channels inputs, directly in the learning goal, or in the ranking metric during model selection. We adapted the proposed algorithm presented in [8] for multi-channel input by combining additional context information with the data in the model's input; and we assessed the model's performance on unseen base stations that were not utilized in the learning process.

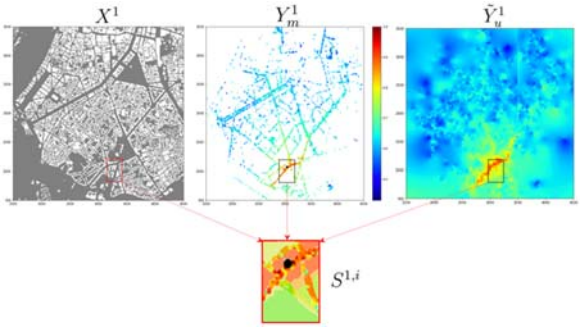


Fig. 1. An example of constituting the training sets for one base station. X^l corresponds to 2D node locations, buildings are shown in white. Y^l_m is RSSI map (true measurements); the base station is shown by a black circle, and $\tilde{Y}^l$ corresponds to interpolated points found by RBF. Colors depict the strength of the signal from dark red (highest) to deep blue (lowest) RSSI values. $S^{(i,j)} = S_l^{(i,j)} \cup S_v^{(i,j)} \cup S_u^{(i,j)}$ is one sub-matrix of partially labeled training data found from $Y^l_m \cup \tilde{Y}^l$.

Here, we supposed to have a small set of n available base stations $(X^j)_{1 \leq j \leq n}$. For each given matrix of base station X^j; $j \in \{1,...,n\}$, let $Y^j \in \mathbb{R}^{H \times W}$ be its corresponding 2D matrix of signal strength values measurements, where $H \times W$ is the size (in number of elements in a grid) of the zone of interest. In practice, we have access only to some ground truth measurements Y^j_m, meaning that $Y^j_m = Y^j \odot M^j$, with $M^j \in \{0, 1\}^{H \times W}$ a binary mask of available measurements, and $\odot$ is the Hadamard's product. Here we suppose sparsity meaning that the number of non-null elements in Y^j_m is much lower than the overall size $H \times W$. For each base station X^j we estimate unknown measurements $\tilde{Y}^l_u$ in Y^j with an RBF interpolation given (X^j_m, Y^j_m), so that we have a new subset $(X^j_u \tilde{Y}_u{}^j)$, where $X^j_m = X^j \odot M^j$ is the associated 2D node locations of Y^j_m in X^j, and the values in $\tilde{Y}^j_u$ are initially given by RBF predictions on X^j_u corresponding to the associated 2D node locations (or equivalently, the cell/pixel coordinates) with respect to the base station X^j which do not have measurements. In our semi-supervised setting, the values for unknown measurements in $\tilde{Y}^j_u$ will evolve by using the predictions of the current NN model during the learning process.

We further decompose the measurements set Y^j_m into two parts: Y^j_l (for *training*), Y^j_v (for *validation*), such that $Y^j_l \oplus Y^j_v = Y^j_m$, where $\oplus$ is the matrix addition operation. Let X^j_l, X^j_v be the associated 2D node locations of Y^j_l and Y^j_v in X^j.

In our experiments the number of base stations n is small, so in order to increase the size of labeled and pseudo-labeled training samples, we cut the initial measurements maps $(Y^j_m \oplus \tilde{Y}^j_u)_{1 \leq j \leq n}$ into smaller matrices which resulted into the sets $(S^{j,i})$: $1 \leq j \leq n$, $1 \leq i \leq m_j$ where the sets $S^{j,i} \subseteq Y^j_m \oplus \tilde{Y}^j_u$; $\forall\, i \in \{1,...,m_j\}$ are shifted with overlapping of the points. Each submatrix $S^{j,i}$ is hence divided into labeled, $S^{j,i}_l \cup S^{j,i}_v$, and pseudo-labeled (first interpolated points using RBF and then using the predictions of the current NN model) $S^{j,i}_u$. To each submatrix $S^{j,i}$ corresponds a 2D location $X^{j,i} \subset X$. Fig. 1 gives a pictorial representation of the notations.

Our objective is to find the optimal architecture for encoder-decoder architecture with skip connections using these side-information and study the generalization ability of obtained models for RSSI map reconstruction.

Side information. As additional context (or side) information, we have considered: (i) information about buildings presence, which was taken from the open-source OpenStreetMap dataset; (ii) number of crossed buildings by signal from base station (BS) to each point of the map; (iii) information about distance from the BS; (iv) information about the relief represented by DSM (digital surface model). This information was taken from the open-source dataset provided by Japan Aerospace Exploration Agency with 30 m accuracy.

Neural architecture search procedure. NAS is performed with a Genetic algorithm is similar to one presented in [8] as it gave better performance result in terms of obtained accuracy.

From the sets $(S^{i,j})$: $1 \leq j \leq n$, $1 \leq i \leq m_j$ we use an evolutionary algorithm similar to [10] for searching the most efficient architecture represented as a Direct Acyclic Graph (DAG). Here, the validation sets $(S_v^{i,j})$: $1 \leq j \leq n$, $1 \leq i \leq m_j$ are put aside for hyperparameter tuning. The edges of this DAG represent data flow with only one input for each node, which is a single operation chosen among a set of candidate operations. We consider usual operations in the image processing field, that are a mixture of convolutional and pooling layers. We also consider three variants of 2D convolutional layers with kernels of size 3, 5 and 7, and two types of pooling layers that compute either the average or the maximum on the filter of size 4.

Candidate architectures are built from randomly selected operations and the corresponding NN models are trained over the set $(S_l^{i,j})$: $1 \leq j \leq n$, $1 \leq i \leq m_j$ and its (possible) combinations with side information. The resulted architectures are then ranked according to pixel-wise error between the interpolated result of the outputs over $(S_v^{i,j})$: $1 \leq j \leq n$, $1 \leq i \leq m_j$ and interpolated

measurements given by RBF interpolation by filtering out the buildings. As error functions, we have considered the Mean Absolute Error (MAE) or its Normalized version (NMAE) where we additionally weight the pixel error according to the distance matrix value. Best ranked model is then selected for mutation and placed in the trained population. The oldest and worst in the rank are then removed to keep the population size equal to 20 models.

Once the NN model with the optimized parameters are found by NAS, f_θ, we consider the following two scenarios for learning its corresponding parameters θ by minimizing

$$\mathcal{L}(f_\theta, S_\ell \cup S^{j,i}_{\hat{y}_t}) = \\ = \frac{1}{n}\sum_{j=1}^{n}\frac{1}{m_j}\sum_{i=1}^{m_j}\Big(\frac{1}{|S^{i,j}_\ell|}\sum_{(x,y)\in S^{i,j}_\ell}\ell(y, f_\theta(x)) + \\ + \frac{1}{|S^{j,i}_{\hat{y}_t}|}\sum_{(x,\tilde{y})\in S^{j,i}_{\hat{y}_t}}\ell(\tilde{y}, f_\theta(x))\Big), \quad (1)$$

where *(x, y)* is a pair of pixel coordinate and corresponding value of this pixel, *l(·)* is a Mean Absolute Error (MAE).

We have considered two scenarios which relate to obtaining model parameters on *labeled and pseudo-labeled* measurements using just RBF interpolated data (scenario 1) or predictions from a first model learnt on these data (scenario 2). The overall learning process is depicted in Algorithm 1.

Algorithm 1: SL_{NAS}^{ind}

Input: A labeled training set with given measurements: $(X^j_m, Y^j_m)_{1\le j\le n}$ and an unlabeled set $(X^j_u)_{1\le j\le n}$

Init: Using $(X^j_m, Y^j_m)_{1\le j\le n}$, find interpolated measurements $(\tilde{Y}^j_u)_{1\le j\le n}$ over $(X^j_u)_{1\le j\le n}$ using the RBF interpolation method;

Step 1: Cut the initial measurements maps $(Y^j_m \bigoplus \tilde{Y}^j_u)_{1\le j\le n}$ into smaller matrices: $(S^{j,i})$: $1 \le j \le n$, $1 \le I \le m_j$;

Step 2: Search the optimal NN architecture using $(S^{j,i})$: $1 \le j \le n$, $1 \le I \le m_j$;

Scenario 1: Find the parameters θ^*_1 of the NN model f_θ:$\theta^*_1 = \text{argmin}_\theta \mathcal{L}(f_\theta, S_l \cup S^{j,i}_u)$ #(Eq. (1))

Scenario 2: Apply f_{θ^*1} on unlabeled data and obtain new pseudo-labeled measurements $S^{j,i}_u$ and find the new parameters θ^*_2 of the NN model f_θ:
$\theta^*_2 = \text{argmin}_\theta \mathcal{L}(f_\theta, S_l \cup S^{j,i}_u)$

Output: $f_{\theta^*_1}$ for scenario 1 or $f_{\theta^*_2}$ for scenario 2.

2. Experimental Results

Datasets. In the area of data-based research and in the field of machine learning singularly, it is usually hard to find large open-source datasets made of real data. In some works, however, alternatively (or as a complement) to using real data, synthetic data can be generated, for instance through deterministic simulations (as already mentioned in Introduction). As the two possible testing environments we considered two distinct datasets of LoRa signals in urban environments: Grenoble (which was collected during this work) with 4 BSs and Antwerp [9] with 9 BSs.

The size of maps is 700×700 pixels for the dataset collected in Antwerp and 368×368 for Grenoble. For that, we aggregated and averaged the power of collected measurements in cells/pixels of size 10 meters ×10 meters by the measured distance from base station location based on local ENU coordinates for Grenoble and Antwerp. After the aggregation, for Grenoble maps the number of available measurements in the zone of interest varies from 2728 to 7266, and for the Antwerp base stations the amount is from 2539 to 8277 (the details are shown in Table 1). As we also consider the generalization task, the algorithm should learn from all the available base stations data simultaneously, but to test its accuracy, we left from each city one base station that was not used in the training process.

Table 1. Summary over the settings for different cities: total amount of available measurements that were used during the learning process

City	Name	Number of points
Grenoble	BS_1	6264
	BS_2	2728
	BS_3	7266
	BS_4 (test)	6836
Antwerp	BS'_1	6060
	BS'_2	5606
	BS'_3	7548
	BS'_4	2539
	BS'_5	2957
	BS'_6	4940
	BS'_7	3154
	BS'_8	8277
	BS'_9 (test)	4335

In our settings, we only have access to several base stations lacking several orders of magnitude in size compared to aforementioned datasets. To artificially overcome this drawback, we created submatrices of the original images by cutting them into smaller ones (we tested over 96 by 96 pixels size because of memory issues during learning of the neural network for the storing of the model weights). We also added the flipped and mirrored images and we also did a shift in 20 pixels meaning that in our dataset there were overlapping between the images. Moreover, if the number of pixels with measurements in the initial cutted image was high enough (more than 3 % of the presented pixels) then we masked out the randomly sampled rectangle of presented measurements similar to the cutout regularization ([11]). By doing this we force the algorithm to do the reconstructions in the zones without measurements (not only locally) and be more robust to the amount of input data.

Matrices of the side information were used in the models as additional channels concatenated with measurements map. Before feeding the data into the algorithm, all the values have been normalized between 0 and 1 in each channel separately before cutting them into smaller sizes to feed into the models.

Evaluation of the results over held out base stations. To evaluate the result, we left one base station out of the initial set of each city to compare further the model's performances with baselines, namely test Antwerp and test Grenoble. To do this, all the points were divided into two parts, namely *train* and *test* points for 90 % and 10 % respectively. Moreover, to highlight the importance of the zones close to base station (as it was mentioned in the Introduction) we compare the performance of the algorithms over different considered circles around the base station location, namely 200 meters and 400 meters radiuses.

We considered state-of-the-art interpolation approaches which are: Total Variation (TV) in-painting by solving the optimization problem, Radial basis functions (RBF) with linear kernel that was found the most efficient, and the k Nearest Neighbours (kNN) regression algorithm. The evolutionary algorithm in the model search phase was implemented using the NAS-DIP package [12, 13]. All experiments were run on NVIDIA GTX 1080 Ti 11 GB GPU.

Results. In our experiments, we are primarily interested in addressing the following two questions: *(a)* does the use of side contextual information aid in the more accurate reconstruction of RSSI maps?; and *(b)* to what extent is the search for an optimum NN design effective in the two scenarios considered (Algorithm 1)?

We begin by envisaging scenario 1 of Algorithm 1 and investigating the impact of side information on the performance of the optimized NN model discovered by NAS. Fig. 2 shows the evolution of MAE inside a circular zone of varied radius for f_{θ^*} for different settings. The use of distance maps to supplement measurements improves predictions.

When the third component of side-information is included, such as height or building counts, we find that the elevation yields better signal estimations than the latter. This is understandable because signal transmission can be severely slowed by building heights.

As a best model obtained by Algorithm 1, scenario 1 we consider the case with three input channels: measurements, distances and elevations and present comparative results with other baselines in Table 2. The lowest errors are shown in boldface. The symbol ↓ denotes that the error is significantly greater than the best result using the Wilcoxon rank sum test with a p-value threshold of 0.01. According to these findings, $f_{\theta_1^*}$ outperforms other state-of-the-art models as well as the Deep Image Prior (DIP) [14] model with a handcrafted architecture. These results suggest that the search of an optimal NN model with side-information has strong generalization ability for RSSI map reconstruction.

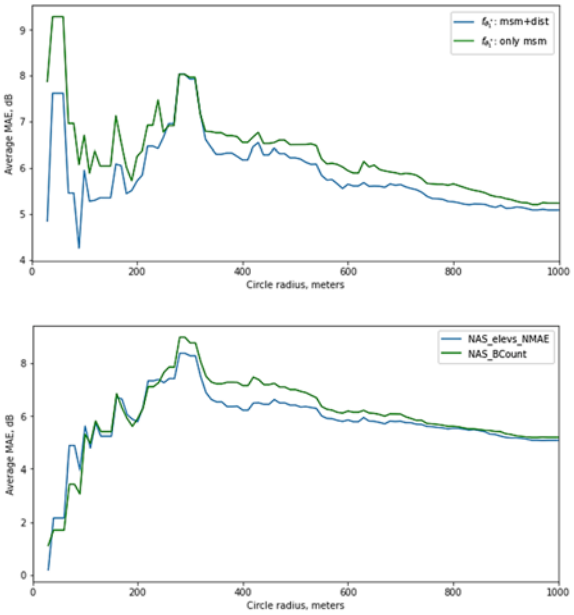


Fig. 2. Cumulative MAE distribution of f_{θ^*} according to the distance to the test base station for the city of Antwerp; with measurements and measurements with distance maps (top) and measurements, distance maps, and building counts or elevation (bottom)

Fig. 3 depicts the average MAE in dB of all models and $f_{\theta_1^*}$ with (iii) and (iv) as well as the NN model $f_{\theta_2^*}$ corresponding to additional learning with pseudo-labeled examples, with respect to the distance to the test BS for the city of Antwerp. For distances between 200 and 400 meters, $f_{\theta_2^*}$ consistently outperforms in terms of MAE. Fig. 4 presents the average MAE in dB of all models with respect to the distance to the test base station BS_4 for the city of Grenoble. These results are consistent with those obtained over the city of Antwerp.

As in paper [8], these findings imply that self-training constitutes a promising direction for RSSI map reconstruction.

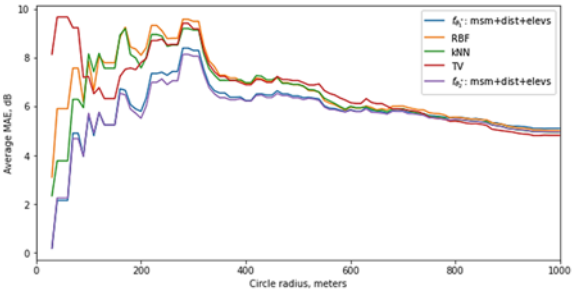


Fig. 3. Average MAE [dB] of all models with respect to the distance to the test BS for the city of Antwerp.

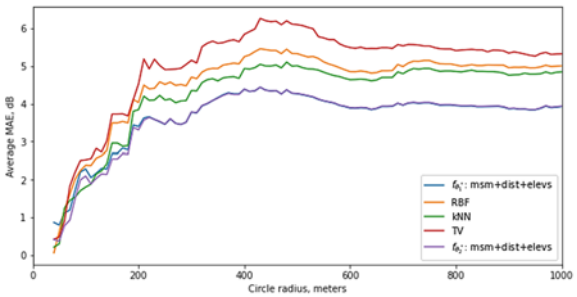


Fig. 4. Average MAE in dB of all models with respect to the distance to the test BS for the city of Grenoble.

Table 2. Comparisons between baselines in terms of MAE with respect to the two distances to Antwerp' test base station BS'$_9$ (top), and Grenoble test base station BS$_4$ (bottom). Best results are shown in bold.

Model	MAE, dB, 200 m	MAE, dB, 400 m
RBF	8.34$^{\downarrow}$	7.04$^{\downarrow}$
kNN	7.98$^{\downarrow}$	7.08$^{\downarrow}$
TV	7.50$^{\downarrow}$	6.97$^{\downarrow}$
DIP	6.55$^{\downarrow}$	6.63$^{\downarrow}$
$f_{\theta^{*1}}$	**5.88**	**6.37**

Model	MAE, dB, 200 m	MAE, dB, 400 m
RBF	4.03$^{\downarrow}$	5.29$^{\downarrow}$
kNN	3.84$^{\downarrow}$	4.92$^{\downarrow}$
TV	4.53$^{\downarrow}$	5.91$^{\downarrow}$
DIP	4.64$^{\downarrow}$	5.50$^{\downarrow}$
$f_{\theta^{*1}}$	**3.40**	**4.32**

4. Conclusions

In this paper, we proposed to find an optimal Neural Network model for RSSI map reconstruction using NAS, showing its generalization ability on two distinct datasets.

References

[1]. W. Choi, Y.-S. Chang, Y. Jung, J. Song, Low-power lora signal-based outdoor positioning using fingerprint algorithm, *ISPRS International Journal of Geo-Information*, Vol. 7, Issue 11, 2018, 440.

[2]. A. Enrico, C. Redondi. Radio map interpolation using graph signal processing, *IEEE Communications Letters*, Vol. 22, Issue 1, 2018, pp. 153-156.

[3]. M. A. Oliver, R. Webster, Kriging: a method of interpolation for geographical information systems, *International Journal of Geographical Information System*, Vol. 4, Issue 3, 1990, pp. 313-332.

[4]. S. Sorour, Y. Lostanlen, S. Valaee, K. Majeed, Joint indoor localization and radio map construction with limited deployment load, *IEEE Transactions on Mobile Computing,* Vol. 14, Issue 5, 2015, pp. 1031-1043.

[5]. R. Levie, C. Yapar, G. Kutyniok, G. Caire, Pathloss prediction using deep learning with applications to cellular optimization and efficient d2d link scheduling, in *Proceedings of the International Conference on Acoustics, Speech, & Signal Processing (ICASSP'20)*, 2020, pp. 8678-8682.

[6]. T. Nagao, T. Hayashi, Study on radio propagation prediction by machine learning using urban structure maps, in *Proceedings of the 14th European Conference on Antennas and Propagation (EuCAP'20)*, 2020, pp. 1-5.

[7]. T. Elsken, J. H. Metzen, F. Hutter, Neural architecture search: A survey, *Journal of Machine Learning Research*, Vol. 20, Issue 55, 2019, pp. 1-21.

[8]. A. Malkova, L. Pauletto, C. Villien, B. Denis, M.-R. Amini, Self-learning for received signal strength map reconstruction with neural architecture search, in *Proceedings of the International Conference on Artificial Neural Networks (ICANN'21)*, 2021, pp. 515-526.

[9]. M. Aernouts, Localization with low power wide area networks, PhD Thesis, *University of Antwerp*, 2022.

[10]. E. Real, A. Aggarwal, Y. Huang, Q. V Le, Regularized evolution for image classifier architecture search, in *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33, 2019, pp. 4780-4789.

[11]. T. Devries, G. W. Taylor, Improved regularization of convolutional neural networks with cutout, *Computing Research Repository (CoRR)*, 2017, abs/1708.04552.

[12]. K. Ho, A. Gilbert, H. Jin, J. Collomosse, Neural architecture search for deep image prior, *Computers & Graphics*, Vol. 98, 2021, pp. 188-196.

[13]. Repository for Neural Architecture Search for Deep Image Prior, https://github.com/Pol22/NAS_DIP)

[14]. D. Ulyanov, A. Vedaldi, V. Lempitsky, Deep image prior, *International Journal of Computer Vision*, Vol. 128, 2020, pp. 1867-1868.

(4180)

Ear Tag Type Tympanic Temperature Sensor for Cattle Considering Communication Range and Its Application

T. Natori [1], H. Nabenishi [2] and N. Aikawa [3]
[1] Tokai University, 9-1-1 Toroku, Higashiku, Kumamotoshi, Kumamoto, 862-8652 Japan
[2] Kitasato University, 35-1 Higashi-23bancho, Towadashi, Aomori, 034-8628 Japan
[3] Tokyo University of Science, 6-3-1 Niijuku, Katsushikaku, Tokyo, 125-8585 Japan
Tel.: +81-096-386-2703
E-mail: t_natori@tsc.u-tokai.ac.jp

Summary: Cattle body temperature is an important source of information in health management, such as predicting estrus and calving time, and detecting fevers in calves. Since it has been shown that body temperature in cattle can be measured near the tympanic membrane in addition to the rectum, the ear tag type temperature sensors for measuring external auditory canal temperature (EACT) are commercially available and used in various studies. In this paper, we develop the ear tag type temperature sensor whose communication range can be easily extended to enable real-time temperature data acquisition. To confirm the usefulness of the developed temperature sensor, seven cows were actually fitted with the sensor and the calving time was predicted using the measured EACT. The results showed that 100 % certainty of notification could be achieved up to 12 hours before calving.

Keywords: Cattle, Tympanic temperature, Calving time prediction, Ear Tag, Sensor, IoT.

1. Introduction

Cattle body temperature is an important source of information in health management, such as predicting estrus and calving time, and detecting fevers in calves. In particular, parturition is an important event for farmers in terms of productivity, and the time of parturition is effective information for preventing parturition accidents. It is generally known that the body temperature of the cattle decreases by 0.2 to 0.5 degree 24 hours before calving. Burfeind et al. measured rectal and vaginal temperatures and reported the details of the temperature difference between the time of calving and 24 or 48 hours before the time of calving [1].

Generally, body temperature of cattle is measured by inserting a thermometer into the rectum and vaginal, however, daily measurement for health management is a very labor-intensive task. As for body temperatures outside the rectum, Guidry et al. investigated the relationship between temperature near the tympanic membrane and rectal, and found that temperature sensed at the tympanic membrane is more suitable than rectal temperature for determining the speed of response of cattle to both internal and external temperature changes [2]. In other words, a system that can easily measure the temperature near the tympanic membrane (EACT: External Auditory Canal Temperature) would enable health management while reducing the burden on farmers.

Several ear tag type temperature sensors are commercially available and studies using them have been reported. Mahendran et al. used an ear tag type temperature sensor called TempVerified (FeverTag, Inc.) to investigate the effects of specific treatments on febrile cattle [3]. TempVerified is measured the EACT every 10 minutes. If the temperature exceeds the threshold value for six hours, an LED inside the device blinks to indicate an abnormality. Randi et al. also investigated the relationship between the EACT and the estrus and ovulation using a commercially available ear tag type temperature sensor called Boviminder [4]. Boviminder measures the EACT every 10 minutes and uploads the data to a base station using frequencies in the ISM band. When the Boviminder exceeds the communication range of the base station, the data is stored in the sensor, and when the sensor returns to the communication range, the data is uploaded again.

In order to solve shortcoming of Boviminder about the communication range, this paper describes the development of an ear tag type temperature sensor that can easily extend the communication range and maintain real-time performance. By using the measured the EACT to predict the calving time, we show that the developed sensor is useful and that the EACT can also be used to predict the calving time.

2. Development of Ear Tag Type Tympanic Temperature Sensor

In this section, we describe an ear tag type temperature sensor for measuring the EACT. Since the EACT is measured by inserting a thermistor directly into the external auditory canal, it is necessary to consider that the external auditory canal of cattle is generally 0.6 cm in diameter and 10-13 cm in length [1]. In addition, the system should have a wireless function to transmit temperature data to the server after acquiring temperature data by AD conversion of the voltage generated in the thermistor. Furthermore, the

communication range must be easily extended to maintain real-time performance, and the power consumption must be low to avoid battery replacement, which can be a burden for farmers. To satisfy these requirements, we select "NXFT15WF104FA2B05" (manufacturer: Murata Manufacturing Co., Ltd.) [5] for the thermistor, "TWELITE" (manufacturer: Mono Wireless Inc.) [6] for the microcontroller, and "CR2032" coin-type batteries for the power supply as components for the sensor developed in this paper. TWELITE is equipped with a 2.4 GHz wireless function and can be used as a repeater simply by changing the settings. In practice, we used a USB-type repeater with TWELITE built in, which can be powered by USB, making it easy to extend the communication range as long as a power supply is available. TWELITE is also characterized by its extremely low power consumption, with a CR2032 battery life of 10 years when the transmission interval is set at 30 seconds.

Fig. 1 shows the developed ear tag type temperature sensor. The lower right circuit board is also equipped with an LED, which enables the sensor to notify abnormalities by light, as in TempVerified. The sensor housing is fixed to a commercially available ear tag with screws. The thermistor is covered with a silicon tube to reduce the strain on the external auditory canal. Fig. 1 shows the ear tag type sensor for calves, however, by extending the thermistor and tubing, it can be used as a sensor for cattle without changing the sensor housing.

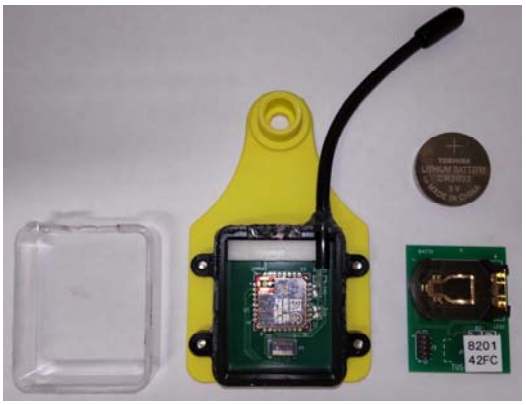


Fig. 1. Ear tag type temperature sensor.

Next, we will discuss the collection system for temperature information. Fig. 2 shows an actual data collection system installed on a farm. The cattle are equipped with the ear tag type temperature sensor shown in Fig. 1. The measured temperature data are collected by a computer for data collection via a repeater. The computer is connected to the Internet, so the temperature data can be viewed as needed.

3. Calving Time Prediction

To confirm the usefulness of the developed ear tag type temperature sensor, the sensor was attached to the ear of cattle with the EACT acquisition interval set to 10 minutes, and the calving time was predicted from the obtained EACT.

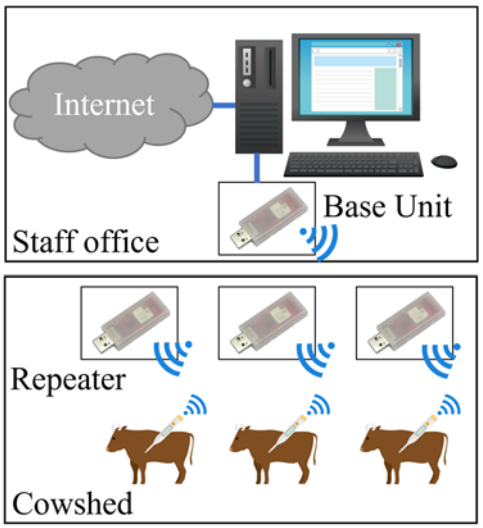


Fig. 2. Data acquisition system.

It is generally believed that the body temperature of cattle decreases about 0.5 ℃ 24 hours before calving. Preliminary experiments have shown that the EACT also decreases before calving. Therefore, a calving prediction algorithm is defined to predict that calving will occur after *t* hours when the difference between the average EACT for the past 3 days and the current EACT is less than *th* degrees. Since the result of calving prediction changes when *th* and *t* are changed, ROC (Receiver Operating Characteristic) analysis is performed to clarify the relationship between the two.

For the ROC analysis, we classify the observed EACT data as the positive or negative data. First, the positive data is the EACT related to parturition, and define as the difference between the average EACT for the past 3 days and the EACT at *t* hours before parturition. Second, the negative data is the EACT unrelated to parturition, and define as the difference between the average EACT for 3 days from wearing the EACT sensor and the EACT for the next day. Finally, the parturition prediction algorithm defined above is applied to the positive data and the negative data for the prediction of parturition.

The outcome of the parturition prediction and the positive / negative data can be summarized in a confusion matrix as shown in Table 1. From the confusion matrix, the sensitivity = TP / (TP + FP) and the false positive rate = FP / (FP + TN) can be calculated. The ROC curve shown in Fig. 3 is a plot of the sensitivity and false positive rate calculated for different *th* degrees. From the ROC curve, it can be seen that if the *th* degrees value nearest to the upper left of the graph is selected, the probability of parturition within *t* hour is the highest because the ideal *th* degrees is when the sensitivity = 1 and the false positive rate = 0.

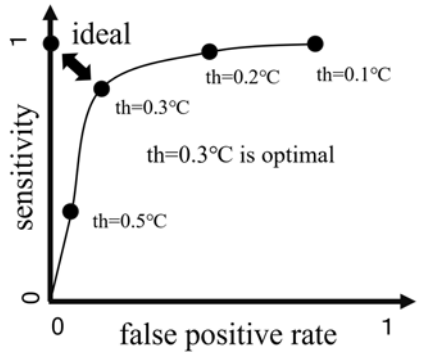


Fig. 3. ROC curve.

Table 1. Confusion matrix.

	Parturition prediction	
	Parturition occurs after *t* hours	Parturition doesn't occur after *t* hours
Positive data	TP(True Positive)	FP(False Positive)
Negative data	FN(False Negative)	TN(True Negative)

Table 2 shows the sensitivity (S) and false positive rate (FPR) of the EACT of seven pregnant Japanese Black cattle for each *th* when *t* = -22 h (calving at 22 hours) and *t* = -12 h (calving at 12 hours) and Fig. 4 explain the ROC curve it.

Table 2. Sensitivity and false positive rate when t = -12 h or t = -22 h.

th (℃)	t = -12 h		t = -22 h	
	S (%)	FPR (%)	S (%)	FPR (%)
0.1	100.0	42.9	100.0	14.3
0.2	100.0	28.6	57.1	0.0
0.3	100.0	0.0	28.6	0.0
0.4	85.7	0.0	0.0	0.0
0.5	71.4	0.0	0.0	0.0
0.6	57.1	0.0	0.0	0.0
0.7	42.9	0.0	0.0	0.0
0.8	14.3	0.0	0.0	0.0
0.9	14.3	0.0	0.0	0.0
1.0	0.0	0.0	0.0	0.0

Fig. 4 shows the ROC curves at *t* = -22 h and *t* = -12 h when th is varied from 0.1 to 1.0 in increments of 0.1. The figure shows that the upper left corner of the graph is closer to the upper left corner for *t* = -12 h than for *t* = -12 h.

Table 2 shows that when *t* = -22, all cattle can be notified at least 22 hours before calving without missing any cattle if *th* is set to 0.1, but there is a 14.3 % false positive rate. When *th* is set to 0.2, false notifications are eliminated, but the sensitivity drops to 57.1 %, so that about half of the cattle cannot be notified at least 22 hours before calving. When *t* = -12 h, setting *th* to 0.3 allows 100 % accuracy of notification up to 12 hours before calving without false notifications. This may be due to the fact that the EACT decreased further than at *t* = -22 h, which facilitated discrimination by the threshold value.

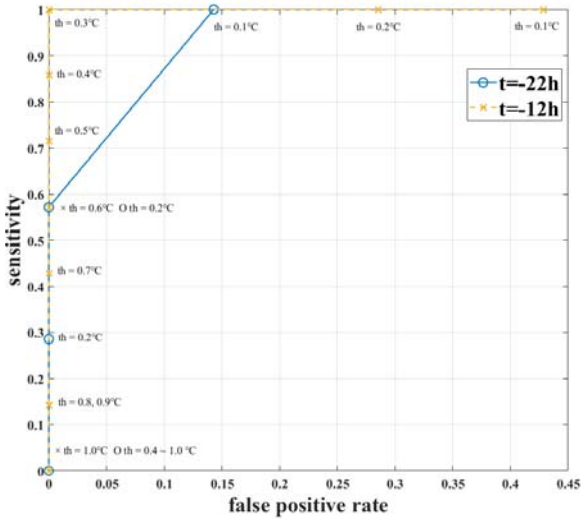


Fig. 4. ROC curve when "t" = -12 h or "t" = -22 h.

4. Conclusions

This paper describes an easy method for developing the ear tag type temperature sensor. The developed ear tag type temperature sensor was actually attached to a cow, and the data collection method was described. The collected temperature data were used to predict calving, and it was shown that calving can be predicted by the EACT.

Acknowledgements

This study was funded by Japan Livestock Technology Association (Livestock and Dairy Production Enhancement Projects.

References

[1]. O. Burfeind, V. S. Suthar, R. Voigtsberger, S. Bonk, W. Heuwieser, Validity of prepartum changes in vaginal and rectal temperature to predict calving in dairy cows, *Journal of Dairy Science*, Vol. 10, Issue 94, 2011, pp. 5053-5061.

[2]. A. J. Guidry, R. E. McDowell, Tympanic membrane temperature for indicating rapid changes in body temperature, *Journal of Dairy Science*, Vol. 49, Issue 1, 1966, pp. 74-77.

[3]. S. A. Mahendran, R. Booth, M. Burge, N. J. Bell, Randomised positive control trial of NSAID and

antimicrobial treatment for calf fever caused by pneumonia, *Veterinary Record*, Vol. 181, Issue 2, 2017, pp. 45-45.

[4]. F. Randi, M. McDonald, P. Duffy, A. K. Kelly, P. Lonergan, The relationship between external auditory canal temperature and onset of estrus and ovulation in beef heifers, *Theriogenology*, Vol. 110, 2018, pp. 175-181.

[5]. Murata Manufacturing, NXFT15WF104FA2B050, https://www.murata.com/ja-jp/products/productdetail?partno = NXFT15WF104FA2B050

[6]. Mono Wireless Inc., https://mono-wireless.com/jp/products/TWE-LITE/index.html

(4207)

A Deep Learning Vision Transformer Using Majority Voting Ensemble for Nematode Classification with Test Time Augmentation Technique

Abdul Qayyum [1], Abdesslam Benzinou [1], Kamal Nasreddine [1], Valentin Foulon [1], Anthonin Martinel [1], Catherine Borremans [2], Randi Hagemann [3] and Daniela Zeppilli [2]
[1] ENIB, UMR CNRS 6285 LabSTICC, 29238, Brest, France
[2] Ifremer, 29280, Plouzané, France
[3] Equinor Energy AS (Company), Norway
E-mail: qayyum@enib.fr

Summary: Nematode worms are one of the most abundant metazoan groups on the earth, occupying diverse ecological functions. Accurate recognition or identification of nematodes is of great importance for pest control, soil ecology, biogeography, habitat conservation, and climate change. Image processing and computer vision-based methods provided a few successes in species recognition of nematodes, however, the lack of a publicly available imaging dataset for diverse identification species of nematodes and the lack of a standard benchmark of state-of-the-art deep learning techniques demands more research on automatic nematode identification. In this paper, we propose a deep learning-based vision transformer model for automatic nematode identification. Further, the post ensemble majority voting technique was used to get the final prediction for nematode classification for 19 classes. The test-time augmentation technique and post-max majority voting ensemble prediction improve the performance of the proposed transformer model. We set up a species recognition benchmark by employing state-of-the-art deep learning networks. We discuss the experimental results, compare the recognition accuracy of different networks, and show the performance comparison of various deep learning models.

Keywords: Vision transformer, Nematode identification, CNN models, Post-ensemble, Majority voting, Test time augmentation.

1. Introduction

Nematode investigation is important due to the following reasons: (1) parasitic nematodes threaten the health of plants on a global scale, causing at least 80 billion US dollars loss per year [1]; (2) interstitial nematodes pervade sediment and soil ecosystems in overwhelming numbers [2], and (3) Caenorhabditis Elegans (C. Elegans) is a favorable experimental model system [2].

Nematodes have millions of different species with diverse organisms and are extremely abundant. This leads to routine the task of identifying nematodes at the species level which is complicated and time-consuming. Rapid identification and accurate recognition of nematodes have great value for soil ecology, pest control, habitat conservation, climate maintenance, and biogeography [3]. Due to vague diagnostic characteristics [4], high morphological variations within a single species, and frequently encountered adolescents [5], the identification of nematodes is a challenging task. Moreover, manual identification of nematodes is labor-intensive and often requires solid training and well-established experience. In a conclusion, the taxonomy of nematode species appears to be a substantial bottleneck for numerous relevant fields. Molecular barcoding [6] has been used as a powerful tool for biodiversity investigation and species identification, however, it heavily depends on the availability of molecular biology facilities, background knowledge, and sequencing instruments. Alternatively, imaging-based nematode identification such as nematode classification [7-9], and nematode (eggs) counting [10], has been used for nematode research. This nematode identification research is based on either nematode image stacks [8, 9], or mainly designed for the model organism C. elegans (Roundworm) [7]. Very few classes involved in the image stacks and low diversity coverage, i.e., species limited to the laboratory cultured C. elegans are the main issues for the identification of nematode task. More importantly, very few research groups are open publicly to the imaging dataset for nematode research and a standard benchmark is still missing. For the automatic and real-time identification of nematodes, we need to develop state-of-the-art deep learning models.

In this paper, the following contribution are addressed:

i. Various vision-transformer-based models are trained for nematode classification. The 10 different vision transformers have been used as a post-ensemble using max-majority techniques to get the final prediction. The test time-augmentation technique has been used after the post-ensemble method to further enhance the performance of the proposed model.
ii. Various CNN pre-trained models have been trained on the nematode dataset and compared the results with proposed vision-based transformer models.
iii. Different performance metrics have been used to assess the performance of proposed models with state-of-the-art methods.

2. Related Work

A few related works have been established for nematode classification, an image stacks, with each stack (or each species) consisting of multiple focal planes of the specimen [8, 9, 11]. Various techniques have been used such as the 3D X-Ray Transform based method [11], 3D X-Ray Transform based multilinear analysis method [12], projection-based methods 8, information fusion-based approaches [13], deep convolutional neural network (CNN) image fusion based multilinear approach [9].

Javer et al. [7] proposed a fully convolutional neural network to distinguish diverse strains of C. elegans using recorded spontaneous activity Zhou et al. [14] proposed an identification method for mutant types and other types using locomotion patterns. Currently, various deep convolutional neural networks (CNNs) have been used as powerful tools for data-driven research. Various state-of-the-art CNN models are trained on the ImageNet dataset and can use be used by fine-tuning or transfer learning to other imaging-based datasets. These pre-trained models are VGG [15], ResNet [16], DensNet [17], and EfficienNet [18] and vision transformer [19]. They are well established for image classification tasks and can learn discriminative features from input images. Hence, Recently, the authors used various pre-trained convolutional neural networks for nematode classification.

3. Material and Methods

3.1 Data Pre-processing

The dataset used in this research is publicly available and can be downloaded from [20]. The dataset has been split into training and validation using the 80/20 ratio. We have used the same splitting of training and testing datasets for a fair comparison with state-of-the-art deep learning models. Data normalization and horizontal and vertical flips were used as data augmentation to train the deep learning models. For training and optimization of the proposed model, all images are resized to 224×224. A few samples of the dataset are shown in Fig. 1.

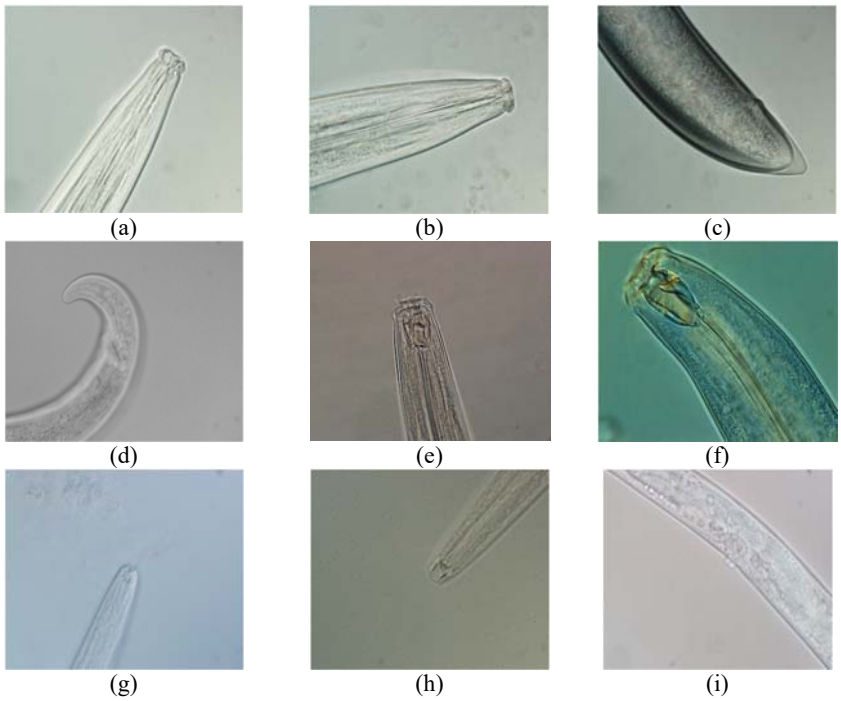


Fig. 1. Few examples of nematode images used in this study. (a-c): Aporcelaimus sp.; (d-f): Mylonchulus sp.; (g-i): Pristionchus sp.; (a), (b), (e), (f), (g), (h): head region; (c, d): tail region; (i): middle body.

3.2. Proposed Deep Learning-based Vision Transformer Model

Transformer architecture has become the de-facto standard for natural language processing tasks, and its applications to computer vision remain limited. Recently, Dosovitskiy et al. [19] proposed a vision-based transformer without any modification in convolutional neural networks and completely replaced convolutional layers with a transformer. A pure transformer was used directly to sequences of image patches and performed very well on image classification tasks. Due to the Self-attention-based mechanism, the transformers [19] have gained popularity in natural language processing (NLP). The transformer produced better performance in NLP and now gained much popularity in image classification tasks with small modifications.

The image is split into patches and provides the sequence of linear embedding of these patches as an input to a Transformer. The authors trained the variety of Transformers in a supervised way on the image

classification task. When trained the Vision Transformer (Vit) was at sufficient scale and transferred to tasks with fewer data points, it achieved excellent performance. The pre-trained transformer models are trained on ImageNet and the public ImageNet-21k dataset. The transformer models achieved accuracy of 88.55 % on ImageNet, 90.72 % on ImageNet-ReaL, 94.55 % on CIFAR-100, and 77.63 % on the VTAB suite 5 of 19 tasks [19]. Based on the success of the transformer model on the ImageNet dataset.

We have chosen a vision transformer for the automatic classification of nematodes. We have trained various transformer models using different hyperparameters on the nematode dataset. The post ensemble approach has been used on different prediction values based on the max-majority voting technique to classify the nematode dataset. Based on experimental results, 10 output predictions obtained from various trained transformer models are used in the max-majority voting technique to get the final prediction. Furthermore, the test-time augmentation (TTA) technique was used to enhance the performance of the proposed model. We gained 1.2 % performance with TTA techniques. To measure the performance of classical machine learning models, we have extracted features from the last layer of the proposed transformer and used these deep features in traditional machine learning models to compare the classification score with the deep learning-based transformer model. We have compared the performance of the proposed transformer with 4 state-of-the-art CNNs which are VGG-19, EfficienetNet-b3, DensNet201, and ResNet-101 models. The comparison with the CNNs model will be discussed in the results section. The overall proposed vision transformer for nematode classification is shown in Fig. 2.

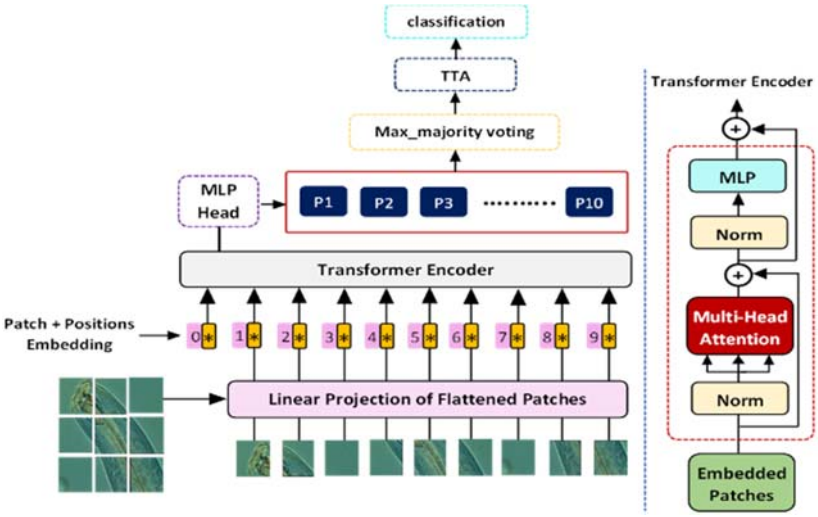


Fig. 2. Flow chart of the proposed solution for nematode classification based on an ensemble vision transformer using the max-majority voting and TTA technique. Where P1 is the prediction probability of the proposed model, P2 is the second prediction, and so on. We took P1—P10 predictions from a single proposed model and ensemble.

3.3. Training and Optimization of the Transformer Model

We have used pre-trained transformers that are initially trained on ImageNet and ImageNet-21k and we have fine-tuned the nematode dataset. The classifier layer has been modified with 19 classes and added a few regularization layers. The Adam optimizer with a 0.0001 learning rate has been used to optimize the weights of the transformers. 32 batch-size with 500 epochs used for training the transformer. The cross-entropy loss function with class weighting has been used. The dataset is highly imbalanced and weighted cross entropy tackled the class imbalance issue. The higher weights are assigned to those classes that have a lower number of samples and are assigned lower weights for a higher number of classes.

The early stopping criteria have been applied for training the proposed model. We trained all models using NVIDIA RTX 1080.

4. Results and Discussion

Accuracy, precision, recall, region of convergence (ROC), and Precision-recall curves (PR) have been used to evaluate the transformer's deep learning models.

Table 1. The proposed majority voting ensemble-based Vis-Transformer model achieved better performance as compared to CNN's based models as well as the simple Vis-Transformer model. Further, TTA was applied to the proposed majority voting ensemble-based Vis-Transformer model to achieve the highest performance. We have applied a two-pair t-test to compare the performance of the proposed algorithms. The lower value produced by our proposed model shows a significant difference and achieved better performance.

In Fig. 3, we give the ROC curve (receiver operating characteristic curve) showing the performance of the classification model at all

classification thresholds. Lowering the classification threshold classifies more items as positive, thus increasing both False Positives and True Positives. The area under the curve (AUC) is the probability that the model ranks a random positive example more highly than a random negative example. The ROC curves show the trade-off between the true positive rate and the false-positive rate for a predictive model using various probability thresholds. High True Positive Rates of a particular class represents that the model has performed well in classifying that particular class.

Table 1. Overall performance of proposed transformer and existing state-of-the-art models.

Methods	Overall Accuracy (%)	Precision	Recall	F-score
Vis-Transformer model	0.8062	0.8043	0.8134	0.8099
Majority voting Vis-Transformer model	0.8162	0.8122	0.8063	0.8129
Majority voting Vis-Transformer model+TTA	**0.8282**	0.8210	0.8133	0.8149
ResNet101 model	0.7753	0.7645	0.7845	0.7788
EfficienetNet-b3 model	0.7599	0.7477	0.7566	0.7463
DensNet201	0.7976	0.7833	0.7793	0.7839
VGG19	0.7422	0.7333	0.7411	0.7310
Xuequan Lu et al. [20]	79.0	--	--	--

The graph shows that the proposed transformer model produced higher AUC curves as compared to other state-of-the-art CNNs models and we can say that our proposed model performed a better job of classifying the positive class in the dataset. In the very imbalance classes dataset, Precision-Recall (PR) curves provided a useful measure of success in a predictive model. Precision is a measure of result relevancy, while recall is a measure of how many truly relevant results are returned. The precision-recall curve shows the trade-off between precision and recalls for different thresholds. A high area under the curve showed both high recall and high precision. High scores for precision and recall show that the classifier is returning accurate results (high precision), as well as returning a majority of all positive results (high recall). PR shows the trade-off between the true positive rate and the positive predictive value for a predictive model using various probability thresholds. The performance of the proposed and existing state-of-the-art deep learning models is shown in Fig. 3.

The grad activation map (GAM) has been used to show the activation of the deep learning models. Fig. 4 shows the GAM based on the proposed Vis-transform and DensNet201 for five validation class samples. By visualization, the proposed Vis-transform produced a more accurate activation map as compared to the DensNet201 model in most cases as shown in Fig. 4 in the second, third, fourth, and fifth rows. This GAM proved that our proposed transformer-based model achieved better performance as compared to CNN's models.

The proposed majority voting ensemble-based Vis-Transformer model with TTA produced more accurate ROC curves as compared to CNN's based models as shown in Fig.3. Similar trends have been noticed for precision and recall curves. The confusion matrix based on the proposed majority voting ensemble-based Vis-Transformer is shown in Fig. 5.

4. Conclusions

In this paper, we have investigated the efficacy of state-of-the-art deep learning models on nematode recognition, a biological computing task. the vision-based transformer model using the post ensemble majority voting method has been applied to the nematode dataset for nematode classification. Further, the test time augmentation technique has been used to enhance the performance of the proposed model. The proposed transformer-based model achieved better performance as compared to CNN's deep learning models. We have conducted a benchmark of adopting various state-of-the-art Convolutional Neural Networks (CNNs) on this dataset. We also discussed and analyzed the results. However, there is still great space to improve, given the current highest overall accuracy of 82.62 %. It is possible to improve performance by designing sophisticated deep learning networks.

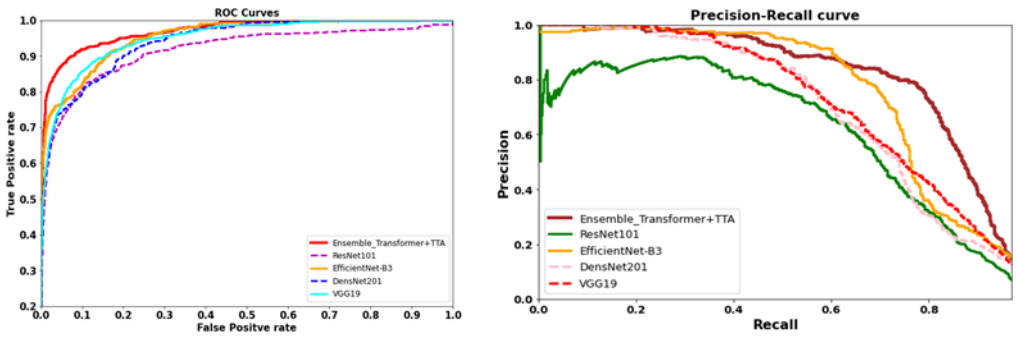


Fig. 3. The ROC and PR curves are based on proposed and existing state-of-the-art models.

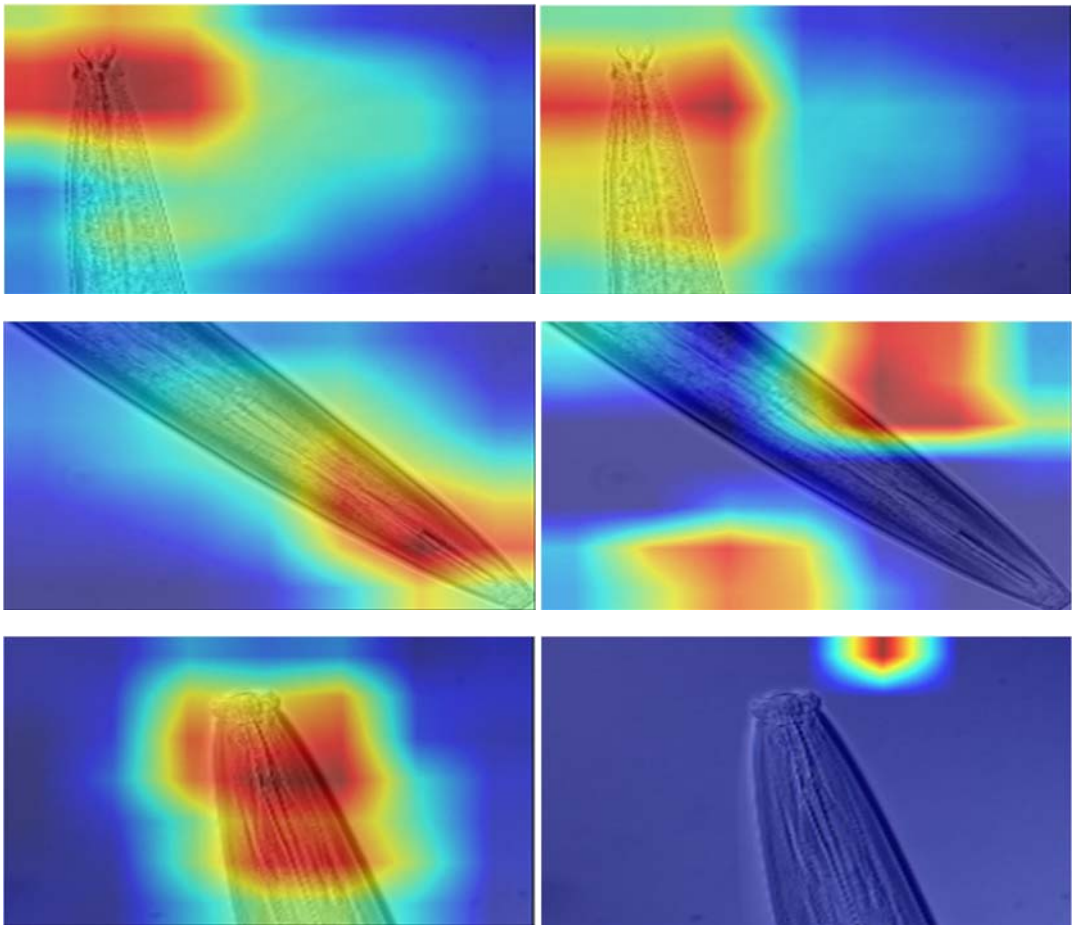


Fig. 4. Grad-activation map (GAM) based on the proposed Vis-transformer and DensNet201 models. The first column shows the GAM based on the Vis-transformer for five classes (Acrobeles, Axonchium, Dorylaimus, Pratylenchus, Rhbiditi). The second column shows the GAM based on the DensNet201 model for five same classes.

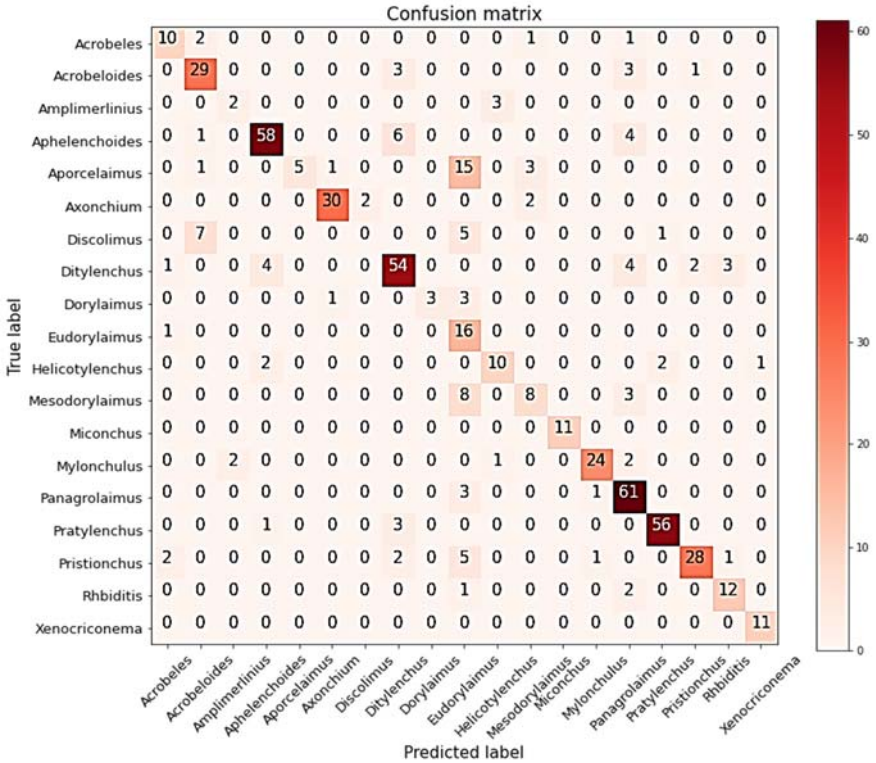


Fig. 5. The confusion matrix based on the proposed ensemble vision transformer with the TTA technique.

Acknowledgments

This work is supported by EUR ISblue (Interdisciplinary Graduate School for the Blue Planet) and Equinor.

References

[1]. J. M. Nicol, S. J. Turner, D. L. Coyne, L. den Nijs, S. Hockland, Z. T. Maafi, Current nematode threats to world agriculture, in Genomics and Molecular Genetics of Plant-Nematode Interactions, *Springer Netherlands*, Dordrecht, 2011, pp. 21-43.

[2]. V. Reinke, *et al.*, A global profile of germline gene expression in C. elegans, *Mol. Cell*, Vol. 6, Issue 3, Sep. 2000, pp. 605-616.

[3]. A. Coomans, Present status and future of nematode systematics, *Nematology*, Vol. 4, Issue 5, 2002, pp. 573-582.

[4]. S. Derycke, G. Fonseca, A. Vierstraete, J. Vanfleteren, M. Vincx, T. Moens, Disentangling taxonomy within the Rhabditis (Pellioditis) marina (Nematoda, Rhabditidae) species complex using molecular and morhological tools, *Zool. J. Linn. Soc.*, Vol. 152, Issue 1, Jan. 2008, pp. 1-15.

[5]. B. Weischer, Nematode parasites of vertebrates. Their development and transmission, *J. Phytopathol.*, Vol. 148, Issues 11-12, Dec. 2000, pp. 637-642.

[6]. M. L. Blaxter, The promise of a DNA taxonomy, *Philos. Trans. R. Soc. London. Ser. B Biol. Sci.*, Vol. 359, Issue 1444, Apr. 2004, pp. 669-679.

[7]. A. Javer, A. E. X. Brown, I. Kokkinos, J. Rittscher, Identification of C. elegans strains using a fully convolutional neural network on behavioural dynamics, in *Proceedings of the European Conference on Computer Vision (ECCV'18)*, 2018, pp. 455-464.

[8]. M. Liu, X. Wang, K. Liu, X. Liu, Multi-focal nematode image stack classification using a projection-based multi-linear method, *Mach. Vis. Appl.*, Vol. 29, Issue 1, Jan. 2018, pp. 135-144.

[9]. M. Liu, X. Wang, H. Zhang, Taxonomy of multi-focal nematode image stacks by a CNN based image fusion approach, *Comput. Methods Programs Biomed.*, Vol. 156, Mar. 2018, pp. 209-215.

[10]. B. H. Holladay, D. S. Willett, L. L. Stelinski, High throughput nematode counting with automated image processing, *BioControl*, Vol. 61, Issue 2, Apr. 2016, pp. 177-183.

[11]. M. Liu, A. K. Roy-Chowdhury, Multilinear feature extraction and classification of multi-focal images, with applications in nematode taxonomy, in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'10)*, Jun. 2010, pp. 2823-2830.

[12]. M. Liu, A. K. Roy-Chowdhury, M. Yoder, P. De Ley, Multi-focal nematode image classification using the 3D X-Ray Transform, in *Proceedings of the IEEE International Conference on Image Processing (ICIP'10)*, Sep. 2010, pp. 269-272.

[13]. M. Liu, X. Wang, H. Zhang, A multi-direction image fusion based approach for classification of multi-focal nematode image stacks, in *Proceedings of the IEEE International Conference on Image Processing (ICIP'17)*, Sep. 2017, pp. 3835-3839.

[14]. B.-T. Zhou, J.-H. Baek, An automatic nematode identification method based on locomotion patterns, in *Proceedings of the International Conference on Intelligent Computing (ICIC'06)*, 2006, pp. 372-380.

[15]. K. Simonyan, A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition, http://arxiv.org/abs/1409.1556

[16]. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'16)*, Jun. 2016, pp. 770-778.

[17]. G. Huang, Z. Liu, L. van der Maaten, K. Q. Weinberger, Densely Connected Convolutional Networks, http://arxiv.org/abs/1608.06993

[18]. M. Tan, Q. V. Le, EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks, http://arxiv.org/abs/1905.11946

[19]. A. Dosovitskiy, *et al.*, An Image is Worth 16×16 Words: Transformers for Image Recognition at Scale, http://arxiv.org/abs/2010.11929

[20]. X. Lu, Y. Wang, S. Fung, X. Qing, I-Nema: A Biological Image Dataset for Nematode Recognition, http://arxiv.org/licenses/nonexclusive-distrib/1.0/

(4288)

Visual Inspection of Patterned Surfaces with Supervised Deep Learning Using Both Real and Synthetic Defects

M. Haselmann, P. J. Krassnig and D. P. Gruber
Polymer Competence Center Leoben GmbH, Roseggerstraße 12, 8700 Leoben, Austria
Tel.: + 433842429627 +44/+749
E-mail: matthias.haselmann@pccl.at, paul.krassnig@pccl.at

Summary: Automated visual inspection plays a major role in industrial quality control. In this contribution we show a use case where defects can be only a few pixels in size and sometimes stand out only slightly from their patterned surroundings. Defect detection methods used in industry should have a high degree of flexibility and adjustability. In addition to training on only defect-free examples, it is advantageous if the method optionally benefits from collected real defects, which allows an improvement of the defect detection accuracy where pure one-class training is not sufficient. In this contribution we propose an approach that uses both, real and synthetic defects, leading to an improvement in the state-of-the-art of defect detection at specific patterned surfaces.

Keywords: Visual inspection, Quality control, Machine learning, Synthetic defects, Anomaly detection.

1. Introduction

Defect detection on patterned surfaces is an important factor in industrial quality control. Humans are good at recognising anomalous structures in patterns, but are poor at reliably scanning larger areas over longer periods of time. To ensure high and consistent quality, automated solutions is a must.

The challenge in the use case shown, is that defects can be only a few pixels in size and sometimes stand out only slightly from their patterned surroundings. State-of-the-art one-class methods such as PatchCore [1], which perform very well on public anomaly detection datasets such as the MVTec dataset [2], perform relatively poorly on the patterned surface images shown in the result section. For industrial use in a visual inspection system, the accuracy is far from sufficient. Furthermore, by their very nature, pure anomaly detection methods cannot benefit from a few collected defect examples. In the industrial environment, however, after initial training on fault-free surface images, defects that were not automatically detected gradually accumulate. The ability to use these defect instances in a subsequent finetuning step is very important for the typical industrial mode operandi. One possibility for pure one-class learning methods such as PatchCore [1] would be to lower the anomaly threshold so that a sufficiently large recall is achieved at the expense of a low precision. The retrieved instances could then be divided into true and false anomalies using a further classifier trained on collected real defects.

Another possibility, which is explored in more detail in this paper, is the use of synthetic defects as proposed in [3]. This approach has already demonstrated to be effective with similarly structured surfaces. Moreover, if real defects are available for training, they can easily be injected into the stream of the synthetic ones. Training in several steps with an increasing collection of real defects that were not recognised in earlier versions is thus straightforward.

This contribution compares PatchCore [1] as a representative of state-of-the-art one-class learning methods [4] with supervised learning on the patterned surface images shown. In the context of supervised learning, training on real defects is compared with training on purely synthetic defects and, for the first time, with training on synthetic and real defects.

2. Methods and Implementation Details

For the supervised approaches, extraction at random positions within the training images provides a virtually endless stream of fault-free patches. Hand-labelled real defects for training are extracted with an initial patch size of 512×512 px. The defects are near the centre and the patches are cropped to 256×256 px just before being used. This way, any image edge created by random affine image transformations during training is avoided. In supervised learning with real defects, the (augmented) defect patches are injected into the stream of fault-free patches with a probability of 50 %. This approach is very similar to the oversampling approach used for supervised training on (strongly) imbalanced datasets. For supervised training with synthetic defects, defects are generated according to [3] on 50 % on the fault-free patches. In the proposed mixed synthetic and real defect training approach, each patch of this 50/50 fault-free and synthetic stream is replaced by an augmented real defect patch with a probability of 1/32. For all supervised learning experiments Resnet18 was used as Deep Convolutional Neural Network (DCNN) classifier in a pretrained initial state (on ImageNet).

3. Experiments

3.1. Dataset Details

All experiments are conducted on patches of size 256×256. For the fault-free training data for PatchCore and the common validation and test data for all tested methods, patches are extracted using a sliding window with stride 160. This results in a number of overlapping patches of 3884 for the training data, of 2014 for the validation data, and 1498 for the test data. Regarding the real defects, 129 patches are available for training, 74 for validation and 95 for testing. All patches used for training, validation and testing originate from different surface samples, so that any correlation between training, validation and test data is prevented.

3.2. Results

In training without real defects (see Table 1, Fig. 1), the synthetic defect method [3] achieved a Matthew-Correlation-Coefficient (MCC) of 0.70 and thus clearly outperformed PatchCore (0.26). Supervised learning using oversampling resulted in a MCC of 0.58 and 0.84 for 32 and 129 real defects, respectively. This was clearly outperformed by the proposed training with mixed real and synthetic defects resulting in a MCC of 0.88 (32 real defects) and 0.92 (129 real defects).

Table 1. Preliminary results including number of real defects used for training, confusion matrix entries (True positives (TP), False Positives (FP), False Negatives (FN), True Negatives (TN) and the Matthews-correlation-coefficient (MCC) for the test data. The best MCC scores in relation to the used number of real defects during training is marked in bold.

Method	#Real def.	TP	FP	FN	TN	MCC
PatchCore [1]	0	28	63	67	1435	0.26
Superv. + O-.Samp.	32	63	51	32	1447	0.58
Superv. + O-.Samp.	129	76	9	19	1489	0.84
Synth. Def. [3]	0	58	12	37	1486	**0.70**
Proposed	32	83	9	12	1489	**0.88**
Proposed	129	90	10	5	1488	**0.92**

4. Conclusions

Although current state-of-the-art anomaly detection methods show top performance on datasets such as MVTec [2], they still can struggle with certain industrial datasets such as demonstrated in this contribution. In such a case, specialized methods such training using synthetized defects [3] still outperform them. In this paper, a training mode with real and synthetic defects is proposed, which shows significant improvement over supervised training with either purely synthetic defects or with purely real defects. With the proposed approach, an inspection model that was initially trained without real defects can be improved step by step over time by real errors, which is an important ability in industry.

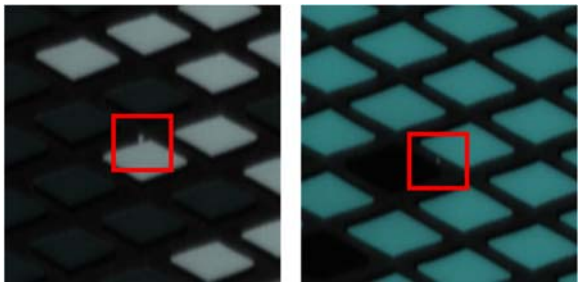


Fig. 1. Example of image patches with small defects.

Acknowledgements

The research work was performed within the COMET-project „Deep on-line learning for highly adaptable polymer surface inspection systems“ (project-no.: 879785) at the Polymer Competence Center Leoben GmbH (PCCL, Austria) within the framework of the COMET-program of the Federal Ministry for Climate Action, Environment, Energy, Mobility, Innovation and Technology and the Federal Ministry for Digital and Economic Affairs. The PCCL is funded by the Austrian Government and the State Governments of Styria, Lower Austria and Upper Austria.

References

[1]. K. Roth, et. al., Towards total recall in industrial anomaly detection, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR'22)*, 2022, pp. 14318-14328.

[2]. P. Bergmann, et. al., The MVTec anomaly detection dataset: A comprehensive realworld dataset for unsupervised anomaly detection, *International Journal of Computer Vision*, Vol. 129, Issue 4, 2021, pp. 1038-1059.

[3]. M. Haselmann, D. P. Gruber, Pixel-wise defect detection by CNNs without manually labeled training data, *Applied Artificial Intelligence*, Vol. 33, Issue 6, 2019, pp. 1-19.

[4]. Y. Chen, et. al., Surface defect detection methods for industrial products: A review, *Applied Sciences*, Vol. 11, Issue 16, 2021, 7657.

(5498)

Drone Radio Signal Detection with Multi-timescale Deep Neural Networks

C. Horn [1], M. Nyfeler [1], Georg Müller [1] and C. Schuepbach [2]
[1] Zurich University of Applied Sciences, ICLS, Schloss 1, 8820 Wädenswil, Switzerland
[2] Armasuisse, Feuerwerkerstrasse 39, 3602 Thun, Switzerland
E-mail: matthias.nyfeler@zhaw.ch

Summary: We develop a multi-timescale deep learning algorithm to detect drones from radio signals. While previous approaches focused on the analysis of high-frequency radio data alone we integrate signals from the higher timescale of the drone communication protocol in an end-to-end architecture. To this end, we develop a new meta-CNN layer, which generalizes the idea of the standard CNN (which slides a single, fully connected kernel along a higher-level input) towards arbitrarily complex kernel models. To detect higher timescale patterns our system uses an LSTM layer in the top layers. As a result, our model is able to extend drone identification abilities significantly toward very small SNRs.

Keywords: Drone signal detection, Deep learning, Multi-timescale modeling.

1. Introduction

As consumer drones are becoming more popular and omnipresent, the need to detect them to ensure privacy and security increases. A promising detection technology for consumer drones is the detection of the radio signals transmitted. This approach is characterised by the following properties: The detection range is in the order of magnitude of the flight range of the drone, the location accuracy is relatively good and the type of signal allows for conclusions to be drawn about the drone model.

Following previous work [1-3], we focus on deep learning-based approaches to separate drone signals from WIFI background and the distinction of different drone types. Our main contributions are the following:

- Introduction of a multi-timescale approach for drone signal detection;
- Implementation of a meta-CNN layer in PyTorch;
- Achievement of an increased sensitivity for drone detection systems.

2. Method

Typical deep learning architectures originally developed for image data continuously compress the input dimension by applying a sequence of convolution and pooling layers. At the same time, the liberated processing power can be used to extract a steadily increasing amount of feature representations.

In the case of radio signals from drones, however, information is present at two distinct timescales: The high-frequency microseconds scale of the radio signal itself and a seconds-level timescale determined by the drone communication protocol. An architecture that tries to learn signals at all intermediate timescales would be very parameter inefficient here. In contrast, our approach makes use of three key performance drivers: The statistical benefit of repeated observations, the detection of temporal patterns, and the power of end-to-end learning.

2.1. Lower Timescale Model

Previous studies found that deep convolutional neural networks with many layers achieve high performance on radio data. Our radio signal model is inspired by the VGG architecture, which was originally developed to process image data. We adapt it to 2-dimensional input data with the convolution applied only along the time dimension (The second dimension represents the real and imaginary parts of the complex numbered input signal). We use 5 blocks, where each block contains 2 convolutional layers, each followed by a batch-normalization and a final pooling layer. The number of filters is increased from 32 for the first block to 256 for the last block.

This type of radio signal network is able to identify the microsecond pattern of different radio signals, as other studies have found before [2]. The model produces a classification output roughly every millisecond, while human evaluation happens at the seconds timescale. This offers the opportunity that based on counting statistics alone we can achieve a much higher accuracy on the operator timescale compared to the model output timescale. To see this, imagine there are n lower-level outputs for each higher-level outputs and we require a threshold of k positive lower-level outputs for a positive classification at the higher timescale. Then the probability to observe a signal at the higher-level timescale, P, given an observation probability of p at the lower-level timescale is determined by the binomial distribution. We can thus apply it to calculate the number of true positive (tp) and false positive (fp) results expected at the higher timescale from the corresponding lower timescale values:

$$P_{tp} = 1 - \binom{n}{k} p_{tp}^{k} (1 - p_{tp})^{n-k},$$

$$P_{fp} = 1 - \binom{n}{k} p_{fp}^{k} (1 - p_{fp})^{n-k},$$

and, since furthermore $P_{fn} = 1 - P_{tp}$ and $P_{tn} = 1 - P_{fp}$ we can derive the higher-level accuracy as:

$$acc = \frac{P_{tp} + P_{tn}}{P_{tp} + P_{tn} + P_{fp} + P_{fn}} = \\ = \frac{1}{2}\left(\begin{matrix} 1 - \binom{n}{k} p_{tp}^{k} (1 - p_{tp})^{n-k} + \\ + \binom{n}{k} p_{fp}^{k} (1 - p_{fp})^{n-k} \end{matrix}\right)$$

As a result, even small lower-level accuracies can give rise to high accuracies at the higher level for sufficiently large *n* and appropriately chosen threshold values *k*. This significantly expands the region where our overall model can achieve near-optimal performance.

However, we do not just want to count the number of lower-level signals but also make use of their temporal patterns. This could be achieved by a meta learning approach, where the output of the radio signal network is used as a new training set for a separate higher-level model.

2.2. Meta-CNN

Our meta-CNN approach offers the additional advantage that we can train our model over multiple timescales end-to-end. For an overview of the full architecture of our approach see Fig. 1. A standard CNN takes a short fully connected layer and slides it over the input which enables it to identify the (lower level) patterns learned by it independently of their location on the higher-level input scale. Our meta-CNN layer generalizes this idea in that it allows to slide any arbitrarily complex kernel model along a higher-level input. In our case, we use our radio signal network as the kernel.

Importantly, the meta-CNN allows for choosing a stride of less than the kernel size which increases the resolution compared to a simple meta learning approach.

2.3. Higher Timescale Model

For the detection of temporal patterns in the drone communication protocol, we use a 1-d convolution and pooling layer to reduce the number of parameters. This is followed by a layer-normalization and an LSTM layer. Layers of LSTM cells are known for their ability to detect patterns in time series data [4]. After flattening of the feature vector, a fully connected layer is used for the final classification.

2.4. Data Preparation and Model Training

We recorded drone radio signal data from DJI Phantom pro 4 drones, while a person was actively manipulating the remote control, as well as Wi-Fi noise background. The signals are all at a carrier frequency of 2.44 GHz. They were with a USRP in an echo-free at a sampling rate of 56 MHz and then downsampled down to 14 MHz using an anti-aliasing filter, to investigate if detection is still possible at lower frequencies.

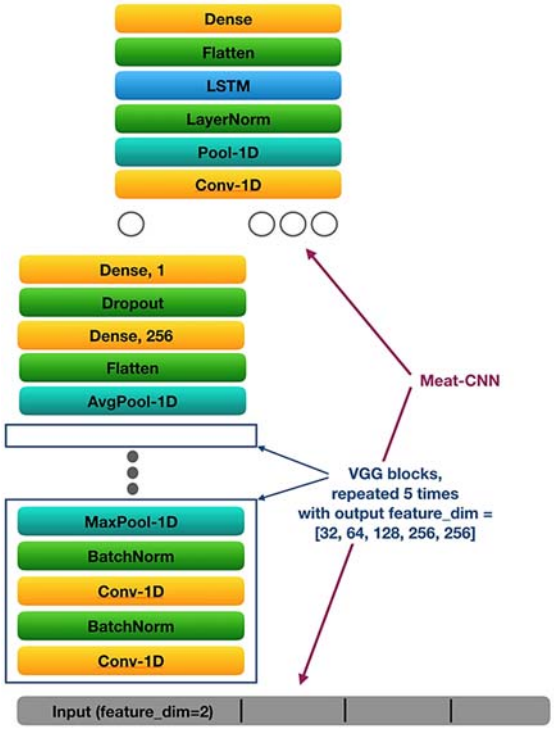


Fig. 1. Complete neural network architecture. The Meta-CNN shifts the input producing a single neuron output (denoted by ◯) for each step. The combined signal of all these neurons is then processed by the higher timescale (upper part) of the network.

The signal has gaps with low power between bursts as shown in Fig. 2 and perform frequency hopping, i.e., the bursts hop between different frequencies in the frequency band as can be seen in Fig. 3.

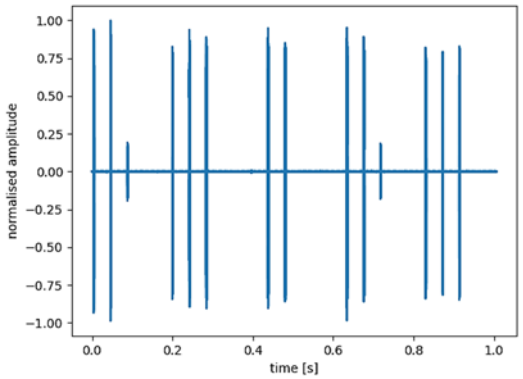


Fig. 2. The real part of a 1 second of a normalised I-Q-radio signals from a DJI phantom pro 4 system in the 2.44 GHz frequency band. The short signal bursts are clearly visible on which the lower timescale model was trained. The meta model then classifies the sequences of outputs from the lower timescale model.

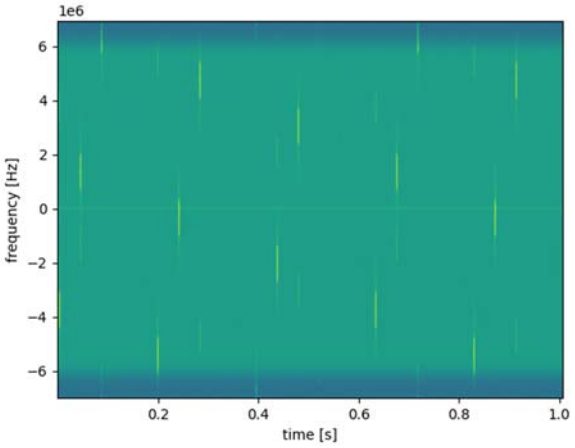


Fig. 3. Spectrogram of 1 seconds of drone signals from the DJI phantom pro 4 in the 2.44 GHz frequency band at a sampling rate of 14 MHz showing the short signal bursts and the frequency hopping

We remove input vectors with an average power of less than 0.001 of the maximums, to facilitate training. All input vectors were normalized individually. We construct noisy signal data by mixing the recorded signal with Gaussian noise and Wi-Fi noise with SNR between -40, and 0 dB. We use equal amounts of signal and background samples in the training set. Additionally, data augmentation is used by constructing multiple datasets from the same background data blended with individual Gaussian noise. We find that this approach increases the robustness of our results. For training our models we use a combination of datasets with SNRs of -30, -20, -10 and 0 dB. The training of the full model (lower-timescale model + meta-CNN + higher timescale model) happened end-to-end over all parameters.

3. Related Work

We compare our performance with previous results on the DeepSig dataset [5] and obtain very similar results as our radio signal network model Deep-six. A comparison on the higher timescale is not possible, since the vectors in the DeepSig dataset are not sequential signals.

4. Results

As a baseline for our higher timescale model, we consider the case where we apply the meta-CNN but simply use a fixed (optimized) threshold on the total count of positive outputs from the lower-level model.

Here we focus on the discrimination ability between drone signals and Wi-Fi backgrounds, leaving the differentiation between drone types for later.

The performance of the three model variants is shown in Table 1. An accuracy of 50 % corresponded to random guessing. While the radio signal model is useless at -30 dB, the meta-CNN baseline operates still at full performance. The dependence of the performance of our full model on SNR is shown in Fig. 4. We observe that the model is still effective even below -35 dB.

Table 1. Accuracy comparison for different models and SNRs. For small SNR values the model performance degrades to 50 %.

SNR [dB]	Lower timescale model only	Meta-CNN with fixed threshold	Full model
0	99 %	99 %	99 %
-10	99 %	99 %	99 %
-20	64 %	99 %	99 %
-30	50 %	98 %	99 %
-35	50 %	62 %	96 %
-40	50 %	50 %	53 %

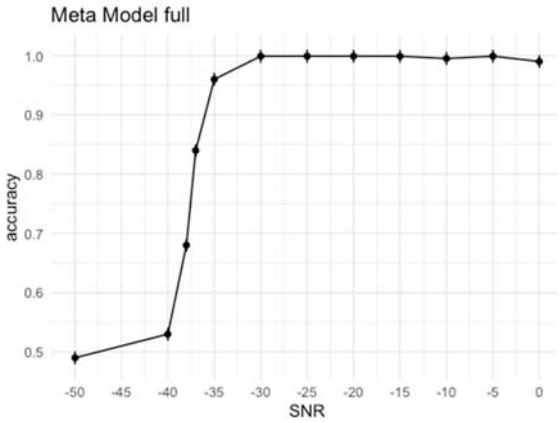


Fig. 4. Performance of the full model versus SNR.

4.1. Importance of Hyperparameters

We tested the influence of stride values for the meta-CNN and found that it had a relatively large effect on model performance. For our final model we used a stride of a quarter of the kernel model input size. Using strides equal to the kernel model input size give significantly worse results with the performance curve (Fig. 4) shifted 10 dB upward. But using even smaller than 1/4 kernel model input strides did not lead to large additional performance gains.

5. Conclusions

We developed a multi-timescale model for the detection of radio signals from drones. The novel neural network architecture we propose combines a 25-layer CNN-based kernel model for the identification of lower timescale signals with a LSTM-based model on the top to extract higher-level communication pattern. As our experiments show this enables a significant extension of drone identification

ability down to SNR values below -35 dB. We note that the idea of a meta-CNN with a complex kernel model might be useful in general for other tasks where relevant information is present at different timescales.

6. Outlook

Further drone signals from different brands (DJI, Futaba, Taranis, Graupner, Turnigy) have been recorded at carrier frequencies of 2.4 GHz, 5.8 GHz and telemetry link at 0.869 GHz and the performance of the model on the new data is under current investigation. Furthermore, a porotype for live-testing in the field is under construction and will be the topic of a further publication.

Acknowledgements

We thank Selina Malacarne and Nicola Ramagnano from the University of Applied Sciences Rapperswil for recording the data and setting up the implementation.

References

[1]. T. J. O'Shea, T. Roy, T. C. Clancy, Over-the-air deep learning based radio signal classification, *IEEE Journal of Selected Topics in Signal Processing*, Vol. 12, 2018, pp. 168-179.

[2]. M. Ezuma, F. Erden, C. K. Anjinappa, O. Ozdemir, I. Guvenc, Microwave detection and classification from RF fingerprints using machine learning techniques, in *Proceedings of the IEEE Aerospace Conference (AeroConf'19)*, 2019, pp. 1-13.

[3]. S. Scholl, Classification of Radio Signals and HF Transmission Modes with Deep Learning, https://arxiv.org/pdf/1906.04459.pdf

[4]. S. Hochreiter, J. Schmidhuber, Long Short-Term Memory, *Neural Computation*, Vol. 9, Issue 8, 1997, pp. 1735-1780.

[5]. RF Datasets for Machine Learning, https://www.deepsig.ai/datasets/

(6248)

Automatic Identification of Extremely Tiny Brain Hemorrhages in Susceptibility Weighted Images using Convolutional Neural Network

S. Obeidavi and G. Hirtz
Chemnitz University of Technology, Reichenhainer Straße 70, Chemnitz, Germany
E-mail: sahereh.obeidavi@etit.tu-chemnitz.de

Summary: Ischemic stroke is an acute cerebrovascular disease that causes long-term disability and even death. Acute lesions that occur in most stroke patients can be eliminated with careful diagnosis and treatment. Although MRI imaging is sensitive to these lesions, it is difficult and difficult for physicians to accurately estimate their location and volume manually, as examining hundreds of MRI-produced incisions takes a long time to examine and can also lead to human error. Be a lot. It is widely accepted by physicians that automated segmentation procedures for ischemic stroke lesions can significantly accelerate the onset of treatment. A considerable number of methods have been introduced for this purpose, but among them, the use of deep neural networks has resulted in better and more accurate results. Of course, these networks themselves have been introduced in a variety of frameworks and hardware, and here we are looking at a new framework for the automatic segmentation of stroke lesions in MRI images. The proposed structure consists of two different networks, the first is the U-Net convolutional neural network and the second is the Inception convolutional neural network, in which the U-Net network forms the main structure of the model, and the Inception network in each U-Net network layers are used. Among the deep learning algorithms, the U-Net algorithm, due to accurate response, high accuracy, high processing and learning speed, no need for large data sets to learn, in recent years the popular algorithm to identify image components and segment them in processing Medical images have become, and the use of the Inception network within the U-Net network has significantly increased segmentation accuracy. In this report, with the focus on the architectures, pre-processing, data pre-preparation and post-processing, we explain the structure of this algorithm and its convolutional network, as well as the most appropriate setting for the parameters and super parameters of this algorithm to optimize and achieve maximum accuracy in segmenting images related to stroke with this algorithm.
Keywords: machine learning, convolutional neural network, brain hemorrhages, stroke, classification.

1. Introduction

Very small accumulations of blood or Cerebral microbleeds (CMB) in the brain are more commonly found in patients with stroke, dementia, and cardiovascular disease [1]. Diagnosing CMBs can help predict stroke. In addition, clinical studies in recent years have shown that CMB can lead to a cognitive impairment such as dementia [2]. Magnetic resonance imaging (MRI) is used to detect CMBs because of its advantages over other imaging modalities. Among the various methods of magnetic resonance imaging, two methods, T2-star and 2SWI, are used. [3] showed that Susceptibility weighted imaging (SWI) increases the number of CMBs detected.

Image processing for automatic detection of CMBs faces challenges: a) different sizes of CMBs between 2 and 10 mm, b) different locations of CMBs [1], and c) there are many pseudo-CMBs that process Makes diagnosis difficult. In the current clinical practice, CMBs are manually labeled. [4] Manual labeling is difficult, time-consuming, and error-prone. Hence, many automated and semi-automated algorithms were created. Existing algorithms have a large number of positive false positives (FP) that reduce their value and continue research in this area. In the following, some examples of previous works are reviewed.

Reference [5], provides a semi-automated method for detecting CMBs. In this paper, first, a threshold algorithm is used to separate the CMB from the pseudo-CMBs and then the support vector classifier (SVM) (4), and finally, manual post-processing is performed. Reference [6] used a random forest classifier in reference [6] and reached a sensitivity of 92.04%. The extraction properties were then classified using the 5ISA network and support vector machine classifier with 89.44% sensitivity. In reference [8] a new structure of convolution neural network (CNN) with a 6RBAP layer was presented And with an accuracy of 97.18%, the reference [9] using the optimal convolution neural network achieved a sensitivity of 99.74%, in the reference [10] a method based on ResNet-50 was proposed and the sensitivity was reported to be 95.71%.

This paper tries to improve the evaluation parameters for CMB detection by various experiments on adjustable convolution neural network parameters. Finally, a network with three convolution layers, two pulling layers, and one fully connected layer is proposed.

The present paper is organized into four sections: In the second section, the proposed algorithm is presented. In the third part, experiments and evaluation of the proposed method are presented and in the fourth part, the work is summarized.

2. Proposed CNN Network Structure

In order to solve the problem of CMB detection based on SWI images, the use of a CNN network was proposed. The proposed algorithm includes a pre-processing step and a convolutional neural network. To improve the performance of the CNN network, experiments were performed to test the factors that can affect the performance of CNN, and based on the results, the parameters for evaluating the network structure were selected. The proposed CNN network structure is shown in Fig. 1.

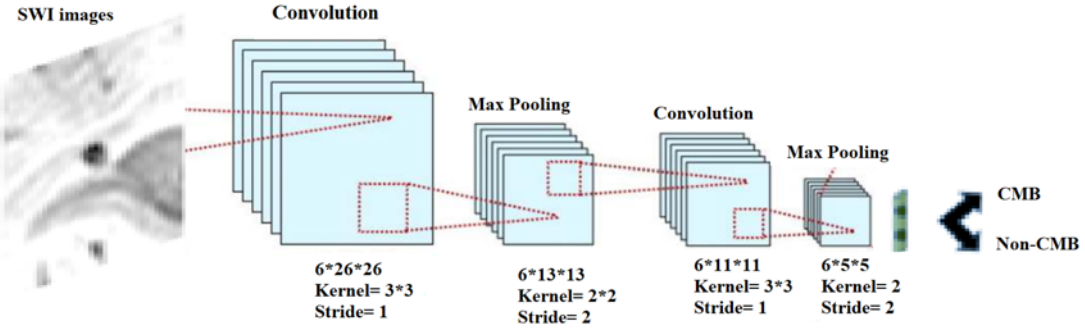


Fig. 1. The proposed CNN network.

Next, convolution and preprocessing neural networks were described as prerequisites.

- **Convolution layer**

The convolution layer is the main network layer that contains a set of learnable filters. Each filter is spatially small but continues along the depth of the inlet mass. Simply put, a filter is a three-dimensional mass that can be defined by any number of filters. But it should be noted that a large number of filters increases the computational load.

The activator function applied to the output of the convolution layer can be linear or nonlinear. Nonlinear functions are used to separate data that cannot be separated linearly. In this research, Rectified Linear Unit (ReLU), Leaky ReLU, and Parametric ReLU (PReLU) functions have been investigated as activation functions, which are defined by relations 1, 2, and 3, respectively.

$$y = max(0,x) \tag{1}$$

$$y = \begin{cases} 0.01x & x < 0 \\ x & x > 0 \end{cases} \tag{2}$$

$$y = \begin{cases} ax & x < 0 \\ x & x > 0 \end{cases} \tag{3}$$

- **Polishing layer**

The polishing layer is usually placed after a canonization layer, and by maximizing 8 or averaging 9, the number of pixels in the width and height is reduced, followed by a decrease in the number of parameters and the complexity of the calculations. The polishing operation is performed independently at each depth.

- **Fully connected layer**

As the name implies, all the neurons in this layer are connected to the neurons in the previous layer. The fully networked layer combines all the features to categorize the input image.

- **Preprocessing**

In the preprocessing step according to relation 4 which is shown below, normalization is used, so that the maximum pixel size is limited to one. The goal of normalization is to have the input pixels have a similar distribution to speed up convergence as the network is trained.

$$X = \frac{X0}{max(X0)} \tag{4}$$

The proposed CNN network consists of 6 layers. Table 1 shows the proposed convolution neural network layers.

Table 1. Proposed network layers.

No.	Layers	Filter size	Filter number	Output dimensions
1.	Pre-processed input patch	-	-	(61,61,1)
2.	Convolution + LeakyReLU	11×11	32	(61,61,32)
3.	Maximum pooling	-	-	(30,30,32)
4.	Convolution + LeakyReLU	3×3	32	(30,30,32)
5.	Maximum pooling	-	-	(15,15,32)
6.	Convolution + LeakyReLU	3×3	32	(15,15,32)

Experiments have shown that the Leaky-ReLU activator function and the maximal-based pooling layer lead to better results, hence they were placed in the proposed network. A fully connected layer was

placed at the end to classify the input image as CMB or non-CMB based on the training algorithm.

3. Tests and Evaluations

In this part of the research, the database for network training, evaluation criteria, and results are introduced. Implementations in the Python programming language are done on Google Colab.

For validation in this research, the 10-folder cross-validation method has been used. In this method, the database is randomly divided into 10 subsets. In each subset, some data is stored as validation data for model testing, and the rest of the data is used as training data. The results are then averaged to produce a single estimate.

3.1. Data Preparation

In this study, 20 volumes of SWI-CMB database images that are available to the public [11] have been used. Most images are 150 × 512 × 512 in size. By separating and saving the slides from the image size, 2982 images were created, which is 63 images from this CMB collection. 10SNP [12] with a 61×61×61 slider was used to create input patches and tags. The reason for choosing the 61×61×61 size for the patch is that this size preserves useful information for detection [9]. The central pixels of the patch, which contained CMB, was labeled one and non-CMB labeled zero. Fig. 2 and Fig. 3 show examples of patches labeled one and zero, respectively.

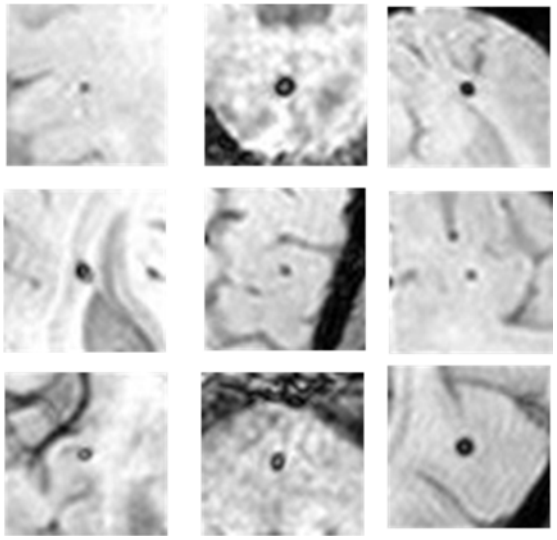


Fig. 2. Sample CMB patches.

Because there were several CMBs in some of the CMB images, the number of CMB patches was reduced to 74 patches after the separation of these CMBs, and to 1776 patches by adding plugin data from symmetry, rotation, and displacement.

From non-CMB images, 1500 non-CMB patches were generated. Finally, the total number of patches for training and evaluation of the network was 3276 patches, of which 90% were used for training and 10 % for evaluation. It is important to note that plugin data for test data has been removed from the tutorial.

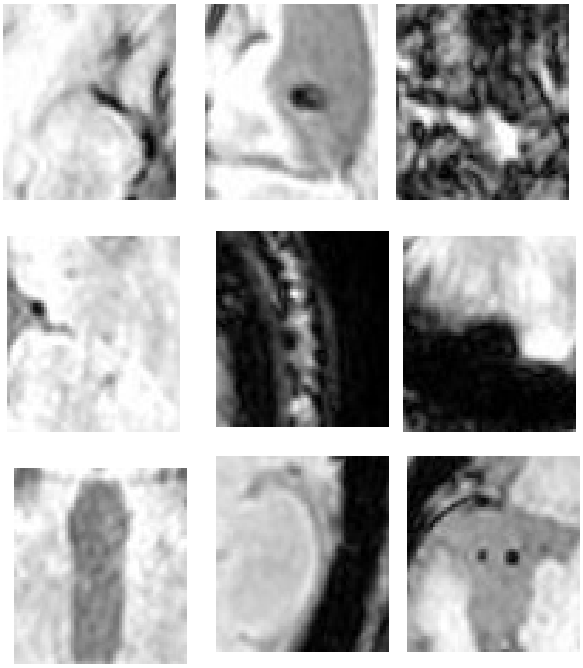


Fig. 3. Sample non-CMB patches.

3.2. Evaluation Parameters

To evaluate the classification results, three indicators were used: sensitivity, specificity and accuracy according to relations 5, 6, and 7. In this study, positive and negative show CMB and non-CMB, respectively.

$$sensitivity = \frac{TP}{(TP + FN)} \quad (5)$$

$$specificity = \frac{TN}{(TN + FP)} \quad (6)$$

$$accuracy = \frac{(TP + TN)}{(TP + TN + FN + FP)} \quad (7)$$

3.3. Experiments

Experiments were performed to achieve the desired structure and evaluate the proposed algorithm. The experiments used the Adam optimizer, which requires little memory and fast convergence. The cost function in this research is considered a reciprocal entropy function. CNN initial weights were randomly adjusted to 13 with a pack size of 128. The softmax function is also used in the fully connected layer.

- **Experiment I: Set the number of IPAC algorithms**

In this experiment, the algorithm was executed 10 times per number of different IPACs and the average evaluation criteria per 10 times performed are shown in Fig. 4. As can be seen from Fig. 4, the number of ipak equals 11 will be a good choice.

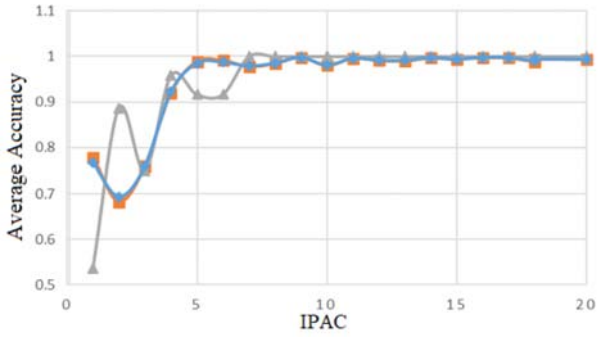


Fig. 4. Set the number of IPAC algorithms.

- **Experiment II: Adjust the number of layers**

Research shows that the number of layers can affect CNN performance, especially convolutional layers used to extract features. The average results obtained from 10 executions for different layer arrangements are given in Table 2. As can be seen from Table 2, the arrangement of the layers in the form of Conv / Pl / Conv / Pl / Conv, respectively, will be a good choice.

Table 2. The effect of layer layout on CNN network performance.

Layers	Average Accuracy (%)	Average Feature (%)	Average Sensitivity (%)
Conv	96/32	97/46	75
Conv/Pl	97/34	98/4	77/5
Conv/Pl/Conv	97/34	98/2	81/25
Conv/Pl/ Conv/Pl	98/10	98/1	95
Conv/Pl/Con v /Pl/Conv	98/3	98/2	96/25
Conv/Pl/Con v/ Pl/Conv/Pl	98/35	98/2	95

- **Experiment III: Activation and pulling function**

Different combinations of the activation and polishing function lead to different network functions, so it is necessary to test different combinations. The algorithm was run 10 times for each combination of the activation and pulling function and the mean values of the evaluation criteria are shown in Table 3. According to Table 3, the efficiency of the average-based polishing layer is less than the maximum-based polishing layer.

- **Experiment IV: Size and number of filters in the convolution layer**

In the first convolution layer, to achieve the best detection performance of CMBs, different numbers and sizes of the corresponding filter in this layer were examined, the algorithm was run 10 times per mode, and the mean values of the evaluation parameters in Tables 4 and 5 is brought. Thus, as shown in Tables 4 and 5, the filter size was set to 5 and the number of filters to 32.

Table 3. The effect of different combinations of activation and pulling functions on CNN network performance.

Activator Function	Pooling based on	Average accuracy (%)	Average feature (%)	Average sensitivity (%)
ReLU	Max.	98/67	98/8	96/25
PReLU	Max.	97/4	97/6	93/75
Leaky ReLU	Max.	99/24	99/26	98/75
ReLU	Average	97/9	98/2	92/5
PReLU	Average	96/01	96/53	86/25
Leaky ReLU	Average	97/84	97/8	98/75

Table 4. The effect of first convolution layer filter size on CNN network performance.

Filter size	Average accuracy (%)	Average feature (%)	Average sensitivity (%)
3	98/98	99	98/75
5	99/43	99/46	99
7	99/05	99	99
9	99/05	96/06	98/75
11	98/98	96/06	97/5

Table 5. The effect of the number of first convolution layer filters with dimensions of 5*5 on the efficiency of the CNN network.

Num. of Filters	Average accuracy (%)	Average feature (%)	Average sensitivity (%)
8	98/48	98/73	93/75
16	99/17	99/13	98/91
32	99/22	99/2	99/9
64	99/24	99/2	99/8
128	98/92	98/86	99/93

- **Experiment V: The proposed network performance**

In order to improve the classification performance of CMBs, experiments were performed to test almost all the factors that could affect CNN performance.

Finally, a convolutional neural network consisting of three convolutional layers, two pulling layers, and a fully connected layer achieved better performance. To avoid randomness, the evaluation results were performed 10 times. The mean and deviation from the mean are shown in Fig. 5.

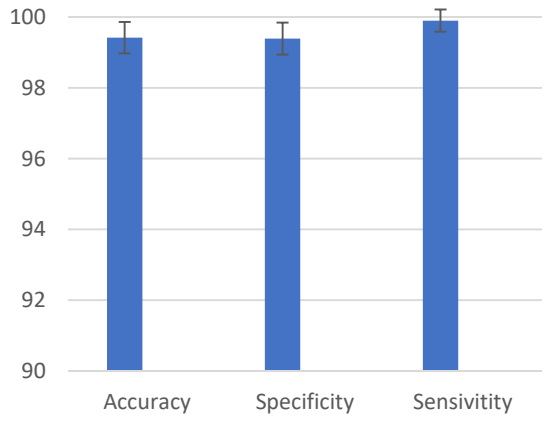


Fig. 5. Evaluation results in 10 times the implementation of the proposed network.

The efficiency of the proposed network with other methods is shown in Table 6. As can be seen, the proposed method has better results than other methods.

Table 6. Comparison of the proposed method with previous articles.

-	[8]	[9]	[10]	Proposed method
Year	2017	2018	2019	2022
Images format	SWI	SWI	SWI	SWI
Accuracy (%)	97/18	98/32	97/46	99/41
Feature (%)	97/18	96/89	99/21	99/39
Sensitivity (%)	96/94	99/74	95/71	99/9

4. Conclusion

The purpose of this article is to automatically detect cerebral hemorrhage. Because manual processing of this issue is difficult, time-consuming, and error-prone, it is essential to provide a way to automate the processing. The extraction of image features is the main and important step of this process, which in this paper uses the high ability of the convolutional neural network. In order to reduce the network parameters and also to maintain sufficient accuracy for classification, a 6-layer convolution neural network with small dimensions has been used. Experiments performed on the SWI image database show that the proposed method has better results than the reference methods.

References

[1]. Y. Wu, T. Chen, An up-to-date review on cerebral microbleeds, *Journal of Stroke and Cerebrovascular Diseases,* 25, 2016, pp. 1301-1306.

[2]. A. Charidimou, A. Krishnan, D. J. Werring, H. Rolf Jäger, Cerebral microbleeds: a guide to detection and clinical relevance in different disease settings. *Neuroradiology,* 55, 2013, 655-674.

[3]. S. Shams, J. Martola, L. Cavallin, T. Granberg, M. Shams, P. Aspelin, L. O. Wahlund, M. Kristoffersen-Wiberg, SWI or T2*: which MRI sequence to use in the detection of cerebral microbleeds? The Karolinska Imaging Dementia Study, *American Journal of Neuroradiology*, Vol. 36, No. 6, 2015, pp. 1089-109.

[4]. J. De Bresser, M. Brundel, M. M. Conijn, J. J. Van Dillen, M. I. Geerlings, M. A. Viergever, P. R. Luijten, G. J. Biessels, Visual cerebral microbleed detection on 7T MR imaging: reliability and effects of image processing, *American Journal of Neuroradiology,* 34, 2013, pp. E61-E64.

[5]. S. R. Barnes, E. M. Haacke, M. Ayaz, A. S. Boikov, W. Kirsch, D. Kido. Semiautomated detection of cerebral microbleeds in magnetic resonance images, *Magn Reson Imaging*, Vol. 29, No. 6, 2011, pp. 844-852.

[6]. A. Fazlollahi, F. Meriaudeau, V. L. Villemagne, C. C. Rowe, P. Yates, O. Salvado, P. Bourgeat, Efficient machine learning framework for computer-aided detection of cerebral microbleeds using the radon transform, in *Proceedings of the 11th IEEE International Symposium on Biomedical Imaging (ISBI)*, 2014, 49.

[7]. Q. Dou, H. Chen, L. Yu, L. Shi, D. Wang, V. C. Mok, P. A. Heng, Automatic cerebral microbleeds detection from MR images via independent subspace analysis based hierarchical features, in *Proceedings of the 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2015, pp. 7933-7936.

[8]. S. Wang, Y. Jiang, X. Hou, H. Cheng, S. Du, Cerebral micro-bleed detection based on the convolution neural network with rank based average pooling, *IEEE Access*, 51, 2017, pp. 16576-16583.

[9]. J. Hong, S. H. Wang, H. Cheng, J. Liu, Classification of cerebral microbleeds based on fully-optimized convolutional neural network, *Multimedia Tools and Applications*, 52, 2018, pp. 1-19.

[10]. J. Hong, H. Cheng, Y. D. Zhang, J. Liu. Detecting cerebral microbleeds with transfer learning, *Machine Vision and Applications*, 53, 2019, pp. 1-11.

[12]. Y. D. Zhang, X. X. Hou, Y. Chen, H. Chen, M. Yang, J. Yang, S. H. Wang, Voxelwise detection of cerebral microbleed in CADASIL patients by leaky rectified linear unit and early stopping, *Multimedia Tools and Applications,* Vol. 77, No. 17, 2018, pp. 21825-21845.

(6444)

A New Real-world Video Dataset for the Comparison of Defogging Algorithms

A. Duminil, J. P. Tarel and R. Brémond
Université Gustave Eiffel, Laboratoire COSYS-PICS-L, F-77454 Marne-la-Vallée, France
E-mail: alexandra.duminil@univ-eiffel.fr

Summary: Video restoration for noise removal, deblurring or super-resolution is attracting more and more attention in the fields of image processing and computer vision. Works on video restoration with data-driven approaches for fog removal are rare however, due to the lack of datasets containing videos in both clear and foggy conditions which are required for deep learning and benchmarking. A new dataset, called REVIDE, was recently proposed for just that purpose. In this paper, we implement the same approach by proposing a new REal-world VIdeo dataset for the comparison of Defogging Algorithms (VIREDA), with various fog densities and ground truths without fog. This small database can serve as a test base for defogging algorithms. A video defogging algorithm is also mentioned (still under development), with the key idea of using temporal redundancy to minimize artefacts and exposure variations between frames. Inspired by the success of Transformers architecture in deep learning for various applications, we select this kind of architecture in a neural network to show the relevance of the proposed dataset.

Keywords: Video dataset, Restoration, Defogging, Fog, Video processing.

1. Introduction

Visibility restoration of single hazy images is a well-known problem in computer vision and image processing. Visual artefacts in the images such as loss of contrast and color shift, contribute to reduced scene visibility which is detrimental to the performances of computer vision tasks such as segmentation, object/scene detection and recognition. Therefore, dehazing algorithms are often needed as a pre-processing step. Many efficient methods are available for single image dehazing, with or without learning. Several datasets are also available with foggy images and a clear image for reference. Due to the scarcity of foggy scenes, most of these databases are produced by adding a synthetic fog. And few works tackle video dehazing.

Working on fog removal in videos is more complicated due to lack of foggy video datasets with a clear video for reference. There are single image and video datasets with synthetic fog. For example, Ren et al. [1] generated a video dataset from the NYU-Depth V2 dataset [2], consisting of video sequences of indoor scenes. However, due to the domain gap between synthetic and real-world foggy dataset, the results on real world images are not convincing. Recently, Zhang et al. [3] proposed a useful hazy video dataset for learning and data evaluation, called REVIDE. Hazy scenes are produced with a fog machine which is more realistic than synthetic fog.

In order to enrich the state of the art, a new video dataset with and without fog is proposed. It includes video sequences of a reduced-scale but real-world scene with several fog densities and several lighting conditions. Earlier works addressed video visibility restoration as an extension of image restoration. However, applying image processing methods to videos introduces artefacts and flickering effects, due to the lack of temporal consistency. The challenge is to get a video restoration algorithm ensuring spatial and temporal coherence. For instance, Wang et al. [4] and Zhang et al. [3] proposed a video restoration architecture with a deformable Convolutional Neural Network (CNN) module, which achieve interesting results. To deal with the temporal information, recent works have been inspired by the Transformers architecture [5] which gives interesting results in video restoration for super resolution and image synthesis applications. While many methods use pure Transformers to solve their tasks, we propose to combine CNN and Transformer architectures to process the temporal information in videos. We have investigated including a Transformer module in a neural network's architecture within a video restoration method. The proposed algorithm, called TCVD, is demonstrated with the proposed VIREDA dataset.

2. Foggy Video Dataset

2.1. Creation Process

The main idea was to collect videos with and without fog as well as associated depth maps. This was achieved by creating a fog chamber to reproduce a hazy scene with the help of a fog machine. This fully controllable system allows to roughly control the fog density and illumination conditions. A series of static objects were included in the chamber, together with a small-scale car which was remotely operated to introduce temporal variations in the otherwise static scene. In order to make sure that the position of the car was exactly the same for all visibility and lighting

conditions, a stop-motion technique was implemented. The following acquisition process was applied for each robot position: First, a video of the scene without fog was recorded while altering the different lightings. Next, fog was fed into the chamber. Then, after a few seconds of waiting for the fog to settle, the video recording was started at the same time as the venting system was started. As fog was slowly evacuate, series of frames with decreasing densities of fog were recorded with each lighting. At each position of the robot, videos with different densities of fog, including without fog, have been acquired with a Kinect v2. This device captures both images and depth maps of the scene thanks to a Time-of-Flight technology, which provides additional information that are useful for some fog removal algorithms. At the end of the process, the frames corresponding to each illumination condition and fog density were splitted, and recomposed into new videos where the car was moving in stable visibility and lighting conditions.

2.2. Dataset Content

The use of LED strips made it possible to produce different illumination conditions. To simulate heterogeneous lighting (such as at night), lamps with different color temperatures were used. In the other five lighting conditions, white LEDs were used to simulate daylight. Fog density was evaluated by monitoring the contrast in a black & white panel. Three fog densities are available in the dataset (see Fig. 1), with contrasts of 0.015, 0.05 and 0.15 (the lower the contrast, the higher the density).

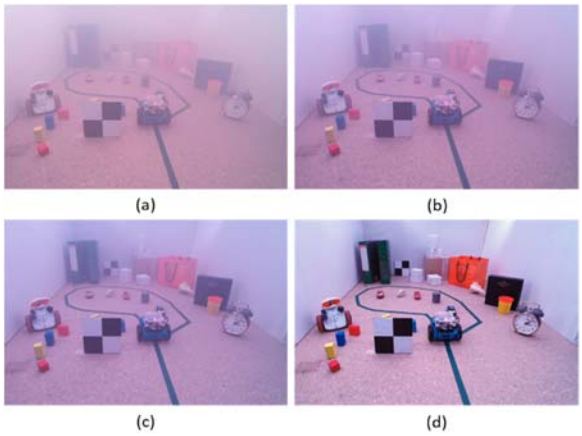


Fig. 1. Frames with three different densities: a) foggy frame with c = 0.015, b) foggy frame with c = 0.05, c) foggy frame with c = 0.15, d) ground truth.

The dataset contains 18 foggy videos and 6 ground truth videos, each containing an average of 180 frames (i.e., about 7 seconds). Additionally, there are depth maps associated with each position of the robot. The dataset is available on: https://github.com/alex-dml/VIREDA-video-dehazing.

2.3. Depth Maps

The use of the Kinect made it possible to collect depth maps via a Time of Flight (ToF) sensor. This sensor measures the time the light takes to travel a distance between the signal emission and its return to the sensor. As it is not sensitive to change in brightness, one depth map per robot position was enough. Depth maps are an additional source of information in the database. Fig. 2 shows the depth map for a robot position and the hazy image.

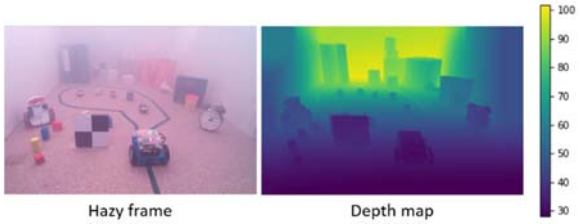


Fig. 2. Hazy frame and the associated false-colour depth map. Graduations are in centimeters.

3. Video Defogging Method

The popularity of the transformer architecture prompts exploration of its use for different types of tasks. Few works are yet inspired by this model for fog removal applications in videos. In this section, we present a multi-step method, called TCVD (Transformer-CNN architecture for Video Defogging), to process spatio-temporal information in video sequences. At each stage, spatial and temporal information are processed through several blocks and are then merged. This method combines two different architectures: CNNs to process spatial information and Transformers to ensure temporal consistency.

The overall architecture is based on an auto-encoder inspired by U-net. Using the transformer architecture is suitable to deal with sequences of data due to its ability to bind data within a sequence and to parallelize tasks. However, the Transformer has some drawbacks including a low inductive bias. The inductive bias represents the set of assumptions made by the model that allows it to learn and generalize beyond the learning data. On the other hand, CNNs have a strong inductive bias, allowing the model to achieve interesting performances with less training data. However, CNNs do not allow capturing the interdependence of images within a video sequence. The joint use of these two architectures can be an advantage.

3.1. Encoder

The encoder part of our architecture consists of four CNN modules and three modules based on a Transformer architecture, called TpFormer. Fig. 3 represents the overall diagram of the encoder.

Each step corresponds to the feature extraction of each image of the sequence. The features are then merged and form the input of the TpFormer blocks at several dimensions (which correspond to 2D convolution layer filters: 32, 64, 128, and 256). At the end of the feature extraction steps, the successive image triplets in the video are concatenated along the time axis.

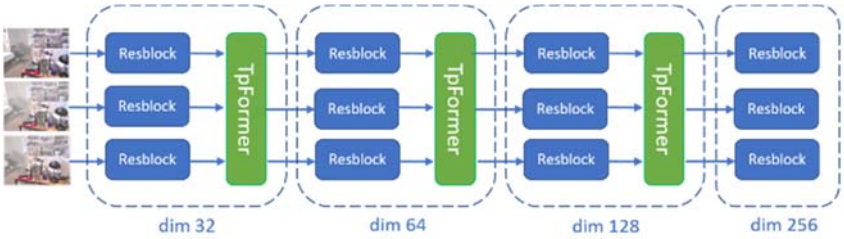


Fig. 3. Encoder architecture of the proposed method with three successive frames as input.

3.2. The TpFormer module

The TpFormer module, detailed in Fig. 4, is composed of several layers including a Multi-Head-Attention (MHA) layer [5]. This layer makes it possible to carry out different attention calculations in parallel. Using attention mechanism allows the network to learn the sequential dependency between frames. The features of the resulting frames are then dissociated and concatenated to the resulting spatial features of the CNN block of the same step. The merged information is the input to the next step.

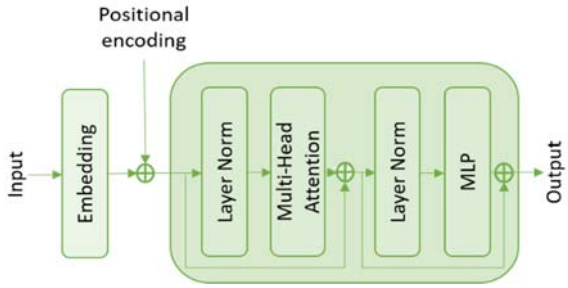


Fig. 4. TpFormer module, inspired by ViT [5].

3.3. Decoder

Once the encoder has extracted all the spatio-temporal characteristics of the images of the sequence, the decoder part will, symmetrically to the encoder, over-sample the encoded data of the central image, so as to obtain an image of dimensions equal to that of the input images. To upsample, Conv2DTranspose layers (Keras) were used, as well as skip connections from the encoder to retain fine details in the images.

4. Implementation

The implementation of the algorithm was carried out with the Keras/Tensorflow frameworks. We use the ADAM optimizer and a learning rate of 0.0001.

The Loss function is a combination between the L1 and SSIM distances. It is defined as follows:

$$L = a * Lssim + b * L1, \quad (1)$$

where a and b are balanced coefficients. We have resized foggy inputs and ground truths to a size of 224×224 and augmented them with horizontal flip and random 90° rotations.

5. Evaluation

Very few methods of video defogging provide publicly available code. Accordingly, our video defogging algorithm, still under development, only served to show some preliminary results on our dataset. TCVD is trained on the REVIDE dataset and TCVD2 is trained on both the REVIDE dataset and the synthetic images dataset (inspired by [1]) in order to have various types of data and be able to generalize. Then, we tested that with our video dataset, VIREDA.

5.1. Quantitative Evaluation

The proposed algorithm is compared to different state-of-the-art algorithms, including two prior-based methods (DCP [7] and MFP [8]) and a learning-based methods (FFA-Net [9]). Due to lacks of available codes for video dehazing, only image dehazing methods are used for comparison. Therefore, only the visual quality of the restored frames is assessed in this evaluation. The FFA-Net algorithm is retrained on the REVIDE dataset. Table 1 shows sample of the quantitative results with the SSIM [6] and PSNR metrics in the case of three fog densities including the different lighting conditions. All frames are resized to 224 x 224. The results illustrate that MFP, FFA-Net, TCVD and TCVD2 methods are competitive. The results obtained with the DCP method can be explained with the visual color shift results seen in the Fig. 5.

Table 1. Quantitative evaluation on different fog densities with the SSIM (top) and PSNR (bottom) metrics.

Fog density	DCP	MFP	FFA	TCVD	TCVD2
0.015	0.57 9.89	**0.74** **15.92**	0.68 14.14	**0.72** 13.29	0.70 **15.95**
0.05	0.60 10.30	**0.80** **16.09**	0.71 14.35	**0.76** 13.97	0.74 **15.73**
0.15	0.66 11.35	**0.86** **16.56**	0.78 **16.22**	0.79 14.68	**0.82** **17.16**

MFP method provides better results than those of our TCVD method with two fog densities: 0.05 et 0.15. Despite this, Fig. 5 shows that MFP is not as efficient as TCVD to removes fog. However, TCVD introduces black artefacts which penalize the results. TCVD2 shows better results with PSNR metric than MFP. Training TCVD2 on both databases appears to reduce artefacts introduced when training with REVIDE only.

5.2. Qualitative Evaluation

Fig. 5 shows sample images of the dataset restored with different dehazing methods. Each row in the figure corresponds to a different illumination condition with a fog density of 0,05. The DCP method seems to deteriorate the visual quality of the images, as shown by the different color shifts. The MFP and FFA-Net methods allow to slowly attenuate dense fog but produce more satisfactory results with lighter fog. The TCVD method, intended for videos, attenuates fog well but leaves black artefacts. This effect may be caused by the high intensity point lights implemented to simulate daylight: the algorithm may interpret these areas as fog to remove. As we noticed in the previous section, TCVD2 produces better quality results than TCVD. The artificial lighting used to create the video dataset produced slightly tainted images despite the use of white LEDs. It produced a colored fog which seems to disturb the restoration process by introducing some artefacts and color shifts.

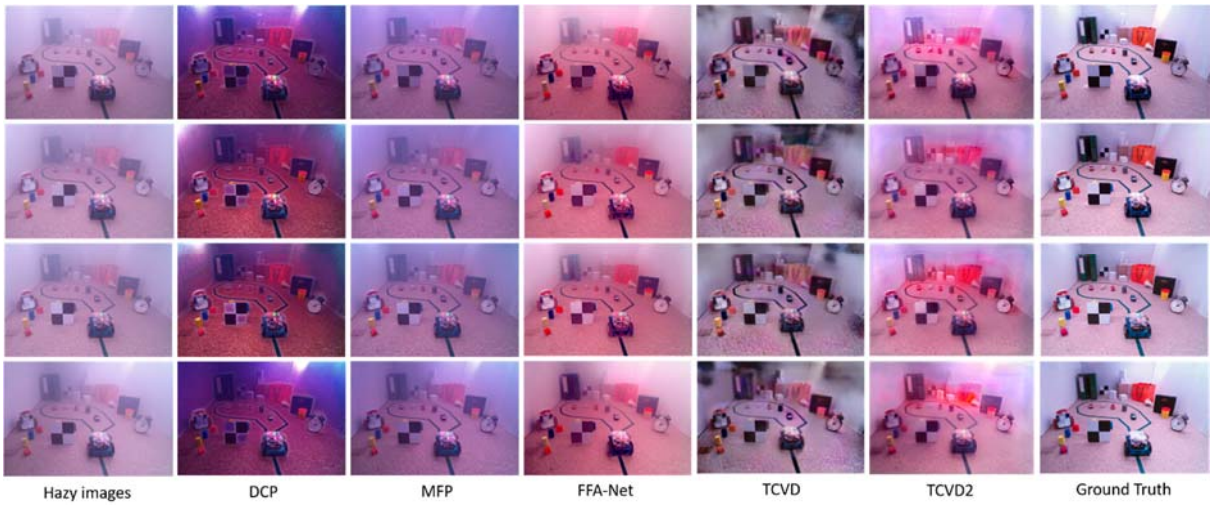


Fig. 5. Qualitative evaluation of dehazing algorithms on the VIREDA dataset with 0.05 as fog density value.

6. Conclusion

A new publicly available test dataset has been generated for the testing of video defogging. Although it is only composed of a single scene, it is available with several lighting conditions and several fog densities, and provides depth maps which may help assessing the quality of some defogging algorithms. For deep learning algorithms, it is useful to propose new sets with different conditions to expect a good generalization. We have implemented a multi-step hybrid method, called TCVD, which combines two architectures: CNNs and Transformers. This method helps reducing fog in realistic videos. However, the small number of publicly available code on defogging methods for videos does not allow to realize a complete benchmark.

References

[1]. W. Ren, J. Zhang, X. Xu, L. Ma, X. Cao, G. Meng, W. Liu, Deep video dehazing with semantic segmentation. *IEEE Transactions on Image Processing*, Vol. 28, Issue 4, 2018, pp. 1895-1908.

[2]. N. Silberman, D. Hoiem, P. Kohli, R. Fergus, Indoor segmentation and support inference from RGBD images, in *Proceedings of the European Conference on Computer Vision (ECCV'12)*, 2012, pp. 746-760.

[3]. X. Zhang, et al., Learning to restore hazy video: A new real-world dataset and a new method, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR'21)*, 2021, pp. 9235-9244.

[4]. X. Wang, et al., EDVR: Video restoration with enhanced deformable convolutional networks, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPR'19)*, 2019.

[5]. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need, *arXiv Preprint*, 2017, arXiv: 1706.03762.

[6]. Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, Vol. 13, Issue 4, 2004, pp. 600-612.

[7]. K. He, J. Sun, X. Tang, Single image haze removal using dark channel prior, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 33, Issue 12, 2011, pp. 2341-2353.

[8]. A. Duminil, J.-P. Tarel, R. Brémond, Single image atmospheric veil removal using new priors, in *Proceedings of the IEEE International Conference on Image Processing (ICIP'21)*, 2021, pp. 1719-1723.

[9]. X. Qin, Z. Wang, Y. Bai, X. Xie, H. Jia, FFA-net: Feature fusion attention network for single image dehazing, *arXiv Preprint*, 2019, arXiv:1911.07559.

(6667)

A Low-cost AI-based Soft Sensor for Food Drying Monitoring

E. Neofotistou [1], N. Stoupas [1] and M. Maniadakis [1,2]
[1] MadeEasy SCE, I. Venezi 36, N. Alikarnassos, Herakleion, Crete, Greece
[2] Foundation for Research and Technology Hellas,
N. Plastira 100, Vasilika Vouton, Herakleion, Crete, Greece
E-mail: mmaniada@ics.forth.gr

Summary: The current work explores the use of AI and low-cost sensor technology for the monitoring and optimization of food drying processes. More specifically, we have considered the use of plug-and-play sensors capturing humidity measurements to monitor and continuously assess the progress of food processing. The proposed developments have focused on (i) the automated termination of the food drying process as a means of saving time and valuable energy resources (ii) the prediction of food drying evolution by combining mathematical modeling and machine learning (iii) the comparative assessment of different drying setups, to support the viability of production. Besides the implementation and assessment of the individual modules, the developed technology has been integrated and tested as a whole on the drying of legume puree and apple slices with very promising results.

Keywords: Machine learning, Soft sensor, Food processing application.

1. Introduction

Food drying is a method of food preservation in which food is dehydrated to inhibit the growth of bacteria, yeasts, and mold through the removal of water [1]. However, the majority of existing small and medium size drying chambers operate in a static and rather blind mode with predefined setting that do not actually take into account the evolving characteristics of the ongoing dehydration process. The main reason for that is that it is difficult to frequently measure food moisture and adjust settings, because drying is disturbed every time a new food sample is taken.

The present work presents a new AI-powered approach for the indirect and real-time monitoring of food moisture that offers significant insight in the drying process and can provide the basis for improving food processing in terms of production costs control, resource efficiency and product quality.

First, we consider that is impractical to directly measure the moisture of food for making decisions on drying issues. To address that, we adopt a soft-sensor approach to contrasts the humidity in the broader atmosphere with the humidity of the air coming out of the drying chamber. We have collected data summarizing the effect of the drying chamber operation on the measured properties of the air at the drying chamber outlet. By using well established Machine Learning (ML) technologies we have been able to consider temperature difference between the broader atmosphere and the drying chamber. In that way we accomplish to indirectly obtain insight into the progress of drying, detect when the evaporated moisture is very low and stop drying when food dehydration reaches the desired result.

In addition, we develop a computational approach to model the evolution of the drying process, as a means to allow making predictions about the condition of the food after a certain time. This is achieved by using machine learning methods to optimize the fitting of mathematical models on collected data to optimize the accuracy of model predictions.

Finally, the two models are combined into a unified approach that enables the production manager examine how different drying parameters affect the efficiency of the dehydration process, thus enabling him/her to make decisions that enhance the viability of production.

2. Experimental Setup

To develop automation around the food drying process, it is essential to know when the food has been dehydrated enough, i.e., food moisture is below a preset upper moisture level [3, 4]. This is not easy to conclude as we will have to take frequent samples and thus stop the drying process. At the same time, measuring the absolute humidity of a food assumes a manual and time-consuming process that is not convenient to repeat often.

The current work aims at the modernization and digitalization of the food drying industry. This is accomplished by stimulating the integration of IoT and related technologies in food production, via the implementation of a soft and smart sensor module that undertakes the monitoring of food drying [2].

To rapidly implement a functional system that is easily deployed in the real world for experimentation, we considered the use of the Raspberry Pi, a credit card-sized, low-cost but fully functional and programmable computer with modern high-definition multimedia capabilities. The Raspberry Pi can be directly interfaced with a bunch of sensors and actuators with the help of the General-Purpose Input/Output (GPIO) pins, the component that truly

enables "physical computing" and the sensory observation of multiple environment states.

Given the particularly low cost of gas sensors is about 10 euros, we considered the use of several possible alternatives that are listed below:

- *SEN0219*: Gravity Analog Infrared CO2 Sensor;
- *Adafruit CCS811*: Air Quality Sensor for VOC, eCO2, Aromatic Hydrocarbons;
- *BME 680*: VOC sensor, temperature sensor, humidity sensor and barometer;
- *MQ5*: gas sensor for H2, CH4, CO, Alcohol;
- *MQ9*: combustible gas sensor.

All sensors have been are sensitive to the changing characteristics of the air coming out of the drying chamber. Thus, we compared them to identify the one providing the most informative measurements using a custom-made Granger causality-like criterion. The PIMORONI BME 680 has been the most informative regarding the prediction of drying parameters. This is a sensor that provides relative rather than absolute humidity measurements.

3. Drying Completion Specification

The PIMORONI BME 680 sensor provides relative rather than absolute humidity measurements which are not directly applicable in the monitoring of food drying since the measured humidity values depend on temperature. To overcome this issue, it is necessary to develop an AI-based mechanism that enables the comparison of relative humidity measurements at different temperatures.

More specifically, in order to be able to decide when the drying of the food is completed, an approximate method has been developed, which is based on measuring the relative humidity of the air leaving the drying chamber and comparing it with the relative humidity of the atmosphere outside the drying chamber, also taking into account the drying chamber mode of operation and the temperature in it. To this end we use two identical PIMORONI BME 680 sensors to collect pairs of humidity and temperature measurements both at the inlet and the outlet of the drying chamber.

This is accomplished by installing one of the sensors in the chimney of the drying chamber, to collect measurements related to the exhaust air. The second sensor (identical to the first) is installed in the inlet of the chamber to provide measurements on the quality of the air coming into the chamber.

We used the above-described setup to collect data representing how the operation of the drying chamber at different temperatures, but without the presence of any food that could emit moisture. Accordingly, since there is no food in the chamber, the absolute humidity of the air at the inlet and outlet of the drying chamber will remain the same but at different temperatures. This will make the two identical sensors measure different relative humidifies on the same quality air.

The data we collected regard the temperature (*t1*) and the humidity (*h1*) at the inlet of the chamber and the temperature (*t2*) and the humidity (*h2*) at the outlet of the chamber. We collected data by examining the operation of the drying chamber at temperatures of 40, 45, 50, 55, 60, 65, 70 degrees every two hours from 07.00 to 23.00, for several nonconsecutive days.

This data has been used to build an LSTM neural network [5] which takes as input the pair humidity *h1* at temperature *t1* and the targeted temperature *t2,* to predict the expected humidity *h2* of the same quality air at temperature *t2*. Accordingly, we have been able to estimate how the operation of the drying chamber shapes the temperature and humidity measurements at the outlet of the chamber if no additional moisture is evaporated due to food drying, with an error that is less than 1 %. This null-food estimate is used as a baseline that critically facilitates the monitoring of the drying process. We call this, the "baseline" humidity at the outlet of the chamber.

In order to be able to decide in an approximate but effective way, when the drying of the products has reached the desired level, we rely on the following two observations:

- When there is fresh food in the drying chamber which is heated and therefore evaporates moisture, then the measurement of the relative humidity sensor in the chimney of the drying chamber will be higher than the "baseline".
- When the food in the chamber has dried and does not have enough extra moisture to excrete, then the measurement of the relative humidity sensor in the chimney of the drying chamber will be very close to the "baseline".

Following the above, by examining the difference between the measured relative humidity and the baseline relative humidity we can have a very good view of the stage in which the drying process is at the given time. In particular, when the difference between the two is below a desired limit, then we may conclude that the food dehydration has completed and we can decide to end the drying process. It is therefore necessary to determine the threshold indicating drying completion. This depends on the type of food being dried, the desired level of food dehydration and the operating temperature of the drying chamber.

Two different types of food were examined, namely, pulse puree and apple slices. For pulse puree we assume that the maximum allowed absolute moisture in the food should be less than 9 %. Similarly, for the case of apple slices the maximum allowed absolute moisture should be less than 14 %. The above are used to specify the threshold limits when estimating the difference between the measured humidity and the predicted baseline humidity for these two foods. The thresholds specified for the different operating temperatures of the drying chamber are shown in the table below.

By using the PMB Food Moisture Analyzer we have been able to accurately measure the moisture in the final, dried product, verifying that in all cases it has been below the desired limit.

Table 1. Humidity thresholds to determine the completion of food drying.

	Relative Humidity Thresholds	
Chamber Temperature	**Legume Puree**	**Sliced Apple**
40	3.2 %	4.5 %
50	2.7 %	3.9 %
60	2 %	3.1 %
70	1 %	2.2 %

4. Predict the Evolution of the Drying Process

In addition to the termination criterion of a drying process, it is particularly useful for the production manager to have a systematic method to predict the basic characteristics of a drying process, such as the duration and the energy consumption. This would allow the comparison of dehydration procedures with different initialization conditions, and thus enable the production manager to take decisions regarding the setup of efficient dehydration processes [4].

To address this issue, a new data collection was carried out which concerned the recording of sensory information in the chimney of the drying chamber throughout several food dehydration processes. Specifically, 14 drying procedures were performed for both the legume puree and the apple slices (two for each of the operating temperatures of the chamber at 40, 45, 50,...., 70 degrees Celsius). The recording of the instantaneous power consumption and the instantaneous humidity and air temperature in the chamber chimney every second is performed using the Raspberry PI.

Interestingly, sensory measurements have a smooth, relatively parabolic form which offers the possibility to be modeled easily and accurately with the use of a polynomial model. Without limiting the degree of polynomial that can be used for modeling, in the present study 4th degree polynomials were used. These are fully described by 5 parameters that represent the coefficients of the polynomial. A machine learning procedure was used to estimate the value of the 5 parameters.

We have used a simple, two-layer stacked LSTM to find the parameters of the polynomial. The LSTM receives as input the type of product to be dried (currently two types are considered, either legume puree or apple slices), its quantity, the temperature and relative humidity of the environment, the temperature at which the drying chamber operates, the average per minute temperature and the relative humidity in the chimney of the dryer over the first 20 minutes of drying. At the end of the LSTM, we add one more layer that implements the 4th degree polynomial and facilitates the direct comparison of the polynomial values with the recorded data. Using the polynomial model, it is possible to predict the values of the humidity sensors even after more than 10 hours of operation of the dryer. Table 2 summarizes the accuracy of the prediction model, showing the mean absolute error (MAE) and the mean average percentage error (MAPE) of the humidity values predicted by using the 4th degree polynomial.

Table 2. Polynomial Model Fitting.

	Humidity Prediction Error	
Product Type	**MAE**	**MAPE**
Legume Puree	0.87	0.021
Sliced Apple	0.83	0.019

5. Specification of Drying Parameters

By combining the models described in Sections 3 and 4, it is possible to compare different configuration setups for the operation of the dryer and find the most suitable for the drying task to be performed at a given time.

To assess the combined performance of the two models, four new pulse puree and sliced apple drying experiments were performed, to estimate the difference between the predicted and actual drying duration as well as the predicted and actual energy consumption. In order to predict drying duration, we use the LSTM model implemented in Section 4 to estimate the parameters of the polynomial representing the evolving dehydration procedure. Then, we assume the current environmental conditions to remain static for the next few hours, and we use the current environmental humidity to estimate the baseline humidity at the output of the chamber.

Following the procedure described in Section 3 we can estimate the moment that the distance between the humidity predicted by the polynomial mentioned above and the baseline humidity will be below the corresponding threshold. The percentage error of the estimated completion time is shown in Table 3, specified as:

$$Error\% = | Predicted - Actual | / | Actual |$$

Additionally, Table 4 presents the predicted energy needs assuming constant consumption on a minute-by-minute basis.

As expected, the predictions regarding the completion of the drying procedures show some variability in comparison to the actual food drying experiments. In the cases that the drying chamber operates in relatively low temperatures the whole procedure is prolonged and thus the prediction error on the completion time increases. In contrast when the drying chamber operates in high temperatures the duration of drying decreases together with the completion time prediction error.

As expected, the energy needs predicted by our model are overall less accurate mainly due to the simplified hypothesis of constant energy consumption every minute. However, it is interesting to note that the accuracy of energy consumption predictions is mostly

at the same level for all the conducted experiments, of a given product type. In particular, if we consider that the estimated energy consumption is directly affected by the time that the drying chamber operates, it seems that instant energy estimates are more stable when the drying chamber operates at lower temperatures. This is explained by the fact that in low temperatures, drying is less affected by environmental conditions. Thus, energy consumption can be predicted more accurately, although the slower change in the physical properties of the food and especially in food moisture, is also affecting the accuracy of the prediction regarding the completion time of drying. Conversely, when drying takes place at a relatively high temperature inside the drying chamber, then the requirement to maintain this temperature conflicts with the state and temperature of the outside environment resulting in more variability in the energy consumed to maintain a constant temperature inside the chamber.

Table 3. Drying Duration Prediction accuracy.

	Prediction Error	
Chamber Temperature	**Legume Puree**	**Sliced Apple**
40	11.3 %	12.1 %
50	10.4 %	11.4 %
60	9.2 %	10.8 %
70	8.2 %	10.1 %

Table 4. Drying Energy consumption accuracy.

	Prediction Error	
Chamber Temperature	**Legume Puree**	**Sliced Apple**
40	14.1 %	16.8 %
50	13.8 %	16.5 %
60	13.8 %	16.4 %
70	13.5 %	16.1 %

Overall, the forecasts for the drying time and the estimated energy consumption have a satisfactory accuracy, thus allowing the production manager to get an overview of the complete drying process and its main characteristics. By examining and comparing the different configuration setups, the production manager can make the right decisions accomplishing to design a drying process that effectively balances between the available time and the desired energy consumption, to achieve the desired quality on the final food product.

4. Conclusions

Our work presents a new AI-powered soft-sensor approach for the indirect monitoring of food moisture in food drying applications and the implementation of computational models that simulate the evolution of the dehydration procedure over time. The combined use of the two models provides a systematic way to examine what-if scenarios regarding the setup of the dehydration procedure and enables the production manager tune and increase the viability of production.

Acknowledgements

This research has been co-financed by the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH—CREATE—INNOVATE (project code: T2EDK- 03242, PRIME).

References

[1]. Z. Berk, Food Process Engineering and Technology, 3rd Ed., *Elsevier Inc*., 2018.

[2]. Y. Jiang, S. Yin, J. Dong, O. Kaynak, A review on soft sensors for monitoring, control, and optimization of industrial processes, *IEEE Sensors Journal*, Vol. 21, Issue 11, 01 June 2021, pp. 12868-12881.

[3]. D., Lauri Pla, R., Kamyar, N. Hashemian, H. Mehdizadeh, M. Moshgbar, Moisture soft sensor for batch fluid bed dryers: A practical approach, *Powder Technology*, Vol. 326, 2018, pp. 69-77.

[4]. S. Jensen, L. da Cruz Meleiro, E. F. Zanoelo, Soft-sensor model design for control of a virtual conveyor-belt dryer of mate leaves (Ilex paraguariensis), Biosystems Engineering, Vol. 108, Issue 1, 2011, pp. 75-85.

[5]. F. Curreri, L. Patanè, M. G. Xibilia, RNN- and LSTM-based soft sensors transferability for an industrial process, *Sensors*, Vol. 21, Issue 3, 2021, 823.

(7010)

Auto-encoder Based Medicare Fraud Detection

Mansour Zoubeirou A Mayaki and Michel Riveill
University Côte d'Azur, CNRS, INRIA, 2004 Rte des Lucioles, 06902 Valbonne, France
E-mail: mansour.zoubeirou-a-mayaki@inria.fr

Summary: In this study, we used deep learning based multiple inputs classifier with a Long-short Term Memory (LSTM) autoencoder component to detect medicare fraud. The proposed model is made of two separate blocks: MLP block and auto encoder feature extraction block. The MLP block extracts high level feature from the invoice data and the auto encoder block extracts high level features from that describes the provider behavior over time. This architecture makes it possible to take into account many sources of data without mixing them. The latent features extracted from the LSTM autoencoder have a strong discriminating power and separate the providers into homogeneous clusters. We use the data sets from the Centers for Medicaid and medicare Services (CMS) of the US federal government. Our results show that baseline artificial neural network give good performances compared to classical machine learning models but they are outperformed by our model.

Keywords: Medicare fraud, Anomaly detection, Deep learning, Auto encoder, Machine learning.

1. Introduction

Insurance fraud results in considerable losses for governments and insurance companies and results in higher premiums from clients. In the European Union, fraud would cost 13 billion per year to European customers and insurance companies [1]. In France for example, over 260 million of medicare fraud transactions are detected each year, mainly due to medicare providers and institutions. In the United States, medicare fraud represents 5-10 % of medicare claims and costs insurance companies between 21 billion and 71 billion per year.

Traditionally, business rule-based systems are used for fraud detection. These methods, although effective, are often very difficult to set up and maintain. Indeed, a rule-based fraud detection system constantly requires the presence of experts in the field and constant updates of the rules. Models based on machine learning make it possible to automatically build patterns and thus detect fraudulent behavior effectively. The main challenge when using the methods in fraud detection is that there not enough data labeled as fraudulent, this leads to an imbalance situation. In the field of medicare, fraudulent transactions represent less than 5 % of all transactions. This high imbalance ratio makes it very difficult for machine learning algorithms to learn as they will tend to favor the majority class.

To detect medicare fraud, we propose a multiple inputs deep neural networks model with a Long-short Term Memory (LSTM) autoencoder component. We call this architecture AE-MFD for Auto Encoder based medicare fraud detection method. This architecture makes it possible to take into account many sources of data without mixing them and makes the classification task easier for the final model. The LSTM autoencoder component plays a dimension reduction role for the provider data and its latent vector describes the provider behavior over time.

We use the publicly available medicare data sets from the Centers for medicare and Medicaid Services (CMS) of the United States federal government for period 2017-2019 [2]. The CMS data sets contain the hospitalization requests, the outpatient care requests etc.

The rest of the paper is outline as follows. The **Related Works** section discusses the other studies related to imbalance data handling, deep learning for anomaly and medicare fraud detection. In the third section, we describe our methodology, the model's architecture and the choice of hyperparameters. Section **Experimental Data Sets** we describe the experimental data sets and pre-processing steps. The results are presented and discussed in the last section.

2. Related Works

To deal with class imbalance issue, there are two main approaches with varying performance depending on the field of application and the complexity of the problem: the resampling approach (or data level) which consists in balancing the classes by adding or removing data from classes and the approach which consists in modifying the learning algorithms so that they take into account the class imbalance (algorithm level).

The Centers for medicare and Medicaid Services (CMS) data has been used in numerous studies to detect medicare fraud. Most of these studies use resampling techniques to overcome the imbalance class issue (Bauder et al. [3] Liu et al. [4], Herland et al. [5]; Johnson et al. [6], Van et al. [7]). In [5], the authors combined also the three parts of CMS data set and showed that it leads to better performance. They used logistic regression, random forest and gradient boosting classifiers to detect fraudster providers in the CMS. Their results show that the performance of all classifiers improves significantly when they used all

parts of the CMS data. In [6], Johnson et al. the CMS data over the period 2012-2016. They used neural network models with algorithm level and data level techniques to predict medicare fraud medicare. They tested random undersampling (RUS), random oversampling (ROS), mean square error (MSE) and Focal Loss techniques among others. Their results suggest that multi layers perceptrons (MLP) classifiers combined with ROS outperforms all other models. They also noted that RUS can improve model's performance up to certain level of imbalance ratio. Thus, RUS improves the performance if the majority class is above 99 %. Lin et al. [10] rewrite the classical entropy loss function by integrating two new parameters: α and γ (gamma). They called the new loss function Focal Loss. The focal lost consist of multiplying the classical cross entropy (CE) by a modulation factor $\alpha(1 - p)\gamma$. Hyper parameter $\gamma \geq 0$ adjusts the rate at which easy examples are down weighted and $\alpha \geq 0$ is a class-wise weight used to give more importance to the minority class [11]. Wang et al. [9] proposed to use a cost matrix with an artificial neural network-based model to predict readmission of patients to a hospital. The cost matrix is defined such that the cost of misclassified readmission is greater than that of misclassified non-readmission. During back propagation, the model penalizes more or pay more attention to the readmission class which is the class of interest. Jason et al. [7], compared different resampling techniques using 11 types of classifiers and 35 different data sets. The imbalance ratio of the experimental data sets varies between 1.33 % and 35 %. The resampling techniques used in this article are: random undersampling (RUS), random oversampling (ROS), one-sided selection (OSS), cluster-based oversampling (CBOS), Wilson's editing (WE), SMOTE (SM) and borderline-SMOTE (BSM). Their results show that RUS tends to give better performance (32 % of the time).

Most of data level experiments studies come to the conclusion that undersampling gives belter results than over-sampling. This goes against what one might have expected as undersampling often leads to a loss of information. One possible explanation is that in cases, adding new artificial data brings more noise than useful information to the model. Algorithm level methods often give better results than resampling methods as they don't alter the training data and don't lead to a loss of information. However, in some cases, when labeled data is limited, oversampling techniques are good way to extend the data set. Moreover, when the majority class distribution is stationary (the samples are very close to each other) undersampling may work very well as we don't lose lot of information by deleting some samples.

3. Methodology

In this section, we present the AE-MFD model architecture and the other classifiers we tested. We refer to the baseline neural network as MLP}.

3.1. Baseline Classifiers

We compared AE-MFD to baseline Multi-Layer Perceptrons (MLP) networks and state-of-the-art classifiers such as logistic regression (LR), random forest (RF), gradient boosting (GB) and XGBoost. These classifiers are good baseline models for classification tasks. They take an invoice as input and predicts if it's fraud or not. The Multi-Layer Perceptrons (MLP) model consists of a single input layer, multiple hidden layers, and an output layer. This model takes an invoice as input and predicts if it's fraud or not. The number of layers and the number of neurons in each layer of the MLP model are variables (hyper-parameters) that must be chosen carefully for neural network models to give good results. These variables remain constant throughout the training process and have a direct impact on the performance of the models. The choice of hyper-parameters is described in Subsection 3.4.

3.2. AE-MFD Model's Architecture

AE-MFD is made up of two different inputs layers. The MLP part input layer receives the claims details and the auto encoder component input layer receives the data relating to the healthcare provider. The model is thus composed of two blocks which meet at the end. Each block consists of an input layer, hidden layers and an output layer. The outputs of the two blocks are then concatenated to form a single vector. Such an architecture makes it possible to simultaneously take into account the details of claims and the healthcare provider behavior separately.

In our version of the multi-input model, the second block is a Long-short Term Memory (LSTM) autoencoder. We first trained the LSTM autoencoder on the provider level data. This autoencoder learns to reconstitute a healthcare provider behavior over time. Then we used the latent vector from the LSTM autoencoder as an input vector for our final model. The final model is thus composed of an input layer which takes as input the claims details and another input layer which makes it possible to inject the latent vectors coming from the autoencoder. In this architecture, the autoencoder plays a dimension reduction role for the provider data and its latent vector describes the healthcare provider behavior. Note that the autoencoder parameters remain constant when learning the final model.

3.3. Performance Metrics

The classifiers are evaluated using the precision-recall curve (PRC). This plot shows precision values for corresponding recall values. It provides a model-wide evaluation like the receiver operating characteristic (ROC) plot or the cost curve (CC). The area under the curve (AUC) score of precision-recall curve, denoted as AUC (PRC), is

likewise effective in multiple-classifier comparisons [11]. The AUC (PRC) measures the entire two-dimensional area under the entire precision-recall curve (by integral calculations) from (0,0) to (1, 1). We don't use AUC (ROC) as most studies do because as Saito et al. [11] show in their study, the AUC (ROC) is not well suited in case of imbalance class. They proved that AUC (ROC) could be misleading when applied in imbalanced classification scenarios instead AUC (PRC) should be used. Their study showed via multiple simulations that AUC (ROC) fails to capture the variation in class distribution contrary to the AUC (PRC).

As the AUC (ROC) is used as performance metric in most studies in the literature, we will give its value for each of our classifiers just as an indication. In order to have more detail on classifiers performance, we also compute the precision and the G-means score. The precision gives the performance of the classifier on the positive class and the G-mean metric makes a compromise between the true positive rate TPR or recall and the true negative rate (TNR).

3.4. Hyperparameters Optimization

We used the mini-batch stochastic gradient descent (SGD) with a batch size of 200. We used the version called SGD Adam which allows to adapt the step of the gradient during the training of the model. This optimizer is known for its better performance compared to other versions of SGD. We kept the default values of the other hyper parameters: $lr = 0.001$, $\beta_1 = 0.9$ and $\beta_1 = 0.999$. The rectified linear unit (ReLU) activation function is used in neurons of the hidden layers, and the sigmoid activation function is used for output layer to estimate posterior probabilities. We choose the best hyper parameters using the KerasTuner library and on a small holdout set (20 %) of the validation dataset. KerasTuner is an easy-to-use, scalable hyperparameter optimization framework that solves the pain points of hyperparameter search [12]. We also added between the hidden layers a Batch Normalization layer followed by a dropout layer. Batch Normalization makes artificial neural network faster and more stable by normalizing and rescaling layers inputs. Dropout consists in "deactivating" randomly some neurons during training [13]. Each neuron being possibly inactive during a learning iteration, this forces each unit to "learn well" independently of the others and thus avoid overfitting.

4. Experimental Data Sets

In order to evaluate our method's capabilities, we compared its performance to those of four state-of-the-art classifiers on four other publicly available bench mark data sets. The data sets are highly imbalanced and present some big data properties. Note that we put more emphasis on the CMS data as our primary focus is medicare fraud detection.

The Centers for medicare and Medicaid Services (CMS) publishes each year a series of publicly available data containing information on the use and payments of medical procedures, services and prescription drugs provided to beneficiaries as well as data on physicians and other actors in the healthcare system. These CMS data combined with the Office of Inspector General's list of excluded individuals and entities (LEIE) [14] containing the list of healthcare providers excluded from the healthcare system for illegal activity allows researchers in the field of statistics and artificial intelligence to propose new methods to fight against fraud in medicare.

The CMS data sets contains mainly two types of information: hospitalization requests (Inpatient Data) and outpatient care requests (Outpatient Data). The Inpatient Data contains information on patients admitted to hospitals. The Outpatient Data gathers information on patients who have visited the hospital without being hospitalized there. We thus have information such as the unique identifier of the healthcare provider (NPI), the refunded amount (AmtReimbursed) etc. Records within the data set also contain various provider-level attributes, e.g., National Provider Identifier (NPI), first and last name, gender, credentials, address etc.

Note that for a fraudster provider, we do not know which of its claims are fraudulent and which are legitimate. To overcome this label issue, we considered that if a provider is fraudster, all its claims are also fraudulent [6]. This assumption makes sense because if a provider has been declared as a fraudster, the decision certainly comes from a deep analysis of his recent activity and his claims reflect illegal activities. We created additional features for the providers by aggregating the variables at the invoice level. For each provider we created new variables by taking the mean, the variance, the sum, the skewness coefficient of the numerical variables per trimester. In order to capture the provider behavior over time, we use a slicing technique called bucketing [9]. The behavior of each provider with respect to each variable can be considered as a time series. Indeed, over the years the provider makes several claims for different patients. For example, we can calculate the total amount paid by the insurance company to the provider each month. The bucketing technique consists of separating the claims from each provider into groups according to time and aggregation indicators are calculated in each group. There are two possible levels of aggregation: first order (first order features) and second order (second order features). The first order indicators are mean, standard deviation, variance, sum, maximum, minimum, skewness and kurtosis. The second order indicators are: Energy (E1), Entropy (E2), Correlation ($\rho_{x,y}$), Inertia and Local Homogeneity (LH). In this study, we only calculate the first order indicators for each numeric variable per trimester. The data of each provider over each year is separated into 4 blocks. In each block, the aggregation

variables are calculated: mean, standard deviation, skewness, maximum, minimum, sum etc. After cleaning and preprocessing, the final data set has a fraud rate of 0.5 %. Table 1 shows a subset of the provider level data and Table 2 a subset of the LEIE data.

Table 1. Subset of the aggregated data set on provider level.

NPI	BeneID count	Deductible AmtPaid mean	InscClaimAmt Reimbursed sum	Fraud
1000051001	24	213.60	104640	No
1000051101	117	502.16	605670	Yes
1000051005	138	2.08	52170	No
1000051007	58	45.33	33710	No
1000051009	36	53.86	35630	No

5. Results and Discussions

The classifiers performances are listed in Table 3. Recall that we refer to the baseline neural network as MLP and our Auto encoder model as AE-MFD. MLP weighted stands for MLP with weighted loss, MLP focal with focal loss, MLP mfe with the mean false error loss and MLP rus the best MLP obtained by random under sampling. Despite the class imbalance in the training data, AE-MFD outperform all other classifiers in terms of AUC (PRC). Our model's AUC (PRC) is 0.765 and that of the second-best classifier is 0.741. Note that the no skill} (random) classifier has an AUC (PCR) of 0.03. Using the ROC (AUC) as performance metric, the baseline neural network with mean false error function (MLP mfe) has the best performance (0.864) but it has a very low precision (0.445) compared to our model (0.77). This is due to the fact that state-of-the-art classifiers (logistic regression, random forest, Gradient boosting and XGBoost) they fail to capture complex structures in sequence datasets and large-scale data [15]. As context matters in fraud detection, the advantage of AE-MFD is that the autoencoder separates the providers into homogeneous groups and creates contextual features. The main disadvantage of our method is that it requires lot of historical data to train the LSTM auto encoder. Thus, the final model's performance depends on the auto encoder performance.

5. Conclusion

Frauds or anomalies are very rare events but results in considerable losses for governments, insurance companies and taxpayers. In this study, we proposed a deep neural network with auto encoder to detect medicare fraud. We also tested some classical classifiers (random forest, logistic regression, gradient boosting) and simple MLP models. We use the publicly available medicare data sets from the Centers for medicare and Medicaid Services (CMS) of the United States federal government. The results of our experiments show that this kind of architecture outperforms a classical machine learning models and multi-layer perceptron models using a single input layer. The Long-short Term Memory (LSTM) autoencoder component learns high level contextual features from the input data. In addition, the capability of the LSTM auto encoder to extract strong discriminating latent features makes the model robust toward the imbalance class issue. Future work will include employing the multiple inputs models with data sampling techniques or algorithm level techniques to combat the imbalanced nature of the data.

Table 2. Subset of LEIE data set.

NPI	CITY	STATE	EXCLTYPE	EXCLDATE
1306111111	GARDEN CITY	NY	1128a1	20180220
1306111111	WARREN	OH	1128b5	20190220
1306111111	PHILADELPHIA	PA	1128b7	20191231
1306111111	FLUSHING	NY	1128a1	20190220
1306111111	SPRINGFIELD	MO	1128b4	20200220

Table 3. Experimental results of the proposed method and some state-of-the-art methods. Mean time refers to the execution time expressed in minutes.

Models	Precision	AUC(ROC)	Gmean	AUC (PRC)	Mean Time
Random	-	0	0.5	0.03	-
LR	0.438	0.827	0.826	0.629	0.51
RF	0.599	0.807	0.796	0.658	2.35
GB	0.713	0.715	0.666	0.617	15.43
XGBoost	0.707	0.713	0.663	0.613	0.53
MLP	0.436	0.863	0.862	0.739	2.54
MLP weighted	0.429	0.860	0.859	0.733	2.96
MLP focal	0.432	0.861	0.861	0.737	2.50
MLP mfe	0.445	0.864	0.863	0.741	6.03
MLP RUS	0.534	0.720	0.668	0.512	1.20
AE-MFD (Ours)	**0.770**	**0.794**	0.762	**0.765**	10

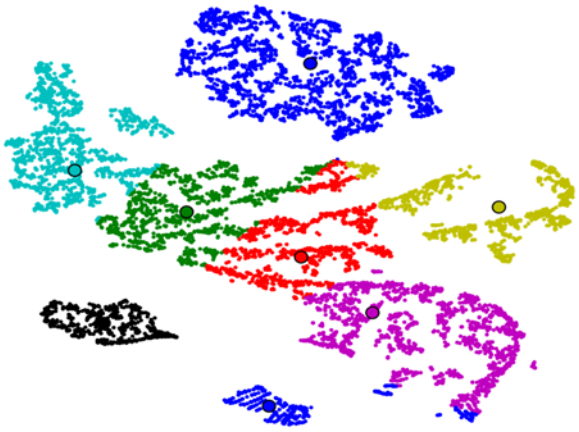


Fig. 1. Mean-shift clustering on the autoencoder latent vector. This output of the autoencoder separates the providers into homogeneous groups.

References

[1]. Fraud Prevention, https://www.insuranceeurope.eu/priorities/23/fraud-prevention

[2]. US Government, Centers for Medicare, Medicaid Services. Medicare Fee-for-service Provider Utilization, Payment Data Physician and Other Supplier Public Use File: A Methodological Overview, https://www.cms.gov/research-statistics-data-and-systems/statistics-trends-and-reports/medicare-provider-charge-data/physician-and-other-supplier

[3]. R. A. Bauder, T. M. Khoshgoftaar, The detection of medicare fraud using machine learning methods with excluded provider labels, in *Proceedings of the IEEE International Conference on Machine Learning and Applications (ICMLA'18)*, 2018, pp. 858-865.

[4]. Q. Liu, M. Vasarhelyi, Healthcare fraud detection: A survey and a clustering model incorporating geo-location information, in *Proceedings of the 29th World Continuous Auditing and Reporting Symposium (WCARS'13)*, Brisbane, Australia, 2013.

[5]. M. Herland, T. M. Khoshgoftaar, R. A. Bauder, Big data fraud detection using multiple medicare data sources, *Journal of Big Data*, Vol. 5, 2018, 29.

[6]. J. M. Johnson, T. M. Khoshgoftaar, Medicare fraud detection using neural networks, *Journal of Big Data*, Vol. 6, 2019, 63.

[7]. J. Van Hulse, T. M. Khoshgoftaar, A. Napolitano, Experimental perspectives on learning from imbalanced data, in *Proceedings of the 24th International Conference on Machine learning (ICML'07)*, 2007, pp. 935-942.

[8]. S. Wang, W. Liu, J. Wu, L. Cao, Q. Meng, P. J. Kennedy, Training deep neural networks on imbalanced data sets, in *Proceedings of the International Joint Conference on Neural Networks (IJCNN'16)*, 2016, pp. 4368-4374.

[9]. H. Wang, Z. Cui, Y. Chen, M. Avidan, A. B. Abdallah, A. Kronzer, Predicting hospital readmission via cost-sensitive deep learning, *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2018, Vol. 15, pp. 1968-1978.

[10]. T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense object detection, in *Proceedings of the IEEE International Conference on Computer Vision (ICCV'17)*, 2017.

[11]. T. Saito, M. Rehmsmeier, The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets, *PloS One*, Vol. 10, 2015, e0118432.

[12]. T. O'Malley, E. Bursztein, J. Long, F. Chollet, H. Jin, L. Invernizzi, et al., Keras tuner, *Retrieved May*, Vol. 21, 2019.

[13]. N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: A simple way to prevent neural networks from overfitting, *The Journal of Machine Learning Research*, Vol. 15, 2014, pp. 1929-1958.

[14]. US Government, Office of Inspector General. List of Excluded Individuals and Entities, https://oig.hhs.gov/exclusions/authorities.asp

[15]. R. Chalapathy, S. Chawla, Deep learning for anomaly detection: A survey, *arXiv Preprint*, 2019, arXiv:1901.03407.

[16]. J. M. Johnson, T. M. Khoshgoftaar, Survey on deep learning with class imbalance, *Journal of Big Data*, Vol. 6, 2019, 27.

[17]. Bilan 2018 des actions de lutte contre la fraude et actions de contrôles, https://www.ameli.fr/sites/default/files/2019-10-01-dp-controles-fraudes.pdf

(7159)

Building Health Monitoring Using Smartphone and Neural Networks

Francisco Vega and Wen Yu
Departamento de Control Automatico, CINVESTAV-IPN (National Polytechnic Institute),
Mexico City, Mexico
E-mail: yuw@ctrl.cinvestav.mx

Summary: In this work, a novel smartphone based SHM system is proposed, taking advantage of the features of smartphones such as that they are easy to handle inside the structure of a building and that they can also communicate through Wi-Fi networks. Sensors embedded in the phones are used to measure the linear acceleration on each floor of a building, this data is analyzed and processed for use in training a fuzzy LSTM network. This network, through its information generalization process, determines the presence of damage in the analyzed building. This network was selected because it has proven to be effective in handling the information, thus contributing to the creation of a reliable and robust SHM.

Keywords: Fuzzy deep neural network, Smartphone, Structural health monitoring.

1. Introduction

A structure can change its behavior depending on a sudden or gradual change in its state, the loading conditions and its response mechanisms, see [1]. Among the main advantages of developing an SHM system are: creating a safe environment, increasing the reliability of the operation, extending the useful life of the components of the structure, reducing the tests and maintenance of the structure, reducing the downtime of the use of the structure and, therefore, reduce the costs involved in its operation, see [2]. Structural health monitoring (SHM) has become a necessity in applications related to aerospace engineering, mechanical engineering, and civil engineering.

In this work we focus only on intelligent systems, NNs and fuzzy systems, because these directly address the uncertainties present in the measurement data through their fusification, defusification and learning processes. In addition, these models are well suited to SHM, because easy to interpret guides can be created due to the use of linguistic variables, speaking directly about fuzzy systems. In addition, diagnoses can be obtained in a relatively short time with a good degree of reliability, compared to other techniques used for the same purpose.

We developed a SHM for structural damage detection in buildings due to earthquakes, which is the reason why most of the previously mentioned articles refer to applications related to buildings. In the aforementioned articles good results were obtained with the use of intelligent systems, so the objective was to create an alternative that is easier to implement and with a lower cost compared to those works. We proposed the use of a new algorithm based on a deep NN known as long-short term memory (LSTM) and fuzzy systems, which is described by [13], it will estimate the damage in buildings with data obtained from smartphones (using the accelerometers that they have internally) located in different parts of the structure in question. All this is explained in the following sections, at the end of the paper an experiment with a test station designed to test structural damage detection algorithms is described and conclusions are given.

2. Structural Health Monitoring Using Smartphone

But the disadvantages of using smartphones for SHM are: it is necessary to design an application for data collection, it relies on Wi-Fi communication between devices (the cellphones, a modem and a computer), and smartphones battery has limit time.

The accelerometers of a smartphone are already configured inside, which can measure the linear acceleration in the X, Y and Z coordinate axes of the device. In our proposal we only consider the linear acceleration present in the X and Y axes, Fig. 1 shows a reference of the acceleration measurement, the acceleration in the Z axis is irrelevant, as the other two axes provide sufficient data for the SHM.

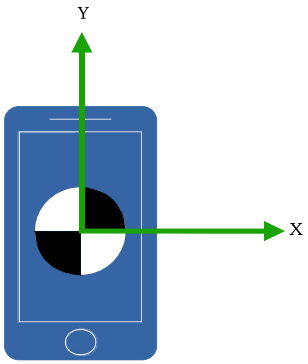


Fig. 1. Acceleration measurement due vibrations using a smartphone.

The structure of our test station has a ground floor and two floors. Two linear actuators located at the base of the station, they are used to emulate the seismic movement for the structure, see Fig. 2. The units in

which acceleration is measured are $\left[\frac{m}{s^2}\right]$. We use two smartphones, which have four acceleration measurements: the linear acceleration in the X and Y axes for the first floor and second floor.

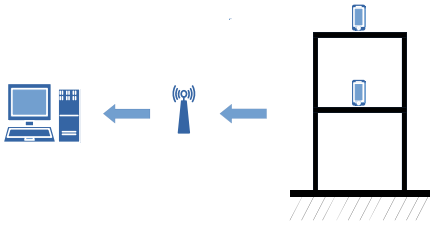


Fig. 2. Smartphone based SHM applied in the test station.

Our smartphone based SHM system is only a wireless system that operates locally. In our application, the operating system of the smartphones is Android, because it is open source and many smartphones in the market use it. Another advantage is that all internal peripheral devices can be accessed directly by using appropriate libraries in Android. For example, the data from the accelerometer and gyroscope in the smartphone can be read in Android operating system via a sensor manager service, which is configured to obtain data at regular intervals.

Before determining the presence of damage in the building, the data obtained from the smartphones must be processed by applying a signal reconstruction method and a filter to reduce measurement noise. Subsequently, the data that have already been processed are subjected to an analysis known as principal component analysis (PCA), which reduces the volume of data to work with and thus facilitates the detection of damage. The following section details the procedure described in Fig. 3.

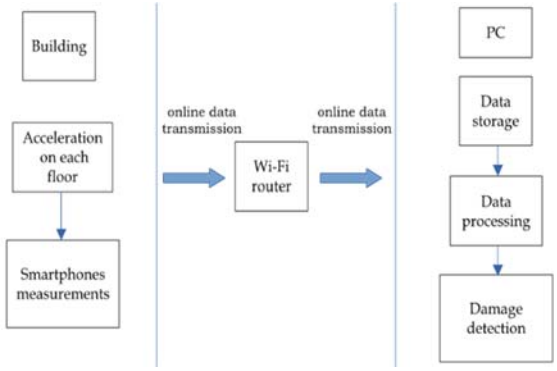


Fig. 3. Scheme of the operation of the smartphone based SHM system.

4. Fuzzy LSTM Networks

Structural health monitoring with deep neural networks is a classification problem. We use the data obtained with PCA to train a neural model, which represent a building as

$$y(k) = \varphi[U_r(k)], \tag{1}$$

where $\varphi(\cdot)$ is an unknown nonlinear difference equation, also the state vector $U_r(k)$, $u(k)$ and $y(k)$ are the input and the output signals for the system, n_y indicates the number of the delayed output signal, n_u indicates the number of the delayed input signal, and m indicates the number of elements u_{rm} in $U_r(k)$.

The concept of the fuzzy system using the LSTM cells is divided in 4 layers: in the first layer the inputs of the network are organized, in the second layer these inputs are fuzzificated, in the third layer the values of the IF and THEN parts are calculated, and in the fourth layer the estimation of the system is made according to (2).

The LSTM network has several stages, which are describe by:

$$\begin{aligned} F(k) &= \sigma\left(W^f U_r(k) + V^f H(k-1)\right) \\ I(k) &= \sigma\left(W^i U_r(k) + V^i H(k-1)\right) \\ S(k) &= \psi\left(W^s U_r(k) + V^s H(k-1)\right) \\ C(k) &= F(k) \otimes C(k-1) + I(k) \otimes S(k) \\ O(k) &= \sigma\left(W^o U_r(k) + V^o H(k-1)\right) \\ H(k) &= O(k) \otimes \psi(C(k)) \end{aligned} \tag{2}$$

where: $F(k)$, $I(k)$, $S(k)$, $C(k)$, $O(k)$ and $H(k) \in R^p$ are sections of the network, they are: the fitness of the internal state, the fitness of the internal input, the internal input, the internal state, the fitness of the output, and the output of the LSTM network, respectively. The synaptic weights are: W^f, W_i, W_s and $W_o \in R_{p\times m}$; V_f, V_i, V_s and $V_o \in R_{p\times p}$ as diagonal matrices or V^f, V^i, V^s and $V^o \in R^p$ as vectors, according to the need. The functions $\sigma(\cdot)$ and $\psi(\cdot)$ are the sigmoid and hyperbolic tangent functions, respectively, $U_r(k) \in R^m$ is the input in (1). From (2), the output of the fuzzy system is

$$\hat{y}(k) = Z_F H(k), \tag{3}$$

where $H(k) = [h_1(k) \cdots h_p(k)]^T$ corresponds to the THEN parts, $Z_F \in R^n$ is the elements of the IF parts. The number of LSTM cells, as well as the number of fuzzy rules, are defined as $p = \kappa^m$ for the case of 1 estimation, for several estimations it has $p = l(\kappa^m)$ where l is the number of estimations (thus $\hat{y} \in R^l$), as was described for (3).

According to function approximation theories of fuzzy systems, the identified nonlinear process (1) can be represented as:

$$Y(k) = Z_F(W^*)H(W^*) + \mu(k) \tag{4}$$

Training of the network

A variation of the BPTT algorithm is chosen to train the fuzzy system. We apply a narrow "window" to execute the training. This window only considers the values generated by the fuzzy LSTM network in the current iteration and its immediate past iteration. In this training method, the values generated by the network in the oldest iterations are forgotten. The training algorithm is defined by:

$$W(k+1) = W(k)+\eta_W \Delta W(k)+\alpha_W \Delta W(k-1) \quad (5)$$

where W is any synaptic weight array of the fuzzy LSTM, ΔW is the weight adjustment, $\eta_W \in (0,1]$ is the learning rate, $\alpha_W \in (0,1]$ is the momentum term for the training algorithm, and $\eta_W > \alpha_W$.

In (5), η_W determines the amount that increases or decreases each weight, while α_W helps to stabilize the modification by considering the past weight adjustment. The modeling error between the desired value and the fuzzy model is defined as:

$$\begin{aligned} \xi(k) &= \tfrac{1}{2} e^T(k)\, e(k) \\ E(k) &= \tfrac{1}{N} \textstyle\sum_{k=1}^{N} \xi(k), \end{aligned} \quad (6)$$

where $e(k)$ is the modelling error between the fuzzy model $y\hat{}(k)$ and the unknown plant $y(k)$, $\xi(k)$ is the instant error energy, $E(k)$ is the total energy during the whole processes, and N is the total number of iterations.

The modelling objective of the fuzzy system is $\min_{W(k)} \xi(k)$. The adjustment of each element of ΔW is defined as follows:

$$\Delta w_{ij}(k) = \frac{\partial \xi(k)}{\partial w_{ij}(k)}, \quad (7)$$

where $\xi(k)$ is defined in (6). A similar calculation is made for the adjustment of W^f, W^s, W^o, V^f, V^i, V^s and V^o. In other hand, for the premise part, for example, the adjustment of χ_j in the membership functions are:

$$\begin{aligned} \Delta\chi_j = &\tfrac{\partial \xi(k)}{\partial e(k)} \cdot \tfrac{\partial e(k)}{\partial \hat{y}(k)} \cdot \tfrac{\partial \hat{y}(k)}{\partial z_{Fj}}, \\ &\cdot \tfrac{\partial z_{Fj}}{\partial \zeta_j(k)} \cdot \tfrac{\partial \zeta_j(k)}{\partial \chi_j} \end{aligned} \quad (8)$$

and in a vectorial form:

$$\begin{aligned} \Delta\chi_j &= (U_r(k) - \chi_j) \otimes \Upsilon_j \otimes D_\chi \otimes e(k)\, H(k) \\ D_\chi &= \exp\left[(U_r(k) - \chi_j)^2 \otimes \left(-\tfrac{1}{2}\Upsilon_j\right)\right] \end{aligned}$$

Also, something similar for Υ_j is done to compute its adjustment. We are only interested in open-loop identification, we assume that the plant (1) is bounded-input and bounded output stable, *i.e.*, $y(k)$ and $U_r(k)$ in (1) are bounded. The following theorem gives a stable gradient descent training algorithm for the fuzzy neural model.

4. Experimental Results

The proposal SHM system was carried out experimentally in a test station (the building) for the development of SHM algorithms in buildings. The smartphones are placed at the center of each floor in a fixed position, the battery life of the devices averaging twelve hours. The devices are SONY Xperia S with Android 2.3, 1.5 GHz Qualcomm Dual Core processor, and 1 G RAM. The parameters of the inertial sensors were determined by the experiments. The 0 g-offset of the accelerometer is 0.12 V (x-axis), 0.071 V (y-axis), 0.094 V (z-axis). The sensors of the phones include an accelerometer, an electromagnetic field sensor used as proximity sensor, an ambient light sensor, a magnetometer, a GPS sensor, and a gyroscope. We only use the accelerometer.

The router is Next Nebula 150 with four 10/100 Mbps LAN ports and Wireless IEEE802.11n ports. The computer we use has Windows 10 as its operating system, the version of MATLAB used is the 2020 version.

The experiment we proposed consists of applying the equivalent of an earthquake movement to the building to record its behaviour, taking conditions of damage and nondamage in the building. With this information we worked as established before. Having two conditions, we created two models with the fuzzy LSTM network, then we compared them to see which one is better to diagnose the damage in the building, making use of data records in different situations in the building.

In both cases, four measurements were taken with a sampling time of 0.002 s, during a time period of 60 s, generating a total of 30,001 iterations for each variable. The variables correspond to the linear acceleration in the X and Y coordinate axes, both for the first and second floor according to the Fig. 3. By applying the methodologies presented for the vibration data, we generated to two vectors of 30,001 elements each, the data of these vectors are dimensionless.

Then, by applying PCA to the measurement data that have been reconstructed and filtered with the LPF, the information is defined by the first two elements of the PCs, out of the four we had, as can be seen in Fig. 1 but only are represented three PCs because the fourth does not provide relevant information, and in Fig. 4 is shown the measurement data as a function of the first two PCs due to these provide the majority of the information of the measurement.

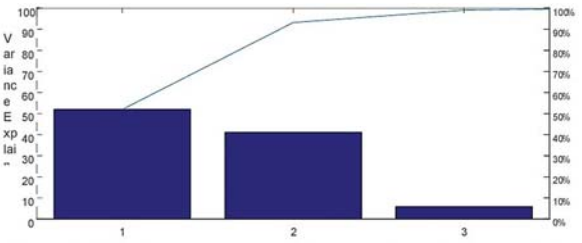


Fig. 4. Variance percentage of each smartphone.

In the comparisons, it turned out that the network fed during its learning process with the undamaged data (information of when the building suffers an earthquake but has no previous damage and is not damaged) is the most suitable for damage detection. This is because the network, when trained with such data, is more sensitive to sudden based SHM system changes in the behavior of the building, which

translates as structural damage. It should be clarified that although each floor is measured, it is not possible to say in which part of the building the damage is with this proposal, the accuracy in the location of the structural damage was sacrificed for the speed of the algorithm to detect damage.

With the new information the fuzzy network was trained, defining its input as $U_r = [u_{r1}(k-1), u_{r2}(k-1)]^T$, each element of the input vector is the delayed value of one iteration of the data shown in Fig. 5. The performance of the network during its training process is shown in Fig. 4, as the process proceeds, the network emulates the process it learns from. The error shown in Fig. 6 corresponds to Eq. (8) and indicates how well the network learned, in this case the process was performed correctly, as the error tends to zero as the process continues data, as these data compared to the previous ones have a different and more abrupt dynamics due to the occurrence of damage to the building structure. This is the part that represents the damage detection, which results in a quick analysis after the data processing given to the measurements obtained from the mobile phones. Taking as a reference Fig. 6, which behaves in a similar way to that established for Fig. 6, the sensitivity of the network to data with damage is better appreciated. At the beginning of the graph, when there is no structural damage, the error remains close to zero, but when damage is generated in the structure, the error increases. This change is what indicates the presence of damage in the building. It is remarkable that despite the high sensitivity of the network to the change in the building dynamics due to the damage, the network still emulates the behavior of the building data in an acceptable way.

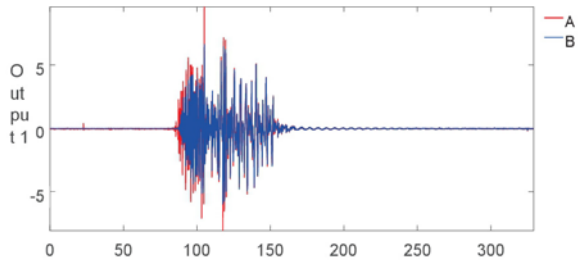


Fig. 5. Represents the measurement data in function of the first smartphone.

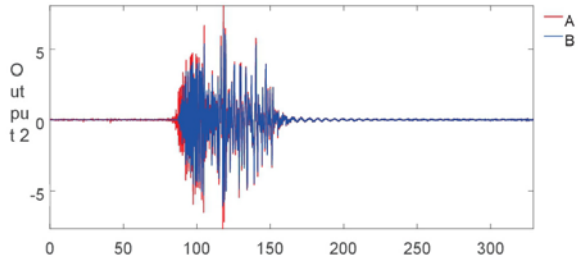


Fig. 6. Represents the measurement data in function of the second smartphone.

4. Conclusions

In this paper, a smartphone-based building structural health monitoring system is developed. To create a low-cost and easy-to-implement alternative, considering online data, we use smartphones to collect data on the movement of a building to determine whether it has suffered any damage. The information obtained is processed through recurrent fuzzy networks of long- and short-term memory, which are responsible for determining the damage to the structure. This proposal was implemented in a test station of structural health monitoring algorithms, obtaining favorable results.

Although this method only indicates the presence of damage, it is simple and quick to apply, as well as low cost and with great flexibility. This may allow future modifications for deeper analysis.

References

[1]. O. Cordon, A historical review of evolutionary learning methods for Mamdani-type fuzzy rule-based systems: Designing interpretable genetic fuzzy systems, *International Journal of Approximate Reasoning,* Vol. 52, Issue 1, 2011, pp. 894-913.

[2]. E. Lee, L. Heung, ANN-based structural damage diagnosis using measured vibration data, in *Proceedings of the International Conference on Knowledge-Based and Intelligent Information and Engineering Systems (KES'04)*, 2004, pp. 373-379.

[3]. I. Lopez, N. Sarigul-Klijn, A review of uncertainty in flight vehicle structural damage monitoring, diagnosis and control: Challenges and opportunities. *Progress in Aerospace Sciences,* Vol. 46, Issue 7, 2010, pp. 247-273.

[4]. J. Shao-Fei, C. Zhang, S. Zhang, Two-stage structural damage detection using fuzzy neural networks and data fusion techniques. *Expert Systems with Applications,* Vol. 38, Issue 1, 2011, 511519.

[5]. L. Meyyappan, et al., Fuzzy-neuro system for bridge health monitoring, in *Proceedings of the 22nd International Conference of the North American Fuzzy Information Processing Society (NAFIPS'03)*, 2003, pp. 8-13.

[6]. M. Gordan, et al., Recent developments in damage identification of structures using data mining, *Latin American Journal of Solids and Structures,* Vol 14, Issue 13, 2017, pp. 2373-2401.

[7]. M. Pawar, R. Ganguli, Genetic fuzzy system for online structural health monitoring of composite helicopter rotor blades, *Mechanical Systems and Signal Processing,* Vol. 21, Issue 5, 2007, pp. 2212-2236.

[8]. S. Ramu, V. Johnson, Damage assessment of composite structures fuzzy logic integrated neural network approach, *Computers & Structures,* Vol. 57, Issue 3, 1995, pp. 491-502.

[9]. J. Sawyer, S. Rao, Structural damage detection and identification using fuzzy logic, *AIAA Journal,* Vol. 38, Issue 12, 2000, pp. 2328-2335.

[10]. K. Soh, S Bhalla, Calibration of piezo-impedance transducers for strength prediction and damage assessment of concrete, *Smart Materials and Structures*, Vol. 14, Issue 4, 2005, 671.

[11]. M. Taha, J. Lucero, A generic fuzzy metric for damage recognition in structural health monitoring systems, in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics (ICSMC'05)*, Vol. 2., 2005, pp. 1518-1523.

[12]. S. Thenozhi, W. Yu, Advances in modeling and vibration control of building structures, *Annual Reviews in Control,* Vol. 37, Issue 2, 2013, pp. 346-364.

[13]. W. Yu, F. Vega, Nonlinear system modeling using the Takagi-Sugeno fuzzy model and long-short term memory cells, *Journal of Intelligent & Fuzzy Systems* Vol. 39, Issue 3, 2002, pp. 4547-4556.

[14]. F. Zhu, Y. Wu, A rapid structural damage detection method using integrated ANFIS and interval modeling technique, *Applied Soft Computing*, Vol. 25, Issue 1, 2014, pp. 473-484.

(8191)

How to Leverage DNN-based Speech Enhancement for Multi-channel Speaker Verification?

Sandipana Dowerah [1], Romain Serizel [1], Denis Jouvet [1], M. Mohammadamini [2], Driss Matrouf [2]
[1] Université de Lorraine, CNRS, Inria-Loria., Nancy, France
[2] Avignon Université, Laboratoire Informatique d'Avignon, Avignon, France
E-mails: name.surname@loria.fr, name.surname@univ-avignon.fr

Summary: Speaker verification (SV) suffers from unsatisfactory performance in far-field scenarios due to environmental noise and the adverse impact of room reverberation. This work presents a benchmark of multichannel speech enhancement for far-field speaker verification. One approach is a deep neural network-based, and the other is a combination of deep neural network and signal processing. We integrated a DNN architecture with signal processing techniques to carry out various experiments. Our approach is compared to the existing state-of-the-art approaches. We examine the importance of enrollment in pre-processing, which has been largely overlooked in previous studies. Experimental evaluation shows that pre-processing can improve the SV performance as long as the enrollment files are processed similarly to the test data and that test and enrollment occur within similar SNR ranges. Considerable improvement is obtained on the generated and all the noise conditions of the VOiCES dataset.

Keywords: Multichannel speech enhancement, Far-field speaker verification, Deep neural network.

1. Introduction

Speaker verification (SV) authenticates a person's identity based on his/her voice characteristics. Despite significant improvements in deep learning-based SV in close-talk or controlled scenarios, SV still suffers from unsatisfactory performance in far-field/ distant scenarios. Speech signals propagating over long distances are subject to fading, absorption, and reflection by various objects, which changes the pressure level at different frequencies and degrades speech quality. In real cases, these acoustic perturbations make far-field SV a challenging task. Several challenges have been organized over the past few years to address this problem, such as VOiCES from a distance challenge [12], Interspeech far-field speaker verification challenge [13], etc. The current state-of-the-art x-vector-based [8] approaches improved the SV performance significantly. But these SV systems still suffer from severe performance degradation in noisy-reverberant scenarios that are typical of hands-free applications.

Speech enhancement can be used to improve the perceptual quality of speech by estimating clean speech signals for signals impacted by acoustic noise and reverberation. Besides denoising autoencoder [14, 15], neural beamforming [16], and dereverberation [17] have been extensively used as front-end processing of speech recognition [16, 18, 19]. But only a few studies have examined the effectiveness of integrating beamforming and dereverberation with multichannel signals for SV in a noisy-reverberant environment [11, 20]. Mosner et al. employed mask-based beamforming combined with WPE to minimize the reverberation effect, but they studied only the reverberation effect, whereas reverberation and noise occur simultaneously in real scenarios [11]. Yang et al. jointly optimized neural networks that supported minimum variance distortion less response (MVDR) beamforming with WPE using a deep speaker embedding model [21]. Taherain et al. used an MVDR beamformer with Rank-1 approximation to search for the optimal beamformer from the variants of ideal ratio mask-based MVDR and generalized eigenvalue (GEV) beamformers [11]. Although often used in a multichannel context, most of these studies use single-channel data as an input to DNN, use matched train/test data, and report poor performance on real data. Moreover, prior works mainly used mask-based beamformers (MVDR or GEV) in the frequency domain, which typically degrades in causal and online scenarios [1] as frequency domain methods lack the reasonable size of frequency resolution and input signal length required for perceivable system latency.

This paper studies the benchmark speech enhancement as a multichannel pre-processing to SV in adverse acoustic conditions where noise and room reverberation distorts the target speech signal. We consider either filtering based on a deep neural network (DNN) or combining DNN and signal processing approaches. The DNN-based approach implements FaSnet [1], a state-of-the-art neural beamforming technique for speech enhancement. The second approach is an integration of FaSnet with Rank 1 multi-channel Wiener filter [2] and a de-reverberation algorithm [3]. We compared our pre-processing approach to the popular state-of-the-art pre-processing approaches from [11]. Furthermore, this work studies the impact of both approaches in different noisy and reverberated acoustic scenarios using various signal-to-noise (SNR) ratios for a multi-channel input signal. We also study the impact of data mismatch, robustness in low SNR scenarios, and generalization to unseen real recorded data. Additionally, we investigate the influence of quality

(in terms of source to distortion and source to interference ratio) of the enhanced signals, which could be helpful in fine-tuning the front end of an SV system.

2. Use Case

The main purpose of this work is to solve the various challenges faced by a mobile security robot in the context of an SV. The performance of SV reduces drastically due to ambient noise leading to low SNR, internal robot noises leading to reverberation further degrading the SNR, and overlapping background speech. The main focus is on developing a multichannel speech enhancement as a pre-processing to the SV in the context of mobile security robot to identify a person on industrial premises during the inactivity period.

3. Problem Formulation

3.1. Signal Model

Considering the mixture of dry speech and noise as recorded by ***K*** microphones can be formulated with the short-time Fourier transform STFT as ***y(T-F) = s(T-F) + h(T-F) + n(T-F)***, where ***y(T-F), s(T-F), h(T-F)*** and ***n(T-F)*** represent the STFT vectors of the noisy speech, dry speech, reverberated speech, and noise.

4. Multichannel Speech Enhancement

This section explains the integrated multichannel speech enhancement approach we developed for far-filed SV.

4.1. FaSNet

FaSnet (filter-and-sum network) is a filter-based beamforming approach suitable for real-time low-latency applications [1]. FaSnet incorporates a two-staged architecture. A beamforming filter for a chosen reference channel is computed in the first stage. The reference channel is randomly selected. The second stage uses the output filter from the first stage to estimate the beamforming filters for the rest of the channels. The input for both stages includes the target channel as well as the output of the normalized cross-correlation between channels as an inter-channel feature. Both stages use the temporal convolutional networks (TCN), enabling the lower latency processing of the FaSnet model. The training objective of the FaSnet model is to select a signal-level loss criterion based on the actual task needed to be solved.

We give noisy multichannel signals as input to FaSnet to separate noise and speech. We used FaSnet to obtain a first estimate of the speech signal s(T-F) and the noise signal n(T-F). These estimates are then used to compute the T-F masks:

$$\boldsymbol{M_s}(\boldsymbol{T}-\boldsymbol{F}) = \frac{|\boldsymbol{S}(\boldsymbol{T}-\boldsymbol{F})|}{|\boldsymbol{S}(\boldsymbol{T}-\boldsymbol{F}) + \boldsymbol{max}(|\boldsymbol{n}(\boldsymbol{T}-\boldsymbol{F})|, \boldsymbol{\varepsilon})}, \quad (1)$$

$$\boldsymbol{M_n}(\boldsymbol{T}-\boldsymbol{F}) = \frac{|\boldsymbol{n}(\boldsymbol{T}-\boldsymbol{F})|}{|\boldsymbol{S}(\boldsymbol{T}-\boldsymbol{F}) + \boldsymbol{max}(|\boldsymbol{n}(\boldsymbol{T}-\boldsymbol{F})|, \boldsymbol{\varepsilon})}, \quad (2)$$

where $\varepsilon = 1\times10^{-16}$.

4.2. Rank-1 MWF

MWF is designed to minimize the mean squared error (MSE) criterion between the record mixture and the target speech.

$$J(w) = E\{|s_1 - w^H y|^2\}, \quad (3)$$

where s_1 is the clean speech signal from the first channel, E is the expectation operator, and.^{H} denotes the Hermitian transpose. The filter $\boldsymbol{w}$ that minimizes the MSE criterion [equation number] is the MWF that can be expressed as below;

$$\widehat{W}_{MWF}(f) = R_{ss}(f) + \\ + R_{nn}(f)^{-1} R_{ss}(f)\, u_1, \quad (4)$$

where R_{ss}, $R_{nn}(f)$ are spatial correlation matrix for the speech and noise, respectively and $u_1 = [1,..., 0]^T$.

It is possible to introduce a trade-off parameter μ which controls the tradeoff between the interference reduction and the desired signal distortion [23]. We then obtained the speech distortion weighted (SDW) MWF that can be expressed as;

$$\widehat{W}_{SDW-MWF}(f) = \\ = R_{ss}(f) + \mu R_{nn}(f)^{-1} R_{ss}(f)\, u_1 \quad (5)$$

If the desired signal comes from a single source, the speech correlation matrix R_{ss} is theoretically of Rank-1. Forcing this matrix to its Rank-1 approximation leads to the so-called Rank-1 version of the filters described above. In the remainder of the paper, we use the Rank-1 approximation of the SDW-MWF.

The computation of MWF requires the estimation of the speech and noise correlation matrices. The estimated T-F masks of speech and noise are used to compute the spatial correlation matrices $R_{ss}(f)$ and $R_{nn}(f)$ that are needed to derive the MWF. The correlation matrices are obtained as

$$R_{ss}(f) = \frac{1}{T}\sum_{t=0}^{T-1} \hat{s}(T-F)\, \hat{s}(T-F)^H \quad (6)$$

Note that the noise correlation matrix can be obtained similarly as in Eq. (6).

4.3. Weighted Prediction Error (WPE)

WPE is used for alleviating degradation performance in speech recognition, mostly in the case of a far-field

scenario. The de-reverberated signal is obtained by subtracting the filtered signal from the observed signal denoted by

$$d(\hat{s}) = \hat{s}(t) - \sum_{k=1}^{N} \hat{w}(k)h(T-k),$$

where $\hat{s}$ is reverberated signal at time t and $d(\hat{s})$ is de-reverberated signal using WPE algorithm. $\hat{s}$ denotes the k^{th} tap of the N-taps. WPE filter is $W = [W_1, ..., W_N]^{\text{T}}$.

5. Datasets

5.1. Synthetic Dataset

We generated a synthetic dataset, namely, RoboVoices simulating real room environments with additive noise and reverberation from dry speech segments. Designing such a dataset is necessary as training speech enhancement approaches require ground-truth knowledge about the target speech and, to some extent, the degradation. This information is not available in the available corpora for far-field SV.

5.1.1. Speech Data

We use the dry speech data from the clean subset of the Librispeech [4] corpus, which is approximately 1000 hours of English speech data collected as part of the Librivox project. We randomly selected around 10000 files from the dry training subset of Librispeech and truncated them to 10 seconds duration for the training set, contributing to 25 hours of speech data.

For the evaluation of the SV system, we use the Fabiole speech corpus [5]. Fabiole is a French speech corpus consisting of around 6882 audio files from 130 native French speakers. The minimum duration of the speech file is 1 second, and the maximum is 46 seconds. The speech data of the corpus is collected from different French radio and television shows. For creating each evaluation set, we have used 1200 speech files from Fabiole representing 2 hrs. of evaluation material.

5.1.2. Noise Data

We have collected realistic office noise from the Freesound platform [6]. The selected noise categories include door, keyboard, office, phone, background noise in the room, printer, fan, door knock, babble, environmental noise, etc. We split the dataset into a training set of 3725 clips and an evaluation set of 1000 clips.

We also evaluate our system's performance using MUSAN noise from the OpenSRL dataset [9]. We convolved the dry speech from Librispeech and noise from Musan with simulated RIR for training. The evaluation protocol is the same as RoboVoices except for the noise samples. The noise categories include dial tones, raindrops, etc.

5.1.3. Room Impulse Response

To simulate room effects, we have generated an RIR corpus of 10000 rooms for training and 3600 for evaluation with the pyroomacoustics toolbox [7]. For training, the room length was chosen between 3-8 m, the width was chosen between 3-5 m, and the height was chosen between 2-3 m. The absorption coefficient was drawn randomly such that the room's RT-60 was between [200-600] ms. The minimum distance between a source and the wall is 1.5 m, and 1 m between the wall and the microphones. The RIR for the evaluation set was generated with the same room dimension as in the training set, but the absorption coefficient was selected to obtain an RT-60 of 400 ms.

The final RoboVoices corpus for training and evaluation is created by first convolving the dry speech and noise with the simulated RIRs. We then added the convolved dry speech and convolved noise to obtain the noisy signal. We randomly select the noise samples from Freesound and the dry speech from Librispeech for the training set. The SNR is drawn randomly with a uniform distribution between 0-10 dB. For the evaluation set, the generation process is similar, except that we draw the SNR values in 5, 10, and 20 dB, and the process is applied to each speech segment from the Fabiole dataset. In total, we have generated 10000 mixtures for training and 3600 mixtures for evaluation.

5.2. VOiCES

We evaluate our approach to the VOiCES challenge 2019 dataset [12]. Among 11 microphone positions in the Eval set, we select three representative positions: 2, 4, and 9. We select the signal from these three microphones confirming all three are in mid-distance from the speaker and are close to building a "virtual" microphone antenna.

6. Experimentation

6.1. Experimental Set-up

The speech and noise signals are sampled at 16 kHz. We provide multichannel speech signal as input to FaSNet with a 4 ms window size and context size of 16 ms. We trained the FaSNet model with SDR loss and SI-SNR (scale-invariant source-to-noise ratio) loss [24]. We employed the dual-path RNN (DPRNN) with an encoder dimension of 50, a chunk size of 50, and a hopping window of 35 dimensions. We use the source-separated outputs from the FaSNet model to compute the target masks. The FaSNet implementation is used from the Asteroid toolbox [25] and replaced the TCN blocks with DPRNN in contrast to the original FaSNet architecture, where TCN is used to predict the beamformed filters.

The SDW-MWF operates on the T-F representation of the signal. STFT is computed with a

window length of 512 samples, a hop size of 256 samples, and a Hann window. A single SDW-MWF is estimated for each speech clip. According to previous experiments, we set the **μ** parameter of the SDW-MWF to 0.1 to limit the amount of distortion introduced by the filter. We use WPE with the following parameters: 10 filter taps, a delay of 3 frames, 5 iterations of WPE, and an alpha of 0.9999.

6.1.1. Speaker Verification

Our SV is an x-vector-based system. The network is trained with data augmentation using different portions of Musan corpus (music, babble, noise, reverberation) [9] with 1 million augmented files from Voxceleb [26] and all the original files from Voxceleb 1 and 2 [10]. We use the Fabiole corpus for tests and enrollment. For enrollment, 3441 files are used, and the remaining files are used for the test. As input to the x-vector network, we extract Mel-frequency cepstral coefficients normalized by Cepstral Mean-Variance Normalization. We removed the non-speech frames with a voice activity detector. The Probabilistic Linear Discriminant Analysis (PLDA) classifier used for scoring is trained on 200k x-vectors extracted from Voxceleb. Before training the PLDA, x-vectors are centered, and their dimensionality is reduced to 128 with linear discriminant analysis. The PLDA scoring system is retrained on the enrollment set. Kaldi[6] is used to process all the steps of SV.

6.2. Evaluation

The SV system is evaluated using an equal error rate (EER). The bootstrap algorithm presents all metrics with a 95 % confidence interval [27]. We compute EER on dry speech and reverberated speech (as a reference point), the input mixture, and the signals estimated with different speech enhancement algorithms.

7. Results and Analysis

Table 1 shows the results of our experiments using different pre-processing approaches on the RoboVoices dataset on various SNR conditions. We implement the BLSTM-based approaches from [11] and consider them as the baseline. FaSnet degrades the SV performance as FaSnet introduces artefacts that distort the signal quality. We observed that integrating FaSnet with signal processing methods consistently improves against the baseline systems BLSTM MVDR Rank-1 and BLSTM GEV-BAN. Error reduction in multi-channel SV is greater, especially in low SNR, with a 7 % reduction at 5 dB showing robustness to low SNR conditions for an unprocessed noisy-reverberated signal. Thus, supporting the argument that Rank-1 MWF is robust to low SNR scenarios.

Table 1. EER (%) on RoboVoices using different pre-processing methods. The confidence interval is 0.1.

Pre-processing/SNR	5	10	20
Unprocessed	34.4	28.0	22.2
FaSnet [1]	45.7	39.0	31.5
BLSTM MVDR Rank1 [11]	32.3	26.6	22.1
BLSTM GEV-BAN [11]	32.5	26.8	21.9
FaSnet Rank1 MWF WPE	**27.1**	**23.2**	**19.7**

Table 2 presents the results obtained on the publicly available VOiCES Eval dataset [12] for various distractor noise conditions. We selected the microphone which was closest to the speaker as a reference microphone. As expected, the condition with no noise distractor (Clean in Table 2) resulted in the best performance across the approaches. The baseline BLSTM-based approaches perform poorly compared to the FaSnet-based approaches in all the noise conditions. With an EER of 9.2 % without any pre-processing, Babble seems to be the most challenging condition due to overlapping speech interference as well as its similarity to the desired clean speech. The proposed system improved the performance of Babble with an EER of 6.3 %. Furthermore, FaSnet Rank-1 MWF WPE achieves the best performance across the noise conditions, demonstrating our approach's efficacy even though the model was trained on synthetic data generated for generic, possibly mismatched, and spatial scenarios. We have also experimented with enrollment in match pre-processing conditions showing its impact in SV.

Table 2. EER (%) on different noise conditions of the VOiCES Eval dataset. The confidence interval is 0.2.

	Clean	Babble	TV	Music
Unprocessed	4.4	9.2	7.9	8.4
FaSnet	4.4	7.8	7.4	7.9
MVDR Rank1	4.3	7.3	6.5	6.9
FaSnet Rank1 MWF WPE	**4.0**	**6.3**	**6.0**	**6.4**

Table 3 reports the performance on the RoboVoices dataset for different pre-processing conditions and depending on the enrollment condition. Performing the enrollment and test with matched acoustic conditions alleviates the effect of reverberation, but this is hardly the case for additive noise. Pre-processing consistently improves the SV performance, but the effectiveness is more evident when the enrollment is done in matched pre-processing conditions (diagonal). FaS Rank-1 MWF WPE obtained the best EER performance for a noisy and reverberated input over the baseline approach.

[6] https://github.com/kaldi-asr/kaldi

Table 3. EER (%) on matched pre-processing conditions on the RoboVoices dataset. We processed both enrollment and test data using the same range of SNR. The average confidence interval is 0.1.

Test data	Dry speech	Reverb. Speech	Noisy	BLSTM MVDR Rank- 1	FaSnet Rank-1 MWF WPE
Dry speech	**14.9**	15.4	16.7	16.1	15.7
Reverb. speech	20.6	**19.8**	20.5	20.4	20.1
Noisy	28.2	24.9	**23.8**	24,9	24,3
BLSTM MVDR Rank -1	27.0	24.2	23.4	**21.3**	22.5
FaSnet Rank-1 MWF WPE	23.3	22.8	21.5	22.4	**19.2**

8. Conclusion

This work presents the benchmark speech enhancement pre-processing approach to multi-channel speaker verification in a far-field noisy-reverberated environment. We experimented with both DNN and a combination of DNN with signal processing methods as a front end to the state-of-the-art x-vector speaker verification system. Experimental evaluations on synthetic and VOiCES datasets show that combining DNN with signal processing methods significantly improves speaker verification performance. Moreover, the combined DNN and signal processing approach show more robustness to low SNR scenarios. Additionally, experimentation with enrollment shows that performing the test and enrollment with matched acoustic conditions alleviates the effect of reverberation. Our approach demonstrated the best performance across the noise conditions on the VOiCES dataset even though the model was trained on synthetic data. This shows that our approach generalizes to unseen real recorded data.

Acknowledgements

French National Research Agency supports this work for the ROBOVOX project (ANR-18-CE33-0014)). Experiments were partially carried out using the Grid5000 testbed supported by a scientific group of INRIA, including CNRS, RENATER, and other Universities and organizations hosted by the University of Lorraine.

References

[1]. Y. Luo, E. Ceolini, C. Han, S-C. Liu, N. Mesgarani, FaSNet: Low-latency adaptive beamforming for multi-microphone audio processing, in *Proceedings of the IEEE Automatic Speech Recognition and Understanding Workshop (ASRU'19)*, Sentosa, Singapore, 14-18 December 2019, pp. 260-267.

[2]. W. Ziteng, V. Emmanuel, S. Romain, Y. Yonghong, Rank-1 constrained multichannel wiener filter for speech recognition in noisy environments, *Computer Speech & Language*, Vol. 49, 2017, pp. 37-51.

[3]. T. Yoshioka, T. Nakatani, Generalization of multi-channel linear prediction methods for blind MIMO impulse response shortening, *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 20, 2012, pp. 2707-2720.

[4]. V. Panayotov, G. Chen, D. Povey, S. Khudanpur, Librispeech: An ASR corpus based on public domain audiobooks, in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'15),* Queensland, Australia, 19-24 April 2015, pp. 5206-5210.

[5]. M. Ajili, J. Bonastre, J. Kahn, S. Rossato, G. Bernard, Fabiole, a speech database for forensic speaker comparison, in *Proceedings of the International Language Resources and Evaluation (LREC'16),* Portorož, Slovenia, 23-28 May 2016, pp. 726-733.

[6]. E. Fonseca, J. Pons, X. Favory, F. Font, D. Bogdanov, A. Ferraro, S. Oramas, A. Porter, X. Serra, Freesound datasets: A platform for the creation of open audio datasets, in *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR'17)*, Suzhou, China, 23-27 October 2017, pp. 486-493.

[7]. R. Scheibler, E. Bezzam, I. Dokmani, Pyroom-acoustics: A python package for audio room simulation and array processing algorithms, in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'18),* Calgary, Canada, 15-20 April 2018, pp. 351-355.

[8]. D. Snyder, D. G. Romero, G. Sell, D. Povey, S. Khudanpur, X-vectors: Robust DNN embeddings for speaker recognition, in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'18),* Calgary, Canada, 15-20 April 2018, pp. 5329-5333.

[9]. D. Snyder, G. Chen, D. Povey, Musan: A music, speech, and noise corpus, *ArXiv Preprint*, 2015, abs/1510.08484, 2015.

[10]. J. S. Chung, A. Nagrani, A. Zisserman, Voxceleb2: Deep speaker recognition, in *Proceedings of the INTERSPEECH Conference*, Hyderanab, India, 2-6 April 2018, pp. 1086-1090.

[11]. H. Taherian, Z. Wang, D. Wang, Deep learning based multichannel speaker recognition in noisy and reverberant environments, in *Proceedings of the INTERSPEECH Conference*, Graz, Austria, 15-19 September, pp. 4070-4074.

[12]. M. K. Nandwana, J. V Hout, C. Richey, M. McLaren, M. Barrios, A. Lawson, The VOiCES from Distance Challenge, in *Proceedings of the Hoot INTERSPEECH Conference*, Graz, Austria, 15-19 September, pp. 2438-2442.

[13]. X. Qin, M. L, Hui Bu, W. Rao, R. K. Das, S. S. Narayanan, H. Li, The INTERSPEECH Conference 2020 far-field speaker verification challenge, in *Proceedings of the INTERSPEECH Conference,* Shanghai, China, 25-29 October 2020, pp. 3456-3460.

[14]. O. Plchot, L. Burget, H. Aronowitz, P. Matejka, Audio enhancing with DNN autoencoder for speaker recognition, in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'16)*, Shanghai, China, 20-25 March 2016, pp. 5090-5094.

[15]. C. Yu, R. E. Zezario, S. Wang, J. Sherman, Y. Hsieh, X. Lu, H. Wang, Y. Tsao, Speech enhancement based on denoising autoencoder with multi-branched encoders, *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, Vol. 28, 2020, pp. 2756-2769.

[16]. J. Heymann, L. Drude, R. Haeb-Umbach, Neural network based spectral mask estimation for acoustic beamforming, in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'16),* Shanghai, China, 20-25 March 2016, pp. 196-200.

[17]. Y. Yoshioka, T. Nakatani, Generalization of multi-channel linear prediction methods for blind MIMO impulse response shortening, *IEEE Transactions on Audio, Speech, and Language Processing,* Vol. 20, 2012, pp. 2707-2720.

[18]. C. Boeddeker, H. Erdogan, T. Yoshioka, R. Haeb-Umbach, Exploring practical aspects of neural mask-based beamforming for far-field speech recognition, in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'18)*, Calgary, Canada, 15-20 April 2018, pp. 6697-6701.

[19]. K. Kinoshita, M. Delcroix, H. Kwon, T. Hori, T. Nakatani, Neural network-based spectrum estimation for online WPE dereverberation, in *Proceedings of the INTERSPEECH Conference*, Stockholm, Sweden, 20-24 August 2017, pp. 384-388.

[20]. L. Mosner, P. Matejka, O. Novotny, J. H. Cernocky, Dereverberation and beamforming in far-field speaker recognition, in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'18)*, Calgary, Canada, 15-20 April 2018, pp. 5254-5258.

[21]. J. Y. Yang, J. H. Chang, Joint optimization of neural acoustic beamforming and dereverberation with X-vectors for robust speaker verification, in *Proceedings of the INTERSPEECH Conference,* Graz, Austria, 15-19 September, pp. 4075-4079.

[22]. S. M. Golan, S. Gannot, I. Cohen, Performance of the SDW-MWF with randomly located microphones in a reverberant enclosure, *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 21, 2013, pp. 1513-1523.

[23]. S. M. Golan, S. Gannot, I. Cohen, Performance of the SDW-MWF with randomly located microphones in a reverberant enclosure, *IEEE Transactions on Audio, Speech, and Language Processing,* Vol. 21, 2013, pp. 1513-1523.

[24]. J. Le Roux, S. Wisdom, H. Erdogan, J. R. Hershey, SDR – half-baked or well done?, in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'19),* Brighton, United Kingdom, 12-17 May 2019, pp. 626-630.

[25]. M. Pariente, S. Cornell, J. Cosentino, S. Sivasankaran, E. Tzinis, J. Heitkaemper, M. Olvera, F. R. Stoter, M. Hu, J. M. Martin-Donas, D. Ditter, A. Frank, A. Deleforge, E. Vincent, Asteroid: The PyTorch-based audio source separation toolkit for researchers, *ArXiv Preprint*, 2020, abs/2005.04132.

[26]. A. Nagrani, J. S. Chung, A. Zisserman, Voxceleb: A large-scale speaker identification dataset, in *Proceedings of the INTERSPEECH Conference*, Stockholm, Sweden, 20-24 August 2017, pp. 2616-2620.

[27]. M. Bisani, H. Ney, Bootstrap estimates for confidence intervals in ASR performance evaluation, in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'04)*, Montreal, Canada, 17-21 May 2004, pp. I-409.

(8235)

A Comparison of Wood Log Dissimilarities to Predict Sawmill Output with k-Nearest Neighbor Algorithms

S. Chabanet, M. Dumas, H. Bril El-Haouzi and Philippe Thomas
CRAN, Université de Lorraine, CNRS, Epinal, F-88000, France
E-mail: sylvain.chabanet@univ-lorraine.fr

Summary: In the sawmill industry, to predict the set of lumber that would be sawed from specific wood logs is a difficult problem. Even if they exist many sawmill simulators able to simulate the sawing process in order to predict these quantities, they can be too slow for large scale industrial problems. Replacing these simulators with machine learning surrogate models, or metamodels, is a promising avenue of research to speed up predictions. One such research direction is based on the computation of pairwise dissimilarities between logs, used, for example, by k-nearest neighbor algorithms. Interesting results have been obtained with the so-called iterative closest point (ICP) dissimilarity who has, however, several undesirable properties. This paper explores another alternative based on ensemble of shape functions.

Keywords: Sawmills, ICP dissimilarity, Ensemble of shape functions, k-Nearest neighbors, Simulation metamodeling.

1. Introduction

Sawmills are key elements of the forest product industry transforming wood logs into various lumber. Several factors, including heterogeneity of the raw material, introduce uncertainty on the mix of lumber that can be obtained from sawing a batch of logs. For this reason, academics and industrials have developed sawing simulators that are able to simulate the sawing of individual logs based on a description of their shapes, and sometimes internal defects. Shape information commonly comes in the form of 3D scans of the full profile of the logs, obtained using laser scanners (Fig. 1).

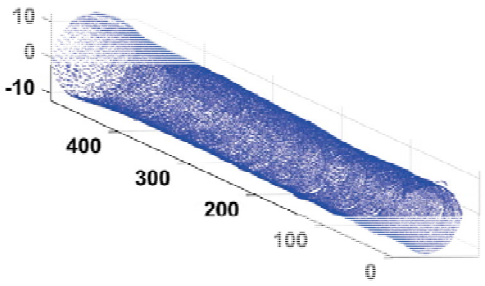


Fig. 1. Full profile 3D scan of a log. The scale is in centimeters.

These simulators can be used to support decision-making by alleviating the uncertainty associated with the sawing process, by predicting sets of lumber that might be obtained from every individual log [1]. In the following of this article, this set of lumber sawed from one log is called its basket of products (BoP).

However, the time taken by such simulation can be too long for practical use for decision problems involving thousands of logs. A single simulation can, indeed, take several minutes, or even more than one hour in some cases.

For this reason, researchers have proposed to replace these simulators with machine learning metamodels, i.e., surrogate models based on machine learning algorithms trained on past simulation results to predict BoP of new logs [2].

Few machine learning algorithms, however, allow making predictions based directly on 3D scans. These scans are, indeed, 3D points cloud, containing unordered points spanning the log surface. The number of points also varies from one scan to another.

Two main approaches have been proposed in past works to predict BoP of logs. The first one, introduced in [2], builds a structured representation of the logs based on a collection know-how features commonly used in the forest-product industry. The second approach, introduced in [3], is to compute a pairwise dissimilarity between 3D logs scans. These dissimilarities can then be used to predict BoP using, for example, a k nearest neighbors (kNN) algorithm.

Such a dissimilarity is a real-valued function $d(x_1, x_2)$, with x_1, x_2 two log scans, that intuitively measure how alike the two scans are. It is, in its usage, similar to a distance, but does not necessarily respect the properties of one, such as symmetry or positivity.

The dissimilarity used in previous works [3-5] to predict logs BoP is the so-called Iterative Closest Point (ICP) dissimilarity. This dissimilarity is a consequence of the Iterative Closest Point algorithm [6], which is classically used to align 3D point clouds. While this dissimilarity led to interesting experimental results, it has several undesirable properties. Most importantly, it is not symmetric, and its results may depend on the point cloud orientations at the algorithm's initialization. This fact motivates the study proposed in this paper, which considers an alternative dissimilarity based on an ensemble of shape functions (ESF) evaluated on the 3D scans.

The remaining of this paper is structured as follows. Section 2 first introduces both the ICP and ESF-based dissimilarities, as well as their advantages. Numerical experiments comparing the performances of these dissimilarities to predict logs baskets of products are presented Section 3. Section 4 concludes and proposes future research directions.

2. Dissimilarities Computation

2.1. ICP Dissimilarity

The ICP dissimilarity is a consequence of the Iterative Closest Point algorithm, which is an iterative algorithm for the fine registration of 3D shapes. This algorithm starts with two points clouds; one usually called the source and the other the target. It then searches for a rotation and a translation to minimize a position-dependent dissimilarity between the point clouds and align the source on the target.

The main steps of an iteration are as follows:

- Pair every point in the sources with its closest neighbor in the target. Points from the target may be selected several times or not at all. This step yield N_S pairs (s, t_s), with N_S the number of points in the source, s a point from the source and t_s its closest neighbor in the target.
- Find a rotation R and a translation T minimizing $D(R,T) = \sum_s (Rs + T - t_s)^2$. A closed-forms solution of this minimization problem can, in particular, be efficiently computed using quaternion theory [6].
- Apply the transformation obtained, go back to the first step, and loop until some ending criterium, such as a maximum number of iterations, is obtained.

It can be shown that the value of $D(R,T)$ decreases at every iteration of the algorithm. It, therefore, converges to a local minimum. The value of $D(R,T)$ obtained at the end of the last iteration of the algorithm is what is kept as the ICP dissimilarity d_{ICP}. Several inconveniences of this dissimilarity should, however, be noticed. Firstly, due to the non-symmetric roles of the target and source in the ICP algorithm, the ICP dissimilarity isn't symmetric, and in general $d_{ICP}(x_1, x_2)$ is not equal to $d_{ICP}(x_2, x_1)$. Secondly, the computation of d_{ICP} is highly dependent on the number of points in the source N_S and target N_T. More precisely, the complexity of one iteration of the ICP algorithm ranges from $O(N_S N_T)$ to $O(N_S \log(N_T))$ depending on the implementation of the closest neighbor search.

2.2. ESF Dissimilarity

The ESF dissimilarity, d_{ESF}, is based on the representation of the logs scan as a collection of q normalized histograms $h_1, \ldots, h_q$. These histograms have to be computed only once for every log, and can be stored and reused for multiple dissimilarities computation.

Every histogram h_j approximates the distributions of values taken by a shape function f_j evaluated over groups of points sampled at random from a scan. Various shape functions can be used. Three common functions were selected to be used in this paper. The first is the Euclidean distance between 2 points, the second is the angle defined by three points, and the third is the area of the triangle defined by three points. An example of these three histograms computed for one log is presented Fig. 2.

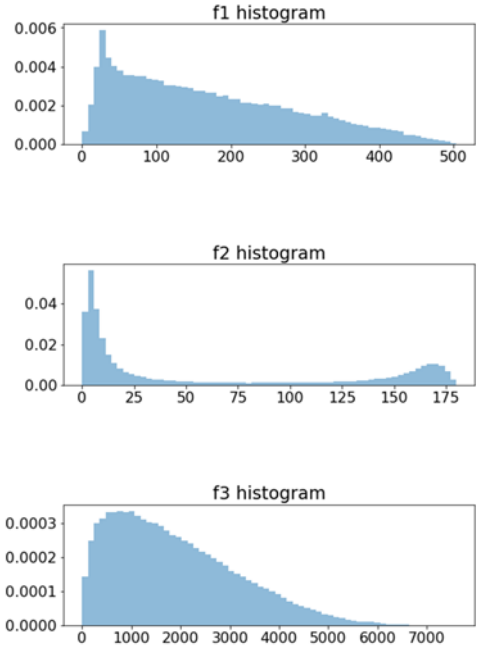


Fig. 2. Histograms of the three shapes functions of a log scan used to compute the ESF dissimilarities.

The ESF dissimilarity between two logs is then defined as the sum of the L_2 distance between the histograms of both logs:

$$d_{ESF}(x_1, x_2) = \sum_{j=1}^{q} \| \mathrm{h}_{1\mathrm{j}} - h_{2j} \|_2, \qquad (1)$$

with $\mathrm{h}_{1\mathrm{j}}$ and $\mathrm{h}_{2\mathrm{j}}$ the histograms of the first and second log respectively.

Interestingly, the complexity of computing d_{ESF} does not depend on the number of points in the scans. It is, instead, governed by two parameters selected by the user. The first one is the number of pair or triplet of points selected at random to estimate the histogram representation of the scan. This parameter was set to 2^{17},in this paper, by trial and error in order to stabilize the histogram estimates. The second is the number of bins in the histograms, fixed to 64 here.

3. Experiments

This section presents numerical experiments comparing the performances of kNN algorithms based

on either the ESF or ICP dissimilarity to predict BoP of logs. The dataset and evaluation scores are presented first, followed by the results.

3.1. Dataset

The dataset used for the experiments described in this paper originates from the Canadian sawmill industry. It contains information over 2219 real softwood logs. The 3D scan of every log is available, as well as their basket of products simulated by the sawing simulator Optitek [7].

The scans of logs are composed of rough ellipsoids spanning the log surface. All scans are i initially oriented around the z-axis, with the first ellipsoid starting at z = 0. This original orientation was kept as initial position when running the ICP algorithm to compute dissimilarities.

The sawmill modeled by the simulator could produce up to 74 types of lumber, characterized by their length, width, thickness, and grade (an evaluation of their quality). To simplify the prediction problem, the products were aggregated by grade, which reduce their number to 47, characterized only by their dimensions. A basket of products is, therefore, modeled as a vector of size 47. The ith component represents the number of lumbers of type I presents in the basket of products.

For experimental purpose, this dataset was repeatedly divided into a training set containing 1500 logs and a test set containing the remaining 719 logs. This dividing was repeated independently 30 times, the training set used as examples set by a kNN regressor algorithm, which is then evaluated on the test set.

Considering that the prediction problem is, here, modeled as a regression problem, the kNN prediction is the average of the baskets of the input log neighbors, and not necessarily a feasible basket. In particular, the kNN can predict non integer lumber quantities. This might not be a problem depending on the usage of the prediction, especially if, has done in [1], the predictions of individual logs are aggregated and used as input to a mix-integer programming problem.

3.2. Evaluation Scores

Several evaluation scores are used in this study to evaluate and compare the predictive performances of the kNN algorithms using either dissimilarity.

The first one is the usual root mean squared error between real and predicted baskets:

$$RMSE = \sqrt{\frac{1}{N}\sum_{l=1}^{N}\sum_{i=1}^{47}(\hat{y}_{li} - y_{li})^2}, \quad (2)$$

with N the size of the training set, $\hat{y}_{li}$ the predicted quantity of lumber of type i for the log l, and y_{li} the real quantity.

Several researchers, however, have stressed that such classic evaluation scores would be difficult to interpret for field experts from the industry and have proposed alternatives, in particular, the prediction-production score ($s^{pre\times pro}$) [2] and a variation of the F_1 score adapted to this problem [8].

In order to define the prediction-production scores, both the prediction score s^{pre} and production score s^{pro} need to be defined. Both are defined on a log-per-log basis.

The prediction score, s^{pre}, is the per-product average of the predicted lumber quantity over the real lumber quantity:

$$s_l^{pre} = \frac{1}{p}\sum_{i=1}^{p}\min(1, \frac{\hat{y}_{li}}{\max(y_{li}, \varepsilon)}), \quad (3)$$

where ε is, here, a very small quantity introduced to avoid dividing by 0. The index l in s_l^{pre} is added to stress the dependency over a specific log. Considering that this score is extremely sensible to $(0, 0)$ predicted-produced pairs which might make this score too optimistically biased due to the sparsity of the produced and predicted basket of product, all such pairs are removed before computing this score. p is the number of non filtered products, which can vary from one log to another.

The production score, s^{pro}, is similarly defined as the per-product average of the real lumber quantity over the predicted lumber quantity:

$$s_l^{pro} = \frac{1}{p}\sum_{i=1}^{p}\min(1, \frac{y_{li}}{\max(\hat{y}_{li}, \varepsilon)}) \quad (4)$$

The prediction-production score is then naturally defined as:

$$s^{pre\times pro} = \frac{1}{N}\sum_{l=1}^{N} s_l^{pre} \times s_l^{pro} \quad (5)$$

Similarly to the prediction-production scores, a variant of the F_1 score has been defined by [8] based on redefinitions of the numbers of True Positive (TP), False Positive (FP) and False Negative (FN) computed on a log per log basis:

- The number of true positives TP_l is the number of lumbers predicted and produced. $TP_l = \sum_{i=1}^{47}\min(\hat{y}_{li}, y_{li})$;
- The number of false positives FP_l is the number of lumbers predicted but not produced. $FP_l = \sum_{i=1}^{47}\max(\hat{y}_{li} - y_{li}, 0)$;
- The number of false negatives FN_l is the number of lumbers produced but not predicted. $FN_l = \sum_{i=1}^{47}\max(y_{li} - \hat{y}_{li}, 0)$.

The F_1 score is then redefined as:

$$s^{pre\times pro} = \frac{1}{N}\sum_{l=1}^{N}\frac{2\times TP_l}{2\times TP_l + FP_l + FN_l} \quad (6)$$

3.3. Results

As detailed previously, the dataset was divided 30 times into a training and a test sets. For every

dividing, two kNN algorithms searching neighbors with the ICP dissimilarity and ESF dissimilarity were trained on the train set and evaluated on the test set. In each case, the parameter k was tuned by 5 folds cross-validation on the training set, using the RMSE as basis for comparisons. k was selected among [1, 5, 10, 20]. All experiments were run on an Intel Core i7 vPRO 10th generation CPU at 2.70 GHz.

Experimental results are exposed Table 1. This table exposes the average and standard deviation over the repetitions of the experiments of the RMSE, prediction-production and F_1 scores for the kNN algorithms based on the ICP and on the ESF dissimilarities respectively.

The kNN based on the ESF dissimilarity has, in average, lower RMSE and higher prediction-production score and F_1 than the kNN based on the ICP dissimilarity. The poor performances of the ICP dissimilarity are, in part, due to the fact that it is not symmetric. In particular, replacing $d_{ICP}(x_1, x_2)$ by $d_{ICP}(x_1, x_2) + d_{ICP}(x_2, x_1)$ give a far lower RMSE, at 2.17. It, however, double the number of ICP computations needed to yield a prediction and is, therefore, not considered in the following.

To confirm this difference between ICP and ESF dissimilarities, scores of both methods were compared by using student statistical test with Nadeau and Bengio correction [9]. This correction aims to take into account the dependency between results obtained for various dividing of the same dataset into a training and test sets. Given two prediction method A and B yielding evaluations a_j and b_j over J independent dividing of a dataset, the statistic of this test is:

$$t = \frac{\sum_j a_j - b_j}{\sqrt{(\frac{1}{J} + \frac{n_2}{n_1})\hat{\sigma}^2}}, \quad (7)$$

with n_1 the size of the training set, n_2 the size of the test set and $\hat{\sigma}^2$ an estimate of the variance of the differences $a_j - b_j$. The p-value of the test is then computed from the usual student distribution with J-1 degrees of freedom.

This test yield, here, a p-value of 3×10^{-3} when applied to the RMSE, 0.6 when applied to the prediction-production scores, and 5×10^{-15} when applied to the F_1. Therefore, the difference can be considered statistically significant for two of the three evaluation scores.

Table 1. Average and standard deviation of the evaluation scores of kNN regressors based on the ICP and ESF dissimilarities, taken over the 30 dividing of the datasets. The best model for each score is highlighted in bold.

Dissimilarity	RMSE	$s^{pre \times pro}$	F_1
ICP	3.05 (0.34)	34.8 (7.1)	35.3 (1.2)
ESF	**2.18 (0.25)**	**39.3 (8.0)**	**50.2 (1.0)**

The ESF dissimilarity might also appear interesting in terms of computational cost. The scan of logs used during the experiments exposed in this paper contains, in average, 18452 points, and the average ICP dissimilarity computation time is 0.1 seconds, with 10 iterations of the algorithm. The implementation was based on Open3D library for python. The ESF dissimilarity was implement from scratch using the NumPy python library. The computation of the collection of histograms for each logs took, in average, 5.8 s per scan. These histograms, however, need to be computed only once for every log. In particular, when predicting the BoP of a log, only its own histograms need to be computed, because the others can be considered to have been computed and stored previously. Computing the ESF dissimilarity from pre-computed histograms is, then, extremely fast. In particular, it took only 0.0014 s in average for the implementation used during experiments. Whether the ICP or ESF dissimilarities would be faster in practice would then depends on the specific user implementation, ability to parallelize the ICP computations and size of the kNN algorithm example set. The ESF dissimilarity appears, however, preferable for large example sets.

4. Conclusion

This paper explores an alternative to the ICP dissimilarity to predict BoP of logs based on their 3D scans. More precisely, it proposes the use of the ESF dissimilarity, based on the computation of an intermediary representation of the scans as an ensemble of histograms. The computation of this representation can take several seconds, but only needs to be computed once for every log. Computation of the distance between the histograms is then far faster than computation of the ICP dissimilarity, which is advantageous for kNN algorithms with large example sets.

Additionally, when predicting BoP of logs with a kNN algorithm, the ESF dissimilarity leads to lower RMSE error and higher F_1 than the ICP dissimilarity.

Others machine learning algorithms, however, have been explored to predict BoP of logs from their 3D scans. These algorithms use a representation of a scan as a vector of dissimilarity toward a small set of preselected representative scans. Whether or not the ESF would still compare favorably for these algorithms needs to be explored in future works.

Acknowledgements

The authors gratefully acknowledge the financial support of the ANR-20-THIA-0010-01 Projet LOR-AI (lorraine intellgence artificielle) and région Grand EST.

We are also extremely grateful to FPInnovation who gathered and processed the dataset we are working with.

References

[1]. M. Morin, *et al.*, Machine learning-based models of sawmills for better wood allocation planning, *International Journal of Production Economics*, Vol. 222, 2020, 107508.

[2]. M. Morin, F. Paradis, A. Rolland, J. Wery, J. Gaudreault, F. Laviolette, Machine learning-based metamodels for sawing simulation, in *Proceedings of the Winter Simulation Conference (WSC'15)*, 2015, pp. 2160-2171.

[3]. C. Selma, H. B. El Haouzi, P. Thomas, J. Gaudreault, M. Morin, An Iterative Closest Point Method for Measuring the Level of Similarity of 3D Log Scans in Wood Industry, in *Proceedings of the 7th Workshop on Service Orientation in Holonic and Multi Agent Manufacturing (SOHOMA'17)*, 2017, pp. 433-444.

[4]. S. Chabanet, P. Thomas, H. Bril El-Haouzi, M. Morin, J. Gaudreault, A kNN approach based on ICP metrics for 3D scans matching: an application to the sawing process, *IFAC-PapersOnline*, Vol. 54, Issue 1, 2021, pp. 396-401.

[5]. S. Chabanet, V. Chazelle, P. Thomas, H. B. El-Haouzi, Dissimilarity to class medoids as features for 3D point cloud classification, in *Proceedings of the IFIP International Conference on Advances in Production Management Systems (APMS'21)*, 2021, pp. 573-581.

[6]. P. J. Besl, N. D. McKay, Method for registration of 3-D shapes, *Proceedings of SPIE*, Vol. 1611 1992, pp. 586-606.

[7]. P. Goulet, Optitek – User's Manual, 2006.

[8]. V. Martineau, M. Morin, J. Gaudreault, P. Thomas, H. B. El-Haouzi, Neural network architectures and feature extraction for lumber production prediction, in *Proceedings of the 34th Canadian Conference on Artificial Intelligence,* Vancouver, Canada, 25-28 May 2021.

[9]. R. R. Bouckaert, E. Frank, Evaluating the replicability of significance tests for comparing learning algorithms, in *Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD'04)*, 2004, pp. 3-12.

(8294)

Deep Feedforward Neural Network Classifier with Polynomial Layer and Shared Weights

Konstantinos Filippou, George Aifantis, Emmanouil Mavrikos and George E. Tsekouras
University of the Aegean, D*ept. of Cultural Technology and Communication*, Mytilene, 81100, Greece
E-mail: cti21003@ct.aegean.gr, cti21004@ct.aegean.gr, cti20006@ct.aegean.gr, gtsek@ct.aegean.gr

Summary: This paper introduces a novel deep feedforward neural network to perform classification tasks. The network encompasses several dense hidden layers followed by a polynomial layer, the nodes of which are inferred using activation functions expressed in terms of Hermite polynomials. The polynomial layer admits linear combinations of the outputs coming from the previous layers and expands them into several truncated Hermite series. The Hermite series are aggregated and fed to the output layer in order to infer the network's decision. The above strategy generates network with improved performance due to the following reasons. The Hermite polynomials are orthogonal over the whole set of real numbers and possess strong modeling and approximation capabilities. In addition, they can effectively deal with highly nonlinear data. Finally, the way the Hermite polynomials are implemented in this paper provides the advantage of using shared weights between nodes within the polynomial layer thus, reducing the number of the design parameters. The performance of the network is assessed by conducting comparative simulation experiments with four additional methods over several datasets. The comparison is carried out in terms of t-test inference statistics. The experimental outcomes indicate a clear superior performance of the proposed network when compared to the rest of the methods.

Keywords: Deep feedforward neural network, Hermite polynomials, shared weights, Classification.

1. Introduction

Deep feedforward neural networks (DFNNs) have been exercised as effective tools in dealing with a large variety of application frameworks [1-4]. Their simple structure provides the advantage of adopting a wide range of learning mechanisms. Truong et al [1] employed a multi-objective evolutionary computation algorithm to train a DFNN and use it to resolve problems related to the optimization of graded beams under static loads and free vibration. Cardoso et al [5] implemented dimensionality reduction mechanisms to study data related to diffuse lung diseases and designed a DFNN for improving the classification accuracy of such kind of data. Chen et al [6] expanded the connections of tree skeletons from hierarchical latent tree models in order to generate an unsupervised training mechanism for the effective development and application of typical DFNNs. In [7], a Gabor convolutional neural network able to perform feature extraction was concatenated with a DFNN that encompassed several fully connected layers to further improve the classification accuracy. In [8], a learning strategy for DFNNs was proposed, which was based on combining the tabu evolutionary computation scheme with a variant of the gradient descent optimization approach.

To enhance the ability of a neural network in effectively dealing with highly nonlinear data, many scholars have used polynomials activation functions, yielding the so called polynomial neural networks (PNNs) [9-11]. Inserting polynomial activation functions into the network's design process constitutes an effective way to approximate the input-output relationships because those functions possess high-order representation capabilities [12-15]. So far, many different types of polynomials have been embedded in the design process of neural networks such as, Hermite polynomials [9, 10], Legendre polynomials [11, 12], and Chebyshev polynomials [13, 14]. PNNs have been considered as trustworthy tools for regression and classification problems, spanning over a wide range of applications related to coastal engineering [10], differential equations [11, 13], energy saving [12], pattern classification [14], medical applications [15, 16], etc.

In [9], a feedforward PNN with one hidden layer was designed, which admitted linear combinations of the input variables and inferred the output in terms of Hermite polynomials activation functions. In [12], the aforementioned linear combinations were normalized within the interval $[-1,1]$ and fed into Legendre polynomial activation functions. An implementation problem related to the above methods is that each linear combination enters one distinct polynomial thus yielding non-exact series expansions of each linear combination thus, compromising the network's efficiency. In [15, 16] a stacked network structures were used to design deep polynomial networks for medical image processing. The difficulty associated with that strategy is that as the number of layers increases the polynomial orders of the input variables also increase. This might result in high computational complexity, while it needs careful design to avoid overfitting problems. Misra et al [17] developed PNNs in terms of the well-known Kolmogorov-Gabor polynomials. The same polynomials were employed in [18, 19] to develop neuro-fuzzy networks. However, the use of such kind of polynomials acts to significantly increase the number of the synaptic weights thus, compromising the computational complexity of the resulting network.

In this paper, a polynomial DFNN is developed, which encompasses several fully connected layers concatenated with a polynomial layer that includes activations functions represented by Hermite polynomials. Hermite polynomials are orthogonal, and they been acknowledged as effective framework to cope with highly nonlinear data [20]. Moreover, they possess certain advantages over other orthogonal polynomials such as the Legendre or Chebyshev polynomials. For example, the Hermite polynomials are orthogonal over the whole set of real numbers, whereas the Chebyshev and Legendre are orthogonal in the interval $[-1,1]$ [20]. Thus, the usage of the latter in designing neural networks requires a normalization process that results in constrained optimization mechanisms [12]. To alleviate the problems discussed in the previous paragraph, the structure of the proposed network follows two basic design principles. First, contrary to the methods in [9, 12], the polynomial layer generates exact truncated Hermite polynomial series for each one of its inputs. Second, each input entering the layer is connected with several polynomial nodes using shared synaptic weights. Thus, contrary to the approaches in [17-19], the resulting number of design parameters lies within acceptable levels. Finally, comparative experimental analysis based of inference statistics verifies the improved performance of the proposed network.

The rest of the paper is synthesized as follow. Section 2 provides the detailed theoretical analysis of the proposed network. The experimental study is described in Section 3. Finally, the paper concludes in Section 4.

2. The Proposed Network

The proposed deep feedforward neural network consists of several fully connected (i.e., dense) layers followed by a polynomial layer that generates a number of truncated Hermite polynomial series, which are fed into the output layer. In what follows, we first describe some properties of the Hermite polynomials providing the nomenclature used in this paper, also. Then, the detailed presentation of the proposed network takes place along with the relative analysis.

The Hermite polynomials are orthogonal satisfying the subsequent differential equation [20],

$$H_n(x) = (-1)^n e^{x^2} \frac{d\left(e^{-x^2}\right)}{dx}, \quad (1)$$

where n defines the polynomial order and $x \in R$. The orthogonality of Hermite polynomials is sustained over the set $(-\infty, +\infty)$having as weight the quantity e^{-x^2}

$$\langle H_n(x), H_m(x)\rangle = \begin{cases} 2^n n! \sqrt{\pi}, if n = m \\ 0, otherwise \end{cases}, \quad (2)$$

where $\langle .,. \rangle$ stands for the inner product operation.

The zero and first order polynomials are equal to $H_0(x) = 1$ and $H_1(x) = 2x$, respectively. As such, any Hermite polynomial of order $n \geq 2$ is calculated in terms of the next recurrence relation,

$$H_n(x) = 2xH_{n-1}(x) - 2(n-1)H_{n-2}(x) \quad (3)$$

It can be easily shown that $P_n(x)$ can be described as [10]

$$H_n(x) = \sum_{k=0}^{n} \phi_{kn} x^k, \quad (4)$$

with

$$\phi_{kn} = \begin{cases} 2\phi_{n-1,k-1} - (k+1)\phi_{n-1,k+1}, if k > 0 \\ -\phi_{n-1,k+1}, if k = 0 \end{cases}, \quad (5)$$

where $\phi_{00} = 1, \phi_{10} = 0$, and $\phi_{11} = 2$.

The overall structure of the proposed network is illustrated in Fig. 1. Specifically, it consists of the input layer, L dense hidden layers and one polynomial layer that feeds the output layer. The detailed functionality of the polynomial layers and its connection to the L-th layer and the output layer is depicted in Fig. 2. The input layer admits p variables, which come in the form of the vector $\boldsymbol{x} = [x_1, x_2, \ldots, x_p]^T \in R^p$. The first layer includes c_1 nodes.

The corresponding bias vector is defined as

$$\boldsymbol{b}^{(1)} = \left[b_1^{(1)} \quad b_2^{(1)} \quad \ldots \quad b_{c_1}^{(1)}\right]^T, \quad (6)$$

and the synaptic weight vectors as

$$\boldsymbol{w}_{i_1}^{(1)} = \left[w_{i_1,1}^{(1)}, w_{i_1,2}^{(1)}, \ldots, w_{i_1,p}^{(1)}\right]^T, \quad (7)$$

with i_1=1, 2, …, c_1. The layer's output $\boldsymbol{g}^{(1)}: R^p \rightarrow R^{c_1}$comes in the form

$$\boldsymbol{g}^{(1)}(\boldsymbol{x}) = \left[g_1^{(1)}(\boldsymbol{x}), g_2^{(1)}(\boldsymbol{x}), \ldots, g_{c_1}^{(1)}(\boldsymbol{x})\right]^T, \quad (8)$$

with $g_{i_1}^{(1)}: R^p \rightarrow R$

$$g_{i_1}^{(1)}(\boldsymbol{x}) = g_{i_1}^{(1)}\left(\boldsymbol{w}_{i_1}^{(1)T}\boldsymbol{x} + b_{i_1}^{(1)}\right), \quad (9)$$

The layer ℓ ($\ell = 1,2,\ldots,L$) encompasses c_ℓ hidden nodes. The bias and the synaptic weights vectors associated with that layer are defined as

$$\boldsymbol{b}^{(\ell)} = [b_1^{(\ell)}, b_2^{(\ell)}, \ldots, b_{c_\ell}^{(\ell)}]^T, \quad (10)$$

and

$$\boldsymbol{w}_{i_\ell}^{(\ell)} = \left[w_{i_\ell,1}^{(\ell)}, w_{i_\ell,2}^{(\ell)}, \ldots, w_{i_\ell,c_{\ell-1}}^{(\ell)}\right]^T, \quad (11)$$

with and i_ℓ=1, 2, …, c_ℓ, thus resulting in the following weight matrix

$$\boldsymbol{W}^{(\ell)} = \left[\boldsymbol{w}_1^{(\ell)T}, \boldsymbol{w}_2^{(\ell)T}, \ldots, \boldsymbol{w}_{c_\ell}^{(\ell)T}\right]^T, \quad (12)$$

The output of the layer $\boldsymbol{g}^{(\ell)}: R^{c_{\ell-1}} \rightarrow R^{c_\ell}$is described by the next vector-based form

$$\boldsymbol{g}^{(\ell)}(\boldsymbol{x}) = \left[g_1^{(\ell)}(\boldsymbol{x}), g_2^{(\ell)}(\boldsymbol{x}) \ldots, g_{c_\ell}^{(\ell)}(\boldsymbol{x})\right]^T, \quad (13)$$

In (13) each element is defined as

$$g_{i_\ell}^{(\ell)}(\boldsymbol{x}) = g_{i_\ell}^{(\ell)}\left(\boldsymbol{w}_{i_\ell}^{(\ell)T}\boldsymbol{g}^{(\ell-1)}(\boldsymbol{x}) + b_{i_\ell}^{(\ell)}\right), \quad (14)$$

with $g_{i_\ell}^{(\ell)}: R^{c_{\ell-1}} \rightarrow R$.

The interaction between the layer L and the output layer is defined in terms of the polynomial layer P, the basic structure of which is depicted in Fig. 2. The main purpose of using the polynomial layer is to generate truncated Hermite polynomial series of linear combinations of the L-th layer's output vector $\boldsymbol{g}^{(L)}(\boldsymbol{x})$. To accomplish that task, the polynomial layer encompasses d groups of nodes, each of which includes n+1 nodes, where the activation function of the first node is the Hermite polynomial of 0 order, the activation function of the second node is the Hermite polynomial of order 1, and finally the activation function of the n+1-th node is the Hermite polynomial of order n of the aforementioned linear combinations. Thus, in total, the layer contains $d(n+1)$ polynomial nodes, defined as follows,

$$G_1^{(P)} = \{g_1^{(P)}(\boldsymbol{x}), \ldots, g_{n+1}^{(P)}(\boldsymbol{x})\} \quad (15)$$

$$\ldots$$

$$G_{i_P}^{(P)} = \{g_{(i_P-1)(n+1)+1}^{(P)}(\boldsymbol{x}), \ldots, g_{i_P(n+1)}^{(P)}(\boldsymbol{x})\} \quad (16)$$

$$\ldots$$

$$G_d^{(P)} = \{g_{(d-1)(n+1)+1}^{(P)}(\boldsymbol{x}), \ldots, g_{d(n+1)}^{(P)}(\boldsymbol{x})\} \quad (17)$$

The bias vector and the synaptic weight matrix associated with the polynomial layer are written as

$$\boldsymbol{b}^{(P)} = \left[b_1^{(P)} \quad b_2^{(P)} \quad \ldots \quad b_d^{(P)}\right]^T, \quad (18)$$

$$\boldsymbol{W}^{(P)} = \left[\boldsymbol{w}_1^{(P)T}, \boldsymbol{w}_2^{(P)T}, \ldots, \boldsymbol{w}_d^{(P)T}\right]^T, \quad (19)$$

where:

$$\boldsymbol{w}_{i_P}^{(P)} = \left[w_{i_P,1}^{(P)}, w_{i_P,2}^{(P)}, \ldots, w_{i_P,c_L}^{(P)}\right]^T \quad (i_P=1, 2, \ldots, d) \quad (20)$$

As such, the above-mentioned linear combinations of the L-th layer's outputs are described as indicated next

$$z_{i_P}^{(P)}(\boldsymbol{x}) = \boldsymbol{w}_{i_P}^{(P)T}\boldsymbol{g}^{L}(\boldsymbol{x}) + b_{i_P}^{(P)}, \quad (21)$$

with $i_P = 1,2,\ldots,d$, and the activation functions for the i_P-th group of nodes, given in equations (15)-(17), are as follows

$$g_t^{(P)}(\boldsymbol{x}) = \begin{cases} H_0(z_{i_P}^{(P)}(\boldsymbol{x})), ift = (i_P-1)(n+1)+1 \\ H_1(z_{i_P}^{(P)}(\boldsymbol{x})), ift = (i_P-1)(n+1)+2 \\ \ldots \\ H_n(z_{i_P}^{(P)}(\boldsymbol{x})), ift = i_P(n+1) \end{cases} \quad (22)$$

where $H_\nu(z_{i_P}^{(p)}(\boldsymbol{x}))$, for $\nu = 0,1,\ldots,n$, are calculated in equations (4) and (5).

A basic property of the polynomial layer is that the synaptic weights are shared between the polynomial nodes that belong to the same group. Indeed, the linear combination in Eq. (21) can be rewritten as

$$z_{i_P}^{(P)}(\boldsymbol{x}) = \sum_{i_L=1}^{c_L} w_{i_P,i_L}^{(P)} g_{i_L}^{(L)}(\boldsymbol{x}) + b_{i_P}^{(P)} \quad (23)$$

Thus, the synaptic weight associated with the activation function $g_{i_L}^{(L)}(\boldsymbol{x})$ $(i_L = 1,2,\ldots,c_L)$ and the i_p-th group of polynomials, given in (22), is the $w_{i_p,i_L}^{(P)}$. Therefore, all polynomial nodes, belonging to the group i_p, share the weight $w_{i_p,i_L}^{(P)}$ as far as the $g_{i_L}^{(L)}(\boldsymbol{x})$ is concerned. The same holds for the corresponding bias $b_{i_P}^{(P)}$. The above analysis is clearly illustrated in Fig. 2.

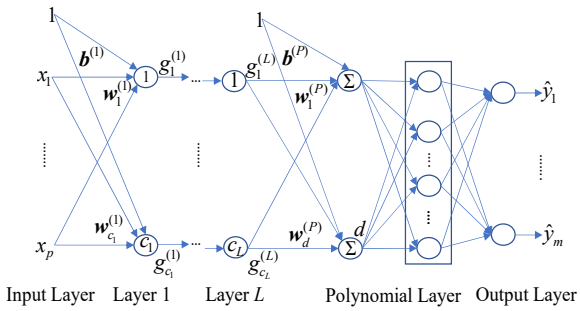


Fig. 1. The structure of the proposed network.

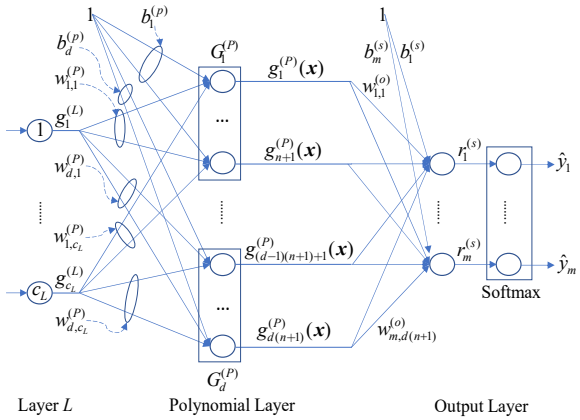


Fig. 2. The structure of the polynomial and output layers (the elliptical curves contain the shared weights).

To this end, the output of the polynomial layer comes in the following vector form

$$\boldsymbol{g}^{(P)}(\boldsymbol{x}) = \left[g_1^{(P)}(\boldsymbol{x}), g_2^{(P)}(\boldsymbol{x}) \ldots, g_{d(n+1)}^{(P)}(\boldsymbol{x})\right]^T \quad (24)$$

The output layer calculates m outputs $\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_m$. The corresponding synaptic weight matrix and the bias vector are

$$\boldsymbol{W}^{(s)} = \left[\boldsymbol{w}_1^{(s)T}, \boldsymbol{w}_2^{(s)T}, \ldots, \boldsymbol{w}_m^{(s)T}\right]^T, \quad (25)$$

$$\boldsymbol{b}^{(s)} = \left[b_1^{(s)} \quad b_2^{(s)} \quad \ldots \quad b_m^{(s)}\right]^T \quad (26)$$

where $\mu = 1,2,\ldots,m$, and each weight vector in (29) is

$$\boldsymbol{w}_\mu^{(s)} = \left[w_{\mu,1}^{(P)}, w_{\mu,2}^{(P)}, \ldots, w_{\mu,d(n+1)}^{(P)}\right]^T \quad (27)$$

The output $\hat{y}_\mu$, with $\mu = 1,2,\ldots,m$, is calculated using the softmax activation function

$$\hat{y}_\mu = \frac{e^{r_\mu^{(s)}}}{\sum_{k=1}^{m} e^{r_k^{(s)}}}, \quad (28)$$

where

$$r_\mu^{(s)} = \boldsymbol{w}_\mu^{(s)T} \boldsymbol{g}^{(P)}(\boldsymbol{x}) + b_\mu^{(s)} \quad (29)$$

Fig. 2 depicts the above analysis, and the way the network's outputs are calculated.

Next, we show that the quantity $r_\mu^{(s)}$ can be written as the sum of d truncated Hermite polynomial series of order n.

From (29) we obtain

$$r_\mu^{(s)} = \sum_{t=1}^{d(n+1)} w_{\mu,t}^{(s)} g_t^{(P)}(\boldsymbol{x}) + b_\mu^{(s)}, \quad (30)$$

which is modified as,

$$r_\mu^{(s)} = \sum_{t=1}^{n+1} w_{\mu,t}^{(s)} g_t^{(P)}(\boldsymbol{x}) + \cdots + + \sum_{t=(d-1)(n+1)+1}^{d(n+1)} w_{\mu,t}^{(s)} g_t^{(P)}(\boldsymbol{x}) + b_\mu^{(s)} \quad (31)$$

Thus

$$r_\mu^{(s)} = \sum_{i_P=1}^{d} \sum_{t=(i_P-1)(n+1)+1}^{i_P(n+1)} w_{\mu,t}^{(s)} g_t^{(P)}(\boldsymbol{x}) + b_\mu^{(s)} \quad (32)$$

For a specific value of the index i in Eq. (22), there is a one-to-one correspondence between the index

$$t = (i_P - 1)(n+1) + 1, (i_P - 1)(n+1) + +2 \ldots, i_P(n+1), \quad (33)$$

and the index $\nu = 0,1,2,\ldots,n$.

In addition, for a specific i_P, the subsequent substitution takes place: $a_{\mu,\nu}^{(i)} = w_{\mu,t}^{(s)}$.Taking into account the activation functions $g_t^{(P)}(\boldsymbol{x})$ in Eq. (22), the Eq. (32) is modified as follows

$$r_\mu^{(s)} = \sum_{i_P=1}^{d} \sum_{\nu=0}^{n} a_{\mu,\nu}^{(i)} H_\nu(z_{i_P}^{(P)}(\boldsymbol{x})) + b_\mu^{(s)} \quad (34)$$

Since the polynomial $H_0(z_{i_P}^{(P)}(\boldsymbol{x}))$ is equal to 1, the $b_\mu^{(s)}$can be absorbed by the weight of one of these polynomials. Finally, we can easily notice that the $r_\mu^{(s)}$is represented as the sum of d truncated Hermite polynomial series of order n.

3. Simulation Experiments Results

In this section, the proposed network is compared with four competitive methods. The first one is a deep feedforward neural network (DFNN), while the other three methods are taken from the WEKA software [21] and they are the well-known J48, the Decision Stump (DS), and the Random Tree (RT) algorithms. In all simulation experiments, the DFNN network utilizes 5 dense layers with 12 ReLu nodes each, while the output layer uses the "SoftMax" activation function.

The optimization task is conducted in terms of the ADAM optimizer [22], while the objective function is the categorical cross-entropy function. The performance index is the Classification Accuracy (CA) normalized in the interval [0,1]. Finally, five data sets are used to conduct the comparative analysis. The data sets were taken from the UCI Machine Learning Repository [23] and their basic properties are provided in Table 1.

Table 1. Properties of the Tested Data Sets.

# Data Set	No of Instances	No of Inputs	No of Classes
1: Abalone	4177	8	3
2: Cardiotocography	2126	21	3
3: Statlog German Credit	1000	24	2
4: Red Wine Quality	1599	11	6
5: Wisconsin Diagnostic Breast Cancer	569	30	2

For each data set we used 5-fold cross-validation analysis. Regarding each fold the proposed network run 10 times with different initializations. Thus, for each data set and each method, 50 samples were generated.

Tables 2 and 3 summarize the mean values and the corresponding standard deviations of the Classification Accuracy (CA) index, obtained by the five competitive algorithms, regarding all tested data sets. The following remarks can be easily derived. First, in all tested data sets the proposed network clearly outperforms the other methodologies. Second, the case where truncated Hermite polynomial series of order n = 2 is used, appear to achieve the best performance in two out of five data sets, while the use of truncated series of order n=3 provides the best results in the remaining three data sets. Third, Table 3 directly

indicates that the DFNN and the J48 are the dominant among the four competitive methods.

Table 2. Classification accuracy index and the respective standard deviations obtained by the proposed network for the five datasets.

#	Proposed (n=2)	Proposed (n=3)	Proposed (n=4)
1	0.5629±0.02	0.555±0.01	0.551±0.02
2	0.899±0.02	0.910±0.03	0.879±0.03
3	0.720±0.03	0.755±0.02	0.733±0.03
4	0.586±0.03	0.583±0.03	0.577±0.04
5	0.942±0.02	0.952±0.02	0.880±0.10

Table 3. Classification accuracy index and the respective standard deviations obtained by the four competitive methods for the five datasets.

#	DFNN	J48	DS	RT
1	0.550±0.01	0.527±0.01	0.529±0.01	0.495±0.01
2	0.881±0.01	0.894±0.16	0.783±0.01	0.859±0.19
3	0.719±0.03	0.729±0.02	0.729±0.02	0.677±0.03
4	0.559±0.04	0.546±0.16	0.490±0.16	0.556±0.16
5	0.942±0.02	0.917±0.11	0.892±0.02	0.930±0.02

To further scrutinize the above results, the five methods are compared in terms of inference statistics. As far as the proposed network is concerned the case of truncated Hermite polynomial series of order n=3 is considered. Descriptive statistics were evaluated regarding the samples obtained for each data set and each method. The resulting box-plots are illustrated in Figs. 3 and 4. Next, we applied the Kolmogorov-Smirnov test, which showed that the data were normally distributed. Thus, to assess the hypothesis that there is a statistically significant difference between the proposed method and the other methods we applied paired t-tests for each dataset individually.

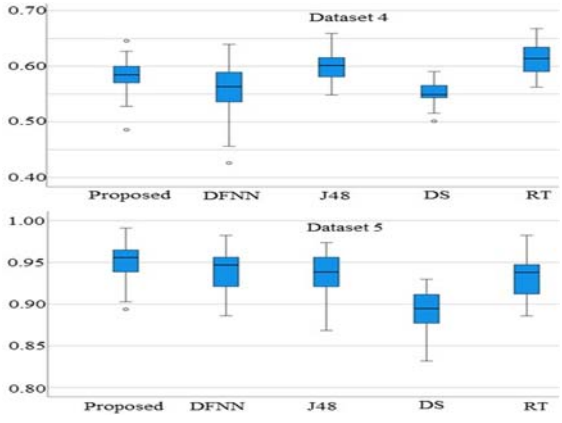


Fig. 3. Box plots of the five methods (proposed n=3) with respect to Dataset 4 and Dataset 5.

The null hypothesis assumes that the true mean difference is equal to zero, while the upper-tailed alternative hypothesis assumes that the difference is greater than zero. Table 4 depicts the 95% confidence intervals of paired differences and the corresponding p-values. Regarding the Dataset 1, the null hypothesis is rejected in favour of the alternative hypothesis for each pair. The proposed method outperforms J48, Decision Stump and Random Tree at significance less than 0.001 and DFNN at significance 0.04.

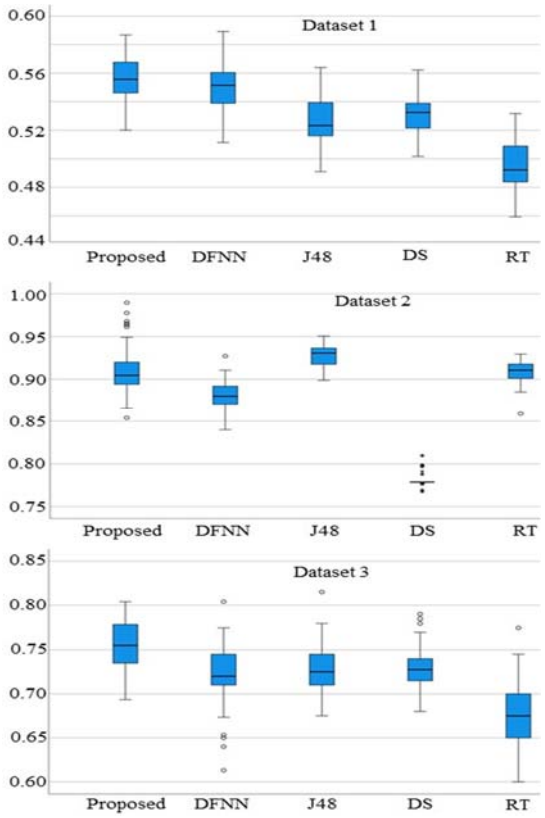


Fig. 4. Box plots of the five methods (proposed n=3) with respect to Dataset 1, Dataset 2 and Dataset 3.

In the Dataset 2, our method has a statistically significant difference against DFNN and Decision Stump, where the null hypothesis is rejected at significance less than 0.001. Meanwhile, J48 outperforms the proposed method at significance less than 0.001 and there was no statistically significant difference between the proposed network and the Random Tree, thus we cannot reject the null hypothesis.

Regarding the Dataset 3, the null hypothesis is rejected in every case at significance less than 0.001 directly indicating the superiority of our proposed network.

For the Dataset 4 the null hypothesis is rejected in every case. In addition, the proposed network performs better than DFNN and Decision Stump (significance less than 0.001), while it achieves inferior performance when compared to the J48 and Random Tree algorithms.

Finally, in the Dataset 5, the null hypothesis is rejected, and the proposed method seems to be superior than the other methods at significance level less than 0.001.

Table 4. Comparative results obtained by the t-test inference statistics.

Pair		95% C. I of Difference		Significance	
	Mean	Lower	Upper	t	p-value
Dataset 1					
Proposed-DFNN	0.0052	-0.001	0.011	1.9	0.040
Proposed-J48	0.0278	0.022	0.336	9.6	$<10^{-3}$
Proposed-DS	0.0255	0.199	0.310	9.3	$<10^{-3}$
Proposed-RT	0.0623	0.0533	0.67	17.9	$<10^{-3}$
Dataset 2					
Proposed-DFNN	0.2960	0.0199	0.0393	6.1	$<10^{-3}$
Proposed-J48	-0.017	-0.0260	-0.0080	-3.8	$<10^{-3}$
Proposed-DS	0.1284	0.1198	0.1369	30	$<10^{-3}$
Proposed-RT	0.0013	-0.0083	0.0109	0.28	$<10^{-3}$
Dataset 3					
Proposed-DFNN	0.0363	0.0243	0.0484	6.1	$<10^{-3}$
Proposed-J48	0.0264	0.0171	0.0356	5.7	$<10^{-3}$
Proposed-DS	0.0255	0.0157	0.0352	5.3	$<10^{-3}$
Proposed-RT	0.0779	0.0661	0.0896	13.3	$<10^{-3}$
Dataset 4					
Proposed-DFNN	0.0237	0.0082	0.0392	3.1	0.002
Proposed-J48	-0.0166	-0.0271	-0.0061	-3.2	0.001
Proposed-DS	0.0323	0.0223	0.0423	6.5	$<10^{-3}$
Proposed-RT	-0.0289	-0.0402	-0.0177	-5.2	$<10^{-3}$
Dataset 5					
Proposed-DFNN	0.0107	0.0039	0.0175	3.2	0.001
Proposed-J48	0.0179	0.0094	0.0264	4.2	$<10^{-3}$
Proposed-DS	0.0605	0.0514	0.0695	13.4	$<10^{-3}$
Proposed-RT	0.0223	0.0134	0.0312	5.03	$<10^{-3}$

To summarize, the above results clearly indicate that the use of the polynomial layer significantly improved the performance of the DFNN, while maintaining a superior behavior when compared to the rest of the tested methods.

4. Conclusions

In this paper, a novel polynomial deep feedforward neural network classifier has been developed. The task was to investigate the potential influence of Hermite polynomial activation functions to the overall classification accuracy. The network consists of several fully connected layers followed by a polynomial layer with a specialized structure based on Hermite polynomials. The very core of the functionality of the polynomial is to admit linear combinations of the outputs coming from the previous dense layer and expand each one of them into truncated Hermite polynomial series. Then, each network's output is realized as the sum of the above-mentioned series. The existence of Hermite polynomial truncated series in the network's inference enhances the corresponding classification performance because it embeds into the design process all the advantages provided by those series such as the strong modelling capabilities, the simplicity, the universal approximation property, etc. In addition, the connection between the polynomial layer and the previous dense layer is realized in terms of shared synaptic weights, which acts to produce a convenient training procedure. The network was compared with four other methods in terms of statistical analysis on the basis of several data sets. The simulation results support the above analysis since the network appeared to achieve superior performance.

5. References

[1]. T. T. Truong, J. Lee, T. Nguyen-Thoi, Multi-objective optimization of multi-directional functionally graded beams using an effective deep feedforward neural network-SMPSO algorithm, *Structural and Multidisciplinary Optimization*, Vol. 63, 2021, pp. 2889-2918.

[2]. A. E. Korchi, Y. Ghanou, Backpropagation issues with deep feedforward neural networks, in Innovations in Smart Cities and Applications (M. B. Ahmed, A. Boudhir, Eds.), *Springer*, 2017, pp. 335-343.

[3]. G. Masetti, F. Di Giandomenico, Analyzing forward robustness of feedforward deep neural networks with leaky ReLU activation function through symbolic propagation, *Communications in Computer and Information Science*, Vol. 1323, 2020, pp. 460-474.

[4]. L. Zhang, T. Luo, F. Zhang, Y. Wu, A recommendation model based on deep neural network, *IEEE Access*, Vol. 6, 2018, pp. 9454-9463.

[5]. I. Cardoso, E. Almeida, H. Allende-Cid, A. C. Frery, R. M. Rangayyan, P. M. Azevedo-Marques, H. S. Ramos, Evaluation of deep feedforward neural networks for classification of diffuse lung diseases, *Lecture Notes in Computer Science*, Vol. 10657, 2018, pp. 152-159.

[6]. Z. Chen, X. Li, Z. Tian, N. L. Zhang, Fast structure learning for deep feedforward networks via tree skeleton expansion, *Lecture Notes in Artificial Intelligence*, Vol. 11726, 2019, pp. 277-289.

[7]. M. M. Lau, J. T. S. Phang, K. H. Lim, Convolutional deep feedforward network for image classification, in *Proceedings of the 7th International Conference on Smart Computing & Communications (ICSCC'19)*, 2019, pp. 1-4.

[8]. T. K. Gupta, K. Raza, Optimizing deep feedforward neural network architecture: A tabu search based approach, *Neural Processing Letters*, Vol. 51, 2020, pp. 2855-2870.

[9]. L. Ma, K. Khorasani, Constructive feedforward neural networks using Hermite polynomial activation functions, *IEEE Transactions on Neural Networks*, Vol. 16, Issue 4, 2005, pp. 821-833.

[10]. G. E. Tsekouras, V. Trygonis, A. Maniatopoulos, A. Rigos, A. Chatzipavli, J. Tsimikas, N. Mitianoudis, A. F. Velegrakis, A Hermite neural network incorporating artificial bee colony optimization to model shoreline realignment at a reef-fronted beach, *Neurocomputing*, Vol. 280, 2018, pp. 32-45.

[11]. S. Mall, S. Chakraverty, Application of Legendre neural network for solving ordinary differential equations, *Applied Soft Computing*, Vol. 43, 2016, pp. 347-356.

[12]. A. Rigos, G. E. Tsekouras, A. Chatzipavlis, A. F. Velegrakis, Modeling beach rotation using a novel Legendre polynomial feedforward neural network trained by nonlinear constrained optimization, in *Proceedings of the Twelfth IFIP International Conference on Artificial Intelligence Applications and Innovations*, Thessaloniki, Greece, 2016, pp. 167-179.

[13]. Z. Hajimohammadi, F. Baharifard, A. Ghodsi, K. Parand, Fractional Chebyshev deep neural network (FCDNN) for solving differential models, *Chaos, Solitons and Fractals*, Vol. 153, 2021, 111530.

[14]. Y. Zhang, Y. Yin, D. Guo, X. Yu, L. Xiao, Cross-validation based weights and structure determination of Chebyshev-polynomial neural networks for pattern classification, *Pattern Recognition*, Vol. 47, 2014, pp. 3414-3428.

[15]. J. Shi, S. Zhou, X. Liu, Q. Zhang, M. Lu, T. Wang, Stacked deep polynomial network-based representation learning for tumour classification with small ultra sound image dataset, *Neurocomputing*, Vol. 194, 2016, pp. 87-94.

[16]. J. Shi, X. Zheng, Y. Li, Q. Zhang, S. Ying, Multimodal neuroimaging feature learning with multimodal stacked deep polynomial networks for diagnosis of Alzheimer's disease, *IEEE Journal of Biomedical and Health Informatics*, Vol. 22, Issue 1, 2018, pp. 173-183.

[17]. B. B. Misra, S. C. Satapathy, B. N. Biswal, P. K. Dash, G. Panda, Pattern classification using polynomial neural network, in *Proceedings of the IEEE Conference on Cybernetics and Intelligent Systems (ICCIS'16)*, 2006, pp. 1-6.

[18]. W. Huang, S. K. Oh, W. Pedrycz, Fuzzy reinforced polynomial neural networks constructed with the aid of PNN architecture and fuzzy hybrid predictor based on nonlinear function, *Neurocomputing*, Vol. 458, 2021, pp. 454-467.

[19]. C. Zhang, S. K. Oh, Z. Fu, Hierarchical polynomial-based fuzzy neural networks driven with the aid of hybrid network architecture and ranking-based neuron selection strategies, *Applied Soft Computing*, Vol. 113, 2021, 107865.

[20]. G. E. Andrews, R. Askey, R. Roy, Special Functions, *Cambridge University Press*, UK, 2000.

[21]. E. Frank, M. A. Hall, I. H. Witten, The WEKA Workbench, Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", 4th Ed., *Morgan Kaufmann*, 2016.

[22]. D. P. Kingma, J. Ba, Adam: A method for stochastic optimization*, arXiv Preprint*, 2014, arXiv:1412.6980.

[23]. UC Irvine Machine Learning Repository, http://archive.ics.uci.edu/ml/index.php

(8410)

Sparse and Irregularly Distributed Geo-data Categorization for Predictive Regression Model Training and Inference – Strategies and Results

Jérémie Farret [1], Bilel Kchouk [2], Roghayeh Soleymani [1], Julien Beaulieu [1], Vanessa Farias [1] and Ibtihel Amara [1]
[1] Mind in a Box Inc. 3575 St Laurent Boulevard, Suite 200 Montreal, Quebec H2X 2T6 Canada
[2] Bell Business Markets
Tel.: +1-833-636-2269
E-mail: jeremie@mindinabox.ai

Summary: Scarce or irregularly distributed geo-data presents specific challenges in order to be used for machine learning model training and inferences. Specific geo-categorization strategies are introduced which enable optimizing the segmentation of the data and corresponding data processing computation, in order to prepare the data for the machine learning model consumption. Through a customer use case, conclusive results are presented which demonstrate the efficiency of some of the geo-categorization strategies introduced in this article.

Keywords: Machine learning, Regression, Geo-data, Geo-categorization, Clustering, Geo-AI.

1. Introduction

Regardless of its source (geo-tracking, IoT, Enterprise Resources Management systems, geo-tracking systems, public/open data, etc.), geo-data often presents irregular distribution and scarcity. Artificial Intelligence techniques heavily rely on data tagging and categorization, and geolocation can be instrumental for categorization of source data and supporting model training purposes. Important variations of the data distribution over time can, in fact, produce unstable results and unpredictable performance for the machine learning models. In a typical customer situation involving machine learning model training, several sources of geo-data are ingested and analyzed by an integrated big data analytics architecture. Geo-categorization of this data is optimized for machine learning models by introducing various strategies, resulting in machine learning model performance increases.

These strategies have been identified based on the following essential requirements:

- To support robust automated qualification methods, since semi-automated or manual evaluation of tridimensional (geolocation + attribute) geo-data is difficult to ensure through visual analysis of 2D representations.
- To enable high performance categorization of historical geo-data for machine learning model training and real time categorization of production data for the model inference.
- To ensure stability over time of geo-categories produced for the sparse datasets, in order to facilitate trend analysis and machine learning model performance continuity.

Three strategies will be presented, each addressing in a specific way the issue of addressing such constraints on sparse geo-data.

2. Geo-data Use Case and Corresponding Geo-categorization Strategies

2.1. Source Geo-data

Source data contains construction permits duration in Ontario and Quebec, both from customer operations and provincial Open Data (for data validation and enrichment). The Open Data, while irregular in its distribution, showed a satisfying volumetry and geo-categorization potential. The operational data revealed to be both scarce outside of major urban areas and adverse to typical geospatial specific categorization approaches because of its distribution in some operational zones of interest.

2.2. Geo-categorization Strategies

Various geo-categorization strategies can be applied on raw location data some based on statistical characteristics, some based on geometric/geomatic attributes (DBSCAN, Cartesian grids, etc.), operational classification (administrative districts, postal codes, etc.) [1]. A mix of administrative zoning based on polygonal boundaries [2] and statistical geo-categorization [1, 3] is the initial approach.

Only a minor part of research in this field is related to sparse data or similar applications. When data density is high, most methods produce similar results [1]. However, when data is sparse and irregularly distributed, the underlying assumptions about the variation among samples may differ, and the choice of a spatial interpolation method and parameters may become critical [1, 4]. The authors of [5] use continuous, possibly concave, iso-contours to delineate set membership, without disrupting the primary layout. Voronoi diagrams are used in [6] to

define and delineate proximal regions around individual data points by using polygonal boundaries [2]. These diagrams have proven to be an excellent resource for large datasets, but not a reliable one for sparse data due to the uncertainty induced by the missing data. Moreover, most of the georeferenced data can be translated to Postal Codes in Canada. However, they are not particularly useful for polygonal zone delimitation around areas, because they do not necessarily follow a geometric pattern and lack location significance. On the other hand, the authors of [7] used a Takagi–Sugeno (TS) fuzzy modelling approach with fuzzy c-means (FCM) clustering to make spatial predictions of the lead concentrations in soil. For modelling under sparse data conditions, the TS fuzzy modelling approach using FCM clustering and constant width Gaussian shaped membership functions did not show any advantages over Inverse Distance Weighting (IDW) and Ordinary Kriging (OK) for the type of data tested. The ability of the three methods to predict outlier points within the respective validation sets appeared to be very similar.

After conventional data cleaning, different clustering algorithms and parametrizations are applied to the geo-data then formally evaluated using two clustering quality indexes, the Dunn index and the Calinski-Harabasz indexes. The next step is geo-fitting. While conventional approaches have been considered, only two strategies revealed applicable to scarce geo-data categorization. Tiles, created by combining the Mercator projection and a square grid, were used when data proved to be too scarce. In zones of relative abundance yet deep overlaps, ellipsoid fitting proved both performant in terms of computational load and geometrically adapted to spatially intersecting clusters. Finally, the distributions of the extracted point sets are compared to the one of the global data. Depending on their divergence, zones of specific interest, or anomalous zones, can then be used for machine learning modelling (see Fig. 1 for an overview of the area of study and Fig. 2 for an overview of the proposed strategy).

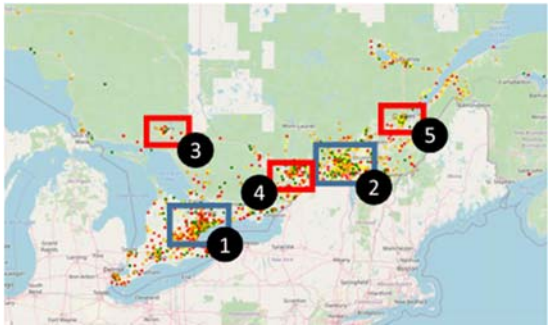


Fig. 1. Area of study: (1) Greater Toronto Area (GTA) – Contour fitting on 3 features: Ellipses. (2) Montreal- 3 features to cluster and fit. (3) Sudbury- Cartesian grid for 3 features. (4) Ottawa- Cartesian grid for 3 features. (5) Quebec - Cartesian grid for 3 features.

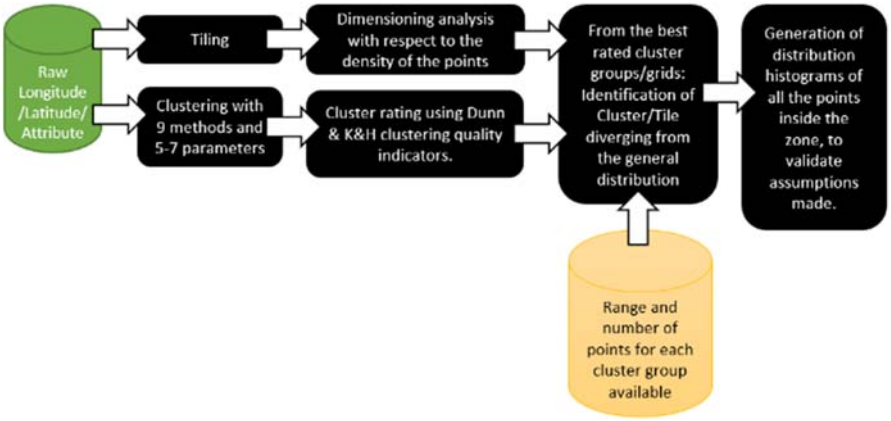


Fig. 2. Diagram of geo-categorization strategy proposed.

3. Data Processing

Several approaches were used in terms of fitting ellipsoids to define a geo-spatial profile around the clustered geo-data.

We evaluated two approaches to support categorization of new location data respective to this dataset.

The first approach consists in identifying, for a given new point, its geo-categorization, associated to a corresponding ellipsoid profile in 2D. A custom distance-based metric was used to calculate this membership. A category feature was then derived from this identification and added for two purposes:

1. Tagging the historical dataset for the machine learning model training purposes, and in particular additional parametrization;
2. Categorizing new points for the machine learning model inference purposes, in operation.

Because of the tridimensional nature of the historical dataset, when visualizing it in 2D, we can witness deep overlaps between geo-fitted profiles. However, this is not necessarily the indication of a

low-quality tridimensional clustering. It is exclusively an artifact of the elected visualization modality.

Consequently, while a visual evaluation of the geo-fitted profile can be employed for manual evaluation, it must include, for deeply overlapping categories, a careful consideration of the third coordinate.

This visual shortcoming is the core motivation behind the use of common practice clustering quality indicators, presented in the next chapter, which are to assist in evaluating the global quality of the clustering strategy with respects to geo-categorization of the source data for AI training and inference.

On the other hand, fitting ellipsoid geo-contours around clusters is not computationally intensive both for categorizing large numbers of points or for real time categorization of new data during operations. Compared to the initial cost of computing high quality clusters for geo-data, fitting contours and using them for categorization only induces negligible computation costs, which represents a significant advantage for the two proposed approaches (see Fig. 3, and Fig. 4 ellipsoids above).

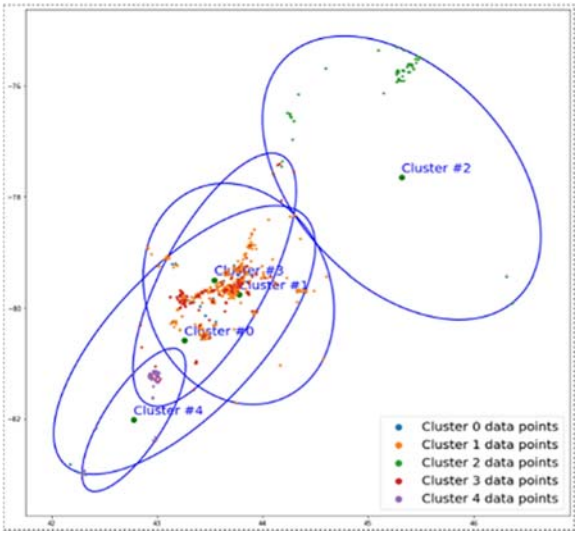


Fig. 3. High level results for candidate zone (GTA zone: Deep overlap of values, as illustrated above, are a challenge for quality geo-categorization, but the demonstrated dataset still allows for a good segmentation for permits). Zone with the largest amount of information available relative to the most significant attribute to categorize (Time to obtain permit). Still, very few applicable data for clustering, even for that best available zone (835 data input points for the whole Ontario province).

Faced with sparse geo-data, the two elected approaches were met with specific challenges. Our first approach would only give one category even if the distance metric values were very similar between two ellipses and a point. In our second approach, a weighted average of the custom metric of all ellipsoids was used. An n dimensional feature vector is returned where n corresponds to the total number of clusters and ellipsoids. This approach yielded better results than the first but was unstable when the geographic zone in question had very sparse data, or when there were no clear patterns among the tridimensional clustering.

In both cases, faced with extremely sparse data in smaller population areas, the clustering resulted in unstable or low-quality clustering, as illustrated by clustering quality indices as well as manual geo-fitting evaluation. These two issues are linked.

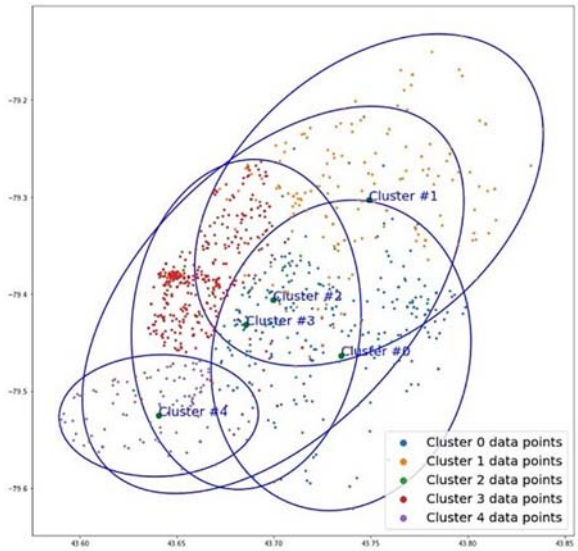


Fig. 4. Illustration of the clustering approach for one candidate zone – Central Toronto Area.

If clusters are of low quality, further segmentation into additional ellipsoids would be required in order to separate them geographically. But doing so requires significantly more input data than we had at our disposal.

Open data enrichment on zones where data was sparse was considered, but after analysis, the correlation between the datasets was too low for us to integrate the open data as input data.

Since under a certain number of values per geo-fitting profile, categorizing with the two strategies presented above did not yield good enough results, a third more stable approach was devised: Cartesian grids.

This method works for sparse data and for quickly eliminating zones where data is not available. The process selects subsets of the original data, allowing for very quick partial visual or automated analysis, without having to work on the original, complete dataset. Grid categories were determined using different statistics such as the mean, standard deviation, min and max values along with select thresholds for each.

Being fully implemented in pandas, the algorithm demonstrates a good level of performance which means that the grid resolution can be high. Generating a grid like the one in Fig. 5 across all of Ontario and Quebec takes only a few seconds to load.

In our final approach, Cartesian grids were used in conjunction with our clustering and point in ellipse approach. For each zone of interest, a zone type (ellipsoid, tile or uncategorized) was automatically determined.

To implement these different geo-categorization strategies:

- With respect to the data science lab environment, the Python language and specifically JupyterLab and Jupyter Notebook were used to script the data ingestion, transformation and curation steps (essentially outlier removal and normalization). Finally, Folium was employed for geo-visualization. Calculation of Dunn and Calinski Harabasz indexes was produced using standard Python libraries such as Sklearn.
- For geo-fitting of ellipsoids on geo-data, Sklearn has been used to implement tridimensional data clustering (geo-position and value). Specific, proprietary implementations of classic geo-fitting algorithms were provided in Python to support ellipsoid categorization.
- For Cartesian grid based geo-data categorization, Pandas was used to implement extraction and analysis.
- In order to convert the geo-categorization results for regression model training and inferences:
- A specific categorization index was produced listing zones of interest with their type (ellipsoid or tile), geometric parameters, associated with a geo-index value. For larger numbers, a geo-hash could be applied.
- In order to support further geo-categorization of new values, a specific function determining if such values belong to specific geo-categories or should be considered as uncategorized, has been implemented (it is both a simple and computationally light operation to determine point inclusion in an ellipse or in a regular tile).

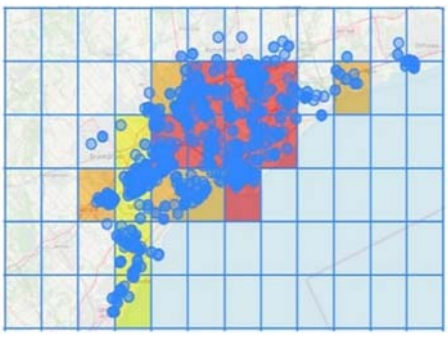


Fig. 5. Cartesian grids applied to the GTA. Grid colors represent different permit lengths.

4. The Regression Use Case Experimental Results

We evaluate the contribution of our engineered features based on geo-zones of interests onto a real-world and client-based machine learning (ML) problem. The main task is to predict the duration of completion of orders (measured in days). For this, we opted for a regression training using efficient machine learning model: LightGBM [8]. The aforementioned technique is a distributed and effective gradient boosting framework based on decision trees algorithms. The main reasons we chose LightGBM are simplicity, training speed, and lower memory usage. It is also a good ML model allowing us to test the quality of our features faster. During our experimentations, we tested other models such as linear regression, XGBoost [9], and CatBoost [10] along with LightGBM. Based on our findings, LightGBM led to the best model performance overall.

The data. The dataset that we used consisted of a large number of features coming from a company that supplies technology-based products and services. It contained more than 10 000 rows describing the available products and services in terms of its characteristics, addresses, postal codes of the delivery, date of order, and many more.

Machine learning training preparation. To prepare our ML model for training, we first proceed in adding our geolocation features that we engineered for our problem from the available data. Details on the technique is explained in the previous sections (see Section 3). After, incorporating these new features, we securely drop feature columns related to generic location information such as civic numbers, addresses, and postal codes, etc. Then we proceed to encoding the dataset. We particularly used one-hot encoding of the geo-zone features. We also used other data transformation techniques on the remaining columns (i.e., features) of the dataset such as label Encoding and Ordinal Encoding. We implemented cyclical feature encoding of dates using sine and cosine. We also implemented feature scaling technique such as StandardScaler that brings all the numerical features into a same interval value for better ML training. All of these methods are straightforward using the Scikit-learn's preprocessing library [11].

Feature importance and selection. These steps are very important when training ML models. Feature importance refers to the process of giving a score for each feature based on the ML model we are training. The score refers to the importance of each feature in our dataset. The higher the score, the higher its impact on the prediction and performance of the chosen ML model. As for feature selection, it is a process of picking the most relevant features from the dataset to use for training the ML model. The main goal is to reduce the size of the dataset while improving the performance of our predictive model. For our implementation, we used common libraries such as the feature importance function provided by LightGBM python library and mlxtend [12] feature selection python library. We particularly implemented the sequential forward feature selection. We provide in Figs. 6 and 7 the results of these steps. Due to data confidentiality, we masked out the real names of the features and renamed them to *feat1, feat2*, *etc.,* and we

only kept those that are related to our proposed geo-features. As we can see from Fig. 6, the geo-zones 5, 11, 14, 13, 9, and 12 are contributing to the performance of our LightGBM model. Likewise, if we look at the result of the feature selection chosen by the mlxtend library in Fig. 7, we can see that the engineered geo-zones are selected and are contributing to the boost of performance of our ML model.

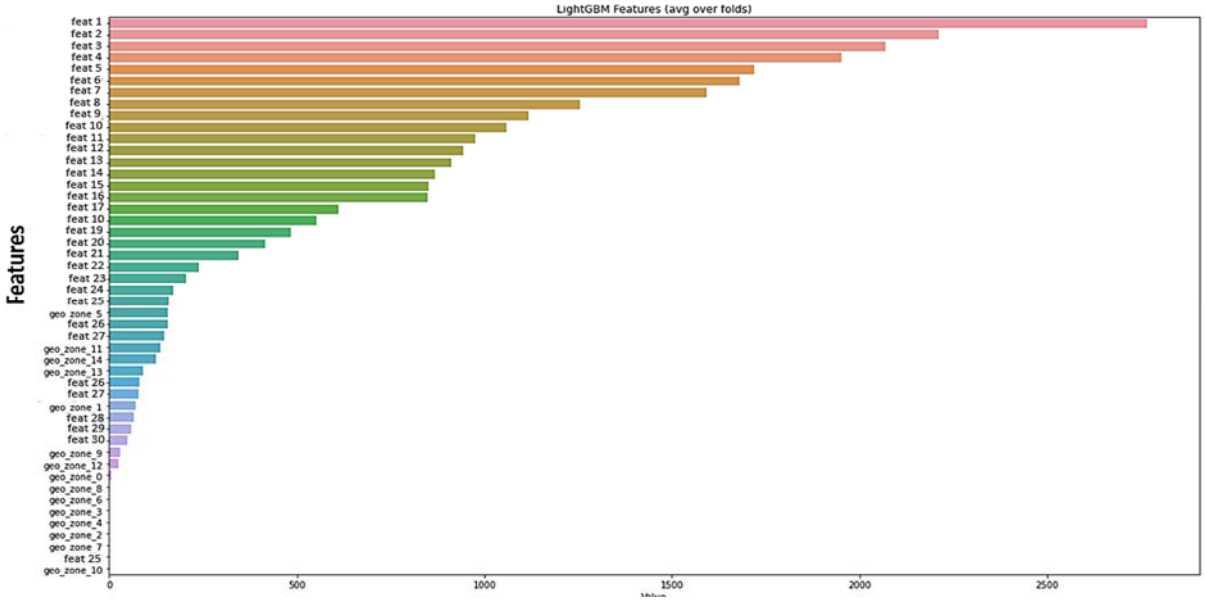


Fig. 6. Feature importance of our encoded dataset using LightGBM feature importance library.

31 features out of a toal 48 features are selected
Best combination of features gives a performance of R2 score of 0.0332
List of selected features below:
Feat 21
Feat 8
Feat 24
Feat 3
Feat 31
Feat 15
Feat 9
Feat 6
Feat 17
Feat 23
Feat 29
Feat 10
Feat 14
Feat 25
Feat 19
Feat 33
Feat 20
Feat 28
geo_zone_1
geo_zone_3
geo_zone_4
geo_zone_5
geo_zone_6
geo_zone_7
geo_zone_8
geo_zone_9
geo_zone_10
geo_zone_12
geo_zone_13
geo_zone_14

Fig. 7. Mlxtend feature selection results. This process chose a combination of 31 features out a total of 48 features with a R2 score of 0.332.

Model evaluation: To evaluate the performance of our model, we used r squared as our machine learning performance metric on a held-out dataset that we call validation set. We also used business metrics, which were defined by comparing the prediction of the model with the real value of the validation set.

The validation set was divided into three categories: Late, On-time, and Early. The success of the prediction with our ML model is when the predicted delivery date matches the target of the actual delivery date (i.e., ground truth), as shown in Fig. 8. To define these business metrics mathematically, we refer to the residual error as the value depicting the difference between the prediction and the actual delivery date (i.e., target label). This is expressed in Equation (1). As for the percentages of samples that falls in each category (late, on-time (or on-target), and early), they are calculated according to Equations (2)-(4).

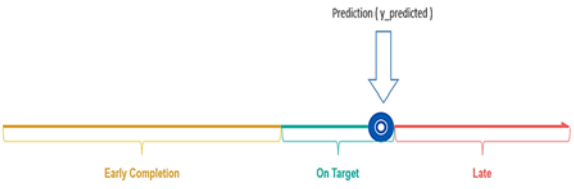


Fig. 8. Measuring the success of a prediction from a business perspective.

$$RE = Y_{true} - Y_{predicted}, \quad (1)$$

$$Late = RE > 0, \quad (2)$$

$$On-time = -21 \leq RE \leq 0, \quad (3)$$

$$Early = 0 < RE \quad (4)$$

Model training. We used an asymmetrical cost function called Quadratic-Quadratic Cost [13]. It is defined in Equation (5). This loss is an asymmetric version of the traditional symmetrical loss the mean squared error (MSE). Two main hyper parameters to this loss function, which are a and b. Setting an equal to be leads to the traditional MSE loss. However, setting a>b we are penalizing more the scenarios where the residuals are higher than zero, as shown in Fig. 9. The idea of incorporating the asymmetrical loss function is to have a model that penalizes over-estimation less than underestimation of delivery dates. In other words, we would like the model to penalize more when the prediction is a lot before than the actual delivery date (lateness of the delivery), since this leads to business risks and alternatively, we would like to penalize less when the predicted date is slightly after than the actual delivery date (earliness of the delivery).

$$QQC(e_i) = \begin{cases} a\, e_i^2, \; if \; e_i > 0; \\ b\, e_i^2, \; if \; e_i \leq 0 \end{cases} \quad (5)$$

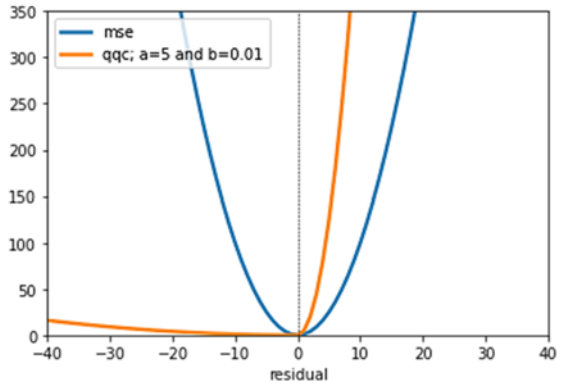


Fig. 9. Quadratic-Quatratic Cost (QQC) function, an example.

Results. To showcase the importance of our features with respect to model performance, we compared our results to a baseline model. This baseline consists of training a LightGBM model with all features of the dataset without any addition of geo-zone features. We provide the results of our experiments in Table 1 and in Fig. 10. From Fig. 10, we could see that by using the geo-zone features, this substantially increased the performance on both ML performance and business success metrics. An increase of on time delivery percentage from 31.2 % up to 32.4 % when geo-zone features were incorporated during training of the data.

Table 1. The performance of the LightGBM model with and without incorporating the Geo-zones of interest after automatic feature selection

Model	R_squared
LGBM without Geo-zones (Baseline)	0.2353
LGBM with Geo-zones	0.332

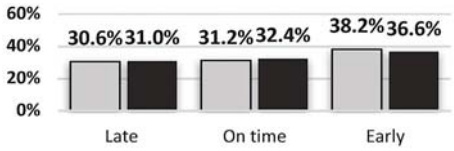


Fig. 10. Performance of the Light GBM regression model in terms of Business metrics (Eq. (1)-(4)). Performance without and with the geo-features in gray and black, respectively.

3. Conclusions

We presented, through the use case of a specific customer situation involving machine learning model training, multiple data geo-categorization challenges that are often met in the industry. Through specific data preparation, conventional clustering strategies associated with a structured qualification, index-based methodology as well as adapted geo-fitting approaches, it was possible to extract from sparse operational datasets relevant geo-categorization, which revealed efficient in contributing to a corresponding regression model's level of performance, while conventional segmentation approaches such as Voronoi diagrams and similar could not be applied efficiently given the characteristics and in particular the scarcity of the source geo-data.

Further investigation could be devised, in order to further automate determination methods for the optimal parametrization of all three proposed approaches. At the time of edition of this publication, an initial visual analysis of the geo-data, associated with iterative evaluation, especially for clustering quality evaluation, is still required. Such automated methods are possible to develop for given use cases, and that appropriate tooling could also be designed to assist and further automate this initial analysis phase.

However, the presented approaches are already applicable to categorization of sparse geo-data and supported positive results for model training and inferences: in the context of the AI supported smart supply chain project for Bell Canada (Bell Business Markets), the performance was improved by incorporating XGB regression model and custom cost optimization as well as hyperparameter tuning of the models.

Acknowledgements

This work was supported by the Scale AI program and Bell Canada.

References

[1]. P. A. Burrough, R. A. McDonnell, Principles of Geographical Information Systems, *Oxford University Press*, Oxford, 1998.

[2]. B. Harrison, Data room management for mergers and acquisitions in the oil and gas industry, *Developments in Petroleum Science*, Vol. 69, 2020, pp. 145-167.

[3]. A. D. Hartkamp, K. De Beurs, A. Stein, J. W. White, CIMMYT, Interpolation techniques for climate variables, *Natural Resources Group*, Mexico, D. F., 1999.

[4]. R. Thomas, U. T. Khan, C. Valeo, F. Talebzadeh, An investigation of Takagi-Sugeno fuzzy modelling for spatial prediction with sparsely distributed geospatial data, *Environments*, Vol. 8, 2021, 50.

[5]. C. Collins, G. Penn, S. Carpendale, Bubble sets: Revealing set relations with isocontours over existing visualizations, *IEEE Transactions on Visualization and Computer Graphics*, Vol. 15, Issue 6, November-December 2009, pp. 1009-1016.

[6]. J. Hou, Y. Yuan, S. Schwertfeger, Area graph: generation of topological maps using the Voronoi diagram, in *Proceedings of the 19th International Conference on Advanced Robotics (ICAR'19)*, 2019, pp. 509-515.

[7]. J. Li, A. D. Heap, A Review of Spatial Interpolation Methods for Environmental Scientists, Record 2008/23, *Geoscience Australia*, 2008.

[8]. G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, T. Y. Liu, LightGBM: A highly efficient gradient boosting decision tree, in *Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS'17)*, 2017.

[9]. T. Chen, T. He, M. Benesty, V. Khotilovich, Y. Tang, H. Cho, K. Chen, xgboost: Extreme Gradient Boosting, R Package Version 0.4-2, https://mran.microsoft.com/snapshot/2015-11-30/web/packages/xgboost/index.html

[10]. A. V. Dorogush, V. Ershov, A. Gulin, CatBoost: Gradient boosting with categorical features support, *arXiv Preprint*, 2018, arXiv:1810.11363.

[11]. F. Pedregosa, et al., Scikit-learn: Machine learning in Python, *Journal of Machine Learning Research*, Vol. 12, 2011, pp. 2825-2830.

[12]. S., Raschka, MLxtend: Providing machine learning and data science utilities and extensions to Python's scientific computing stack, *Journal of Open-Source Software*, Vol. 3, Issue 24, 2018, 638.

[13]. K. Dress, S. Lessmann, H.-J. von Mettenheim, Residual value forecasting using asymmetric cost functions, *International Journal of Forecasting*, Vol. 34, Issue 4, 2018, pp. 551-565.

(8759)

Convolutional Recurrent Network Design for Real-time Speech Enhancement

Amir Rajabi and Mohammed Krini
Laboratory for Signal Processing, Aschaffenburg University of Applied Sciences,
Aschaffenburg, Germany
E-mail: amir.rajabi@th-ab.de, mohammed.krini@th-ab.de

Summary: In this contribution, the realization of an end-to-end speech enhancement system is presented. The proposed method is based on the convolutional recurrent network (CRN) proposed in [1]. In the first stage, network hyperparameter tunings of the CRN U-Net were utilized as our baseline. In addition, a new skip connection to the network was implemented. In the second stage, the architecture of CRN U-Net was extended using deeper encoder and decoder architecture and bidirectional LSTM layers, with the goal of achieving a better and more generalized pattern recognition of any form of seen and unseen dynamic noise. Furthermore, an effective approach is proposed to reduce the complexity of the proposed network while simultaneously enhancing noise reduction and speech quality. Finally, the proposed approaches are analyzed objectively and compared.

Keywords: Convolutional Recurrent U-Net for Speech Enhancement (CRUSE), Bidirectional LSTM layer, Encoder decoder architecture, End-to-end speech enhancement.

1. Introduction

The goal of a speech enhancement system in real-time communication is to improve the speech quality and intelligibility. The additive background noise in the speech signal can negatively influence audio/video communication, hearing aids and speech or speaker recognition [2]. However, it is challenging to preserve undistorted speech while reducing noise, resulting in the performance of speech enhancement systems being bounded by the trade-off between speech distortion and noise suppression [3].

In the present work, the CRN U-Net will be analyzed as a starting reference. After that, some optimization techniques for CRN U-Net will be proposed. Furthermore, the new suggested neuronal network, which is an extended version of CRN U-Net, will be illustrated. It has a deeper encoder-decoder structure and in between two stacked long-short term memory (LSTM) followed by two bidirectional LSTM layers were utilized.

2. State of the Art and Related Work

Over the past years many solutions and algorithms for speech enhancement have been proposed. The most widely implemented algorithms for decades were spectral subtraction [4] or approaches based on minimum mean-square error (MMSE) [5]. In stationary noise reduction, short-time spectral estimation is utilized to adaptively compute suppression gains for the time-frequency spectral coefficients of a noisy signal [6]. However, because of the statistical variance of the short-time spectral estimates, noise artefacts are known as musical noise that may be quite annoying and possibly more diverting than the original noise could occur [7].

In the past few years, the use of deep learning for noise suppression has increased dramatically, allowing speech enhancement to make tremendous progress [1, 8, 9]. Therefore, in this work the deep learning approach was employed as the basis. While most of the denoising deep learning models show a significant improvement in the estimation of the magnitude spectrum compared to the classical methods, the noisy phase is still used to transform the spectrum back to a waveform in such models. However, this is not a limitation of the generality in the resynthesized enhanced audio at high Signal Noise Ratio (SNR) level, where the magnitude plays a more important role [10], because the noisy phase in this SNR range could be considered as clean phase and be used for reconstructing the enhanced signal in time domain. In addition, the phase cannot be substantially modeled, so a direct estimate of the phase is not straightforward [11].

The focus of the paper is set on estimating the enhanced spectrum magnitude. As already mentioned, with the motivation of the CRN and recent works [1, 9, 12], the goal is the extension of a noise and speaker independent end-to-end model, which optimizes the magnitude spectrum of the estimated clean speech. Regarding the implementation and extension of a speech enhancement model, the application of LSTM layers, which are especially good at processing sequences such as speech, has been investigated [1, 12]. Furthermore, the implementation of BLSTM as proposed in [13] was realized in this work. The motivation was the overlay of the performance of BLSTM against LSTM for speech recognition [14-16] and video content analysis [17]. Additionally, a new type of add-convolutional skip connection will be proposed.

The structure of this paper is arranged as follows: First, the signal model and feature extraction are

presented. Afterwards, the training objective is defined and an overview of the feature generation and enhanced audio resynthesis is described. In addition, the topology and architecture of the model will be considered in more detail. At the end, the results are shown leading to a conclusion.

3. Model and Learning

3.1. Signal Model

It is to be assumed that the microphone captures the clean speech signal *s(t)* and the additional uncorrelated pure noise *d(t):*

$$x(t) = s(t) + d(t), \tag{1}$$

where t is the time index. For further analysis, the audio signal has to be transformed into the frequency domain by Short-Time Fourier Transform (STFT). For this purpose, the noisy signal was first segmented into the overlapping frames and subsequently a Hann window function hhann, k was applied. Afterwards discrete Fourier Transformation (DFT) was applied to the weighted segments according to:

$$X_\mu(n) = \sum_{k=0}^{N-1} x(nr - k) h_{hann,k} e^{-j\Omega_\mu k}, \tag{2}$$

where *n*, *N, r* and $\mu = 0,1,..., N-1$ represent the frame index, frame length, frame shift and frequency band, respectively. The observed signal is represented in the STFT domain as follows:

$$X_\mu(n) = S_\mu(n) + D_\mu(n), \tag{3}$$

where $S_\mu(n)$ denotes the clean speech magnitude and $D_\mu(n)$ the noise magnitude signal.

The aim of a full speech enhancement system is the estimation of the enhanced speech signal using a mapping function $G_{\theta,\mu,n}$:

$$G_{\theta,\mu,n}: X_\mu(n) \rightarrow S_\mu(n), \tag{4}$$

where θ represents the model parameters. This nonlinear mapping function is implemented by the proposed neural network, which is explained in more detail in the next section.

3.2. Training Objective

In this work, the magnitude spectrum of clean and noisy signals according to Eq. (2) was used as input. In the first part of the experiment, where the baseline CRN is analyzed, the 161-dimensional spectrum was used, which corresponds to half of the entire spectrum. In the second part of the experiment, the 257-dimensional magnitude input was fed to the extended version of CRN U-Net. Further information will be presented in Section 4.2.

Firstly, it was trained with mean-squared error (MSE), which is the most commonly used loss function for regression-based deep learning. However, further examinations showed that the mean-absolute error (MAE) provides better results, which can be explained by a better fitting to the signal/noise distribution.

3.3. Bidirectional Recurrent Network

In this work, to build a deep recurrent part in CRN U-Net the LSTM layers were extended with BLSTM. The dynamic interaction between the input and output of the BLSTM Layer is described below. For illustration purposes it is assumed that the following input feature sequences

$$\boldsymbol{Z_\mu}(n) = \{Z_\mu(0), Z_\mu(1), ..., Z_\mu(N_T)\}, \tag{5}$$

is fed into a unidirectional LSTM. The analysis is performed on the sequential relations in forward direction and the output $\boldsymbol{y_t}$ is calculated as:

$$\begin{aligned} \boldsymbol{h_t} &= \propto(\boldsymbol{Z_\mu}(\boldsymbol{n})^T \boldsymbol{W_x} + \boldsymbol{W_h h_{t-1}} + \boldsymbol{b_h}), \\ \boldsymbol{y_t} &= \propto(\boldsymbol{h_t W_y} + \boldsymbol{b_y}) \end{aligned} \tag{6}$$

Here $\boldsymbol{W_x} \in N_T \times N_n$ contains the connection weights for the inputs of the current frame, where N_n denotes the number of neurons used within the unidirectional LSTM. $\boldsymbol{W_h} \in N_n \times N_n$ contains the connections weights for hidden units. Besides, $\boldsymbol{W_y} \in N_n \times N_T$ includes the connection weights the connection weights for the outputs of the previous frame. Moreover, the $\boldsymbol{b_h}$ is a vector of size of the number of neurons containing each neuron's bias term and $\propto$ represent the hidden layer activation functions. Note that the inclusion of the LSTM layers does not change the causality of the system, since the pastime sequences always are used in further processing. The idea of investigating the input sequence in both the forward and backward direction, to capture the global temporal structure, leads to the implementation of the BLSTM, which then provides a bilateral denoising.

In this layer the input is fed into forward and backward recurrent layers, both connected to the same output layer. The BLSTM calculates the output y_t with the iteration of the forward layer from $n = 0$ to N_T and the backward layer from $n = N_T$ to 0. (see Eq. 7).

$$\begin{aligned} \boldsymbol{h_t^f} &= \propto(\boldsymbol{Z_\mu}(\boldsymbol{n})^T \boldsymbol{W_x^f} + \boldsymbol{W_h^f h_{t-1}^f} + \boldsymbol{b_h^f}), \\ \boldsymbol{h_t^b} &= \propto(\boldsymbol{Z_\mu}(\boldsymbol{n})^T \boldsymbol{W_x^b} + \boldsymbol{W_h^b h_{t+1}^b} + \boldsymbol{b_h^b}), \\ \boldsymbol{y_t} &= \propto(\boldsymbol{h_t^f W_y^f} + \boldsymbol{h_t^b W_y^b} + \boldsymbol{b_y}). \end{aligned} \tag{7}$$

3.4. Evaluation Metrics

To compare the performance of the trained models, a suitable criterion has to be defined. This optimization

criterion has to provide a fair evaluation of the speech quality and speech intelligibility by objective analysis. The distorted speech is divided into two categories: speech distortion, which influences the original speech signals, and noise distortion, which influences background noise. Therefore, an effective enhancement system should be able to provide the correction of both types of distortion. For that reason different types of metrics should be considered in a combined criterion.

Due to the high correlation between subjective and objective evaluation that was examined in [18], [19], Perceptual Evaluation of Speech Quality (PESQ) is used. In [20] it was shown that the Log-Likelihood Ratio (LLR) [21] and the frequency-weighted segmental SNR (fwSNRseg) [22] are good alternatives for PESQ to evaluate the quality of speech and noise distortion. Including the overall quality measure used in [9] with scale-invariant Signal-to-Distortion Ratio (siSDR), the heuristic optimization equation ζ is represented as follows:

$$\zeta = PESQ + 0.2*siSDR + 0.3*fwSNRseg - LLR \quad (8)$$

It has to be mentioned, that the short-time objective intelligibility (STOI) [23] was calculated for investigating the speech intelligibility. However, due to its very small change interval but it wasn't considered in the optimization criterion. The speech enhancement system is depicted in Fig. 1, with the training and test scenarios placed adjacently in order to visualize the entire process. In the test mode, the phase of the noisy signal is used to reconstruct the enhanced signal.

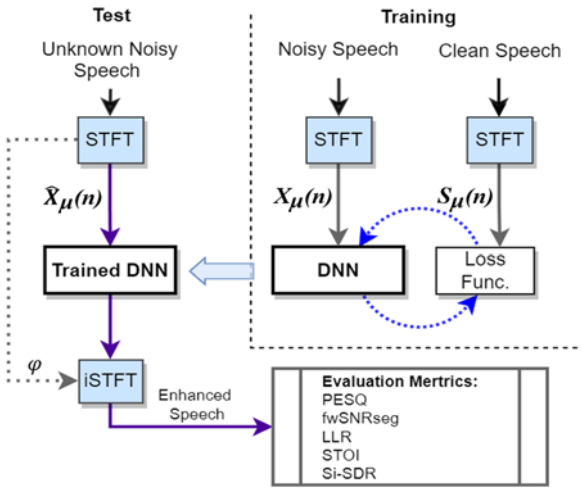


Fig. 1. The procedure from feature extraction to the reconstruction of the data in the time domain and evaluation of the results.

4. Proposed Architecture and Topology

In this section, the proposed architecture as illustrated in Fig. 2 will be presented. The extended model has six convolution layers in the encoder and six convolutional transpose layers in the decoder. The kernel numbers or channels start from 16, 32, 64, 128, 256 and 128 per layer in the encoder and in the decoder from 128, 64, 32, 16, 8 and 1 in the output layer. All convolutional layers have a kernel of size (2, 4) in time and frequency dimensions. Moreover, with a stride of (1, 2) the convolutional kernels downsample the input feature along the frequency dimension. It is to be mentioned that such a design will decrease the number of model parameters and the computing complexity in the recurrent component of the CRN without loss of prediction accuracy. Besides, the activation function of all convolutional layers is ELU except the output layer, which has a PRELU activation.

In [1], [12] a stack of multiple LSTM layers was implemented, which has been extended in this work with the implementation of the BLSTM. To feed these stacked layers the required input shape by the LSTM should be prepared by flattening the frequency and the depth dimension of the encoder. Therefore a 2D feature map of the recurrent part of the suggested model along the time axis could be achieved. Subsequently, the output of the BLSTM layer will be re-reshaped in the form of a suitable input shape for the decoder. In addition, after every LSTM layer, a dropout layer was used to avoid any possible overfitting. Furthermore, the skip connection was equipped with a convolutional layer, which will be presented in more detail in the following subsection. Its implementation in the proposed architecture in Fig. 2 was found to produce better and more effective enhancement results.

4.1. Causality and Online Task

For BLSTM layers allow per mathematical definition non causality. In reality the information of the future is problematic and causality can be violated, especially when it comes to online tasks. However, because of the assumption that only the sequences from $n = 0$ until $n = N_T$ should be considered, the causality is no longer a problem. It means that h_{t+1} must be set equal to zero in the Eq. 7. Furthermore, in order to consider the denoising problem in examination more deeply, a delay is a simple solution to apply this theory on real world temporal problems. With this delay between inputs and the corresponding output, the network is given the opportunity to analyze the past sequences with the influence of present sequences. In this work to ensure the causality of the system in real-time, BLSTM layers are encoded framewise along the time axis. In Section 6, the results from the above assumptions, which further processes the extracted features of the encoder along time dimensions, will be presented.

4.2. Convolutional Skip Connection

In both networks [1, 12], each convolutional layer of the encoder is connected to its corresponding

decoder layer with a skip connection. There are two methods on how the skip connection can be implemented. One is the concatenation, which means duplication of the number of inputs from the decoder layers. Another is the addition of the outputs from the encoders and inputs from the decoders. In [12], a channel-wise scaling in add-skip connections is recommended, implied by the causal convolutional layer (1, 1), and showed an improvement in performance. In this work, this aspect was investigated with different kernel sizes. A descending kernel size from the first encoder layer to the last layer led to better results. That means, the first encoder-layer, using a convolutional layer with C_1 channels and the (v, v) kernel size, was connected with the last encoder-layer, where v corresponds to the number of encoder layers.

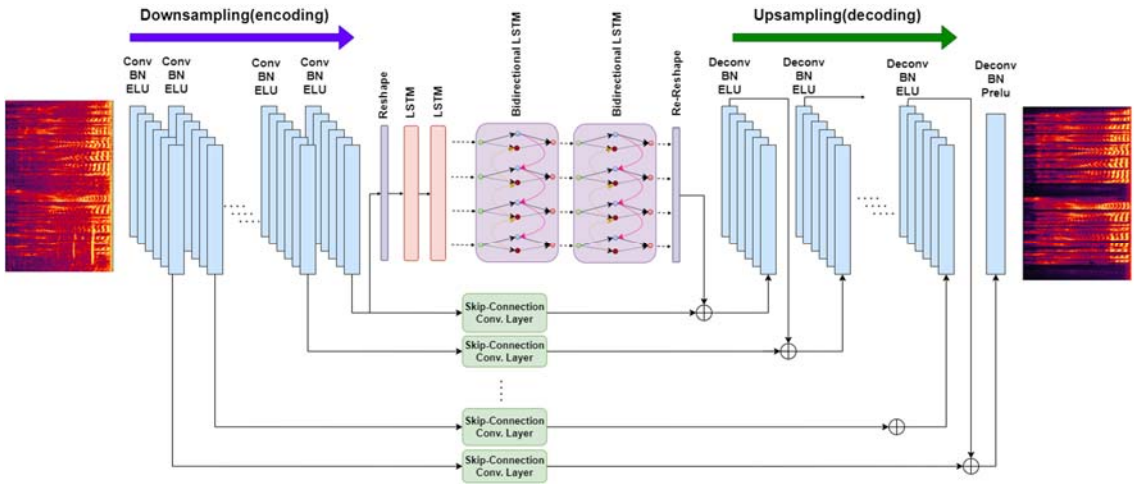


Fig. 2. The architecture of the new CRN approach. As it was mentioned in Section 3, the BLSTMs are implemented to continue the processing of the extracted features along the frequency axis. The BN and ELU/PReLU represent batch normalization and activation function respectively.

5. Experimental Setup

5.1. Dataset

To demonstrate and compare the effectiveness of the proposed network against the CRN U-Net as the baseline, The Edinburgh Data Share [24, 25] is applied, which includes 28 speakers- 14 male and 14 female. From each speaker, there are 400 utterances available. In [25] it was mentioned that in the training data a combination of artificially generated noises (speech-shaped noise and babble) and eight types of real recorded noises from the Demand Database [26] was used. The other eight selected noises were: a domestic noise (inside a kitchen), an office noise (meeting room noises), three public space noises (cafeteria, restaurant and train station), two transportation noises (car and subway) and the street noise (busy traffic intersection). SNR-levels used for the training were: 15, 10, 5 and 0 dB. In summary, a total of 40 noise conditions were produced due to 4 SNR and 10 noise types. For an even tougher performance test of the speech enhancement system, 5 noises that were not employed in the training phase were used in the test data set to ensure the noise independence. In addition, two other speakers were selected to ensure the speaker independence of the enhancement system. Otherwise, SNRs of the test set were slightly higher, namely 17.5, 12.5, 7.5 and 2.5 dB.

5.2. Training Setup

The mixed dataset was downsampled to 16 kHz. The input of the neural networks as mentioned in Section 2 were the 161- and 257-dimensional magnitude spectrum (half of the entire spectrum) calculated from STFT, while the Hann windows lengths with respect to the input sizes were 320, 512 with a 50 percent overlap in both cases. All models are trained with Adam optimizer [27] and MAE loss function. The learning rate was reduced during the training, starting from 0.0001 and reduced by a factor of 0.8 if the validation loss did not decrease after 5 epochs. In addition, an early stopping function was employed to avoid overfitting.

6. Result and Discussion

The results for the convolutional-add-skip connections, which are discussed in the last section are shown in Table 1. In view of this table, it is seen that by interesting cheap descend convolutions in the proposed model an improvement can be noticed.

The experimental results of the performance of the optimized model as well as the baseline CRN U-Net (mentioned as CRN-Softplus-concat-skips) are considered in Table 2 and Table 3.

In Table 2, it is necessary to mention that the input size was a 161-dimensional magnitude spectrogram. All kernel sizes have been set by (2, 3) with a stride of

(1, 2). After implementation and evaluation of the baseline model, the first optimization was performed with an activation function PReLU and add-skip connection. The result can be seen in the second line of Table 2. The optimization was continued with the implementation of 2 BLSTM layers after the LSTM layers and the mentioned add-convolutional skip connection. Every CRN model has 4, 5 or 6 encoder decoder dimensions and 2 LSTM layers in between and they are extended 2 BLSTM layers.

In the second part of the investigation a larger input dimension 257 was chosen. It might be like other classical machine learning tasks that the larger the dimension, the more difficult the task becomes for the neural network.

In contrast, it was also clear that in larger dimensions or larger FFT more speech information per every time frame is available. The results of this investigation are shown in Table 3.

Table 1. Comparison between the improvements of the model versus the complexity of the model (number of trainable parameters). With the description *add conv descending* is meant that the kernel size is decreasing in (1, 1) steps from (5, 5) in the last decoder layer to (1, 1) in the first decoder layer.

Skip Connection	Δ computational complexity[k]	Δ ζ
Add skips	0	0
Add conv (1,1) skips	87.7	0.1
Add conv (3,3) skips	768.1	0.17
Add conv (5,5) skips	2182.2	0.33
Add conv descending	192.2	0.3

Table 2. Results of evaluation metrics for CRN U-Net(second row) and extended models. The Softplus and PReLU represent the activation function of the last layer of the network. The third through fifth rows contain the optimized and the extended models, respectively. The green line represents the best result in this comparison.

Model	PESQ	fwSNRseq	Stoi	LLR	Si-SDR
Noisy speech	1.97	10.97	0.92	0.59	8.44
CRN5-Softplus-concat skips	1.86	11.95	0.93	1.23	16.52
CRN5-PReLU-add skips	2.4	12.92	0.92	0.87	16.64
CRN5-PReLU+2BLSTM-add skips	2.47	13.72	0.92	0.84	16.84
CRN5-PReLU+2BLSTM-descending conv-add skips	2.79	15.01	0.94	0.53	16.87
CRN6-PReLU+2BLSTM-descending conv-add skips	2.74	15.26	0.94	0.8	16.99

Table 3. Results of different analyzed models, which has 4, 5 or 6 encoder-decoder layers. The input size was a 257-dimensional magnitude spectrogram. *Here the stacked LSTMs and BLSTMs have a PReLU activation function. The last row contains the result of the proposed model.

Model	PESQ	fwSNRseq	Stoi	LLR	Si-SDR
Noisy speech	1.97	10.97	0.92	0.59	8.44
CRN4-Softplus-concat skips	1.76	11.16	0.93	1.31	16.41
CRN4+2BLSTM-C = {16,16,32,64}-descending-conv-add-skips	2.81	15.3	0.94	0.43	17.79
CRN4+2BLSTM-C = {16,16,32,128}-descending-conv-add-skips	2.84	15.49	0.94	0.42	17.88
CRN5+2BLSTM-C = {16,32,64,128,128}-descending-conv-add-skips	2.84	15.33	0.94	0.46	17.99
CRN6+2BLSTM-C = {16,32,64,128,256,256}-descending-conv-add-skips*	2.82	15.53	0.94	0.38	17.84
CRN6+2BLSTM-C = {16,32,64,128,256,128}-descending-conv-add-skips*	2.86	15.61	0.94	0.39	17.87

In addition, it should be taken into consideration that the complexity of the model plays a decisive role compared to the improvement of the enhanced signal. To illustrate this assumption more clearly, consider Fig. 3. Here all selected models were represented on the basis of the improvement of the enhanced signal, which was obtained by defined criterion, see Eq. (5). in the comparison of the computational effort. It can be seen that the proposed model demonstrates a major improvement in the quality of the enhanced signal in comparison to CRN U-Net, although the complexity of the model is significantly lower. Besides, Fig. 3 also shows a classic Wiener filter, which is actually an unfair comparison, but it shows the benefit of using deep neural networks for speech enhancement. As can be seen in Fig. 3 and Table 3, the proposed model (CRN6+2BLSTM-C= {16, 32, 64, 128, 256, 128}-descending-conv-add-skips) consistently outperforms all other architectures considered in this study.

7. Conclusions and Future Work

All experimental results have shown that the proposed model architecture outperforms CRN U-Net and can be generalize to unseen sounds as well. The proposed model has demonstrated the effectiveness of the proposed model in speech enhancement. In summary, the whole proposed neural network that combines CRN U-Net, descending convolutional-add-skip connections and bidirectional neural network can

be considered as a non-linear end-to-end filter for noise reduction in the frequency domain. In further steps, the improvement of the phase should also be taken into consideration, therefore the extension of the proposed model with a phase correction part is of major interest.

For the visual representation of the differences between the proposed model and the other models, the comparison of the spectrograms of the representative models are shown in Fig. 4.

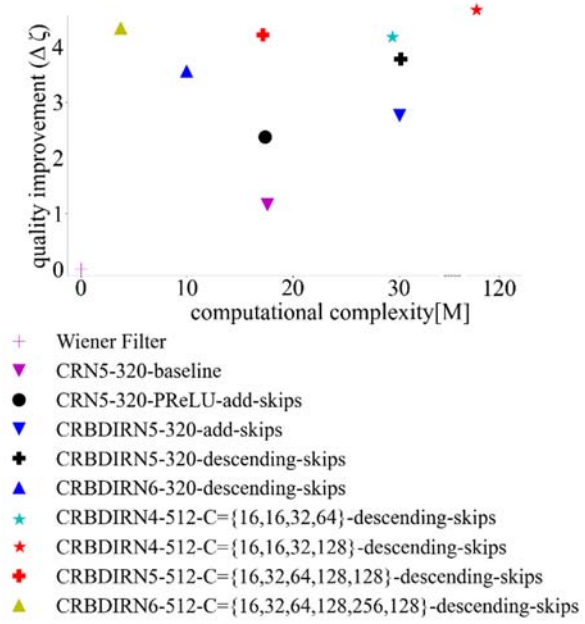


Fig. 3. Quality improvement vs. computational complexity. The green triangle shows the proposed model and the purple triangle displays the CRN U-NET considered as the baseline.

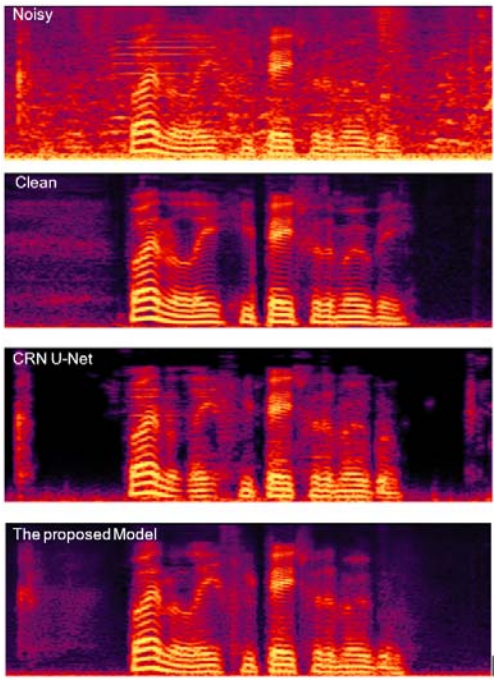


Fig. 4. The comparison of the spectrogram of noisy and clean test sample, together with the denoised spectrograms using CRN U-Net and the proposed Model.

References

[1]. K. Tan, D. Wang, A Convolutional Recurrent Neural Network for Real-Time Speech Enhancement, in *Proceedings of the Conference of the International Speech Communication Association (INTERSPEECH'18)*, 2018, pp. 3229-3233.

[2]. P. C. Loizou, Speech Enhancement: Theory and Practice, *CRC Press*, 2013.

[3]. S. Vihari, A. S. Murthy, P. Soni, D. C. Naik, Comparison of speech enhancement algorithms, *Procedia Computer Science*, Vol. 89, 2016, pp. 666-676.

[4]. S. Boll, Suppression of acoustic noise in speech using spectral subtraction, *IEEE Transactions on Acoustics, Speech and Signal Processing*, Vol. 27, Issue 2, 1979, pp. 113-120.

[5]. Y. Ephraim, D. Malah, Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator, *IEEE Trans. Speech and Audio Process*, Vol. 32, Issue 6, 1984, pp. 1109-1121.

[6]. J. Thiemann, Acoustic noise suppression for speech signals using auditory masking effects, PhD Thesis, Department of Electrical & Computer Engineering, *McGill University*, Montreal, Canada, July 2001.

[7]. A. Lukin, J. Todd, Suppression of musical noise artefacts in audio noise reduction by adaptive 2D filtering, in *Proceedings of the 123rd Audio Engineering Society Convention,* New York, USA, October 5-8, 2007.

[8]. T. Gao, J. Du, L. -R. Dai, C.-H. Lee, Densely connected progressive learning for LSTM-based speech enhancement, in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'18)*, 2018, pp. 5054-5058.

[9]. S. Braun, H. Gamper, C. K. A. Reddy, I. Tashev, Towards efficient models for real-time deep noise suppression*, arXiv Preprint*, 2021, arXiv:2101.09249.

[10]. D. Wang, Jae Lim, The unimportance of phase in speech enhancement, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. 30, Issue 4, August 1982, pp. 679-681.

[11]. P. Mowlaee, J. Kulmer, J. Stahl, F. Mayer, Single Channel Phase-Aware Signal Processing in Speech Communication: Theory and Practice, *Wiley*, Dec. 2016.

[12]. Z. Zhang, Z. Sun, J. Liu, J. Chen, Z. Huo, X. Zhang, Deep recurrent convolutional neural network: Improving performance for speech recognition, *arXiv Preprint*, 2016, arXiv:1611.07174.

[13]. M. Schuster, K. K. Paliwal, Bidirectional recurrent neural networks, *IEEE Transactions on Signal Processing*, Vol. 45, Issue 11, Nov. 1997, pp. 2673-2681.

[14]. A. Graves, N. Jaitly, A.-R. Mohamed, Hybrid speech recognition with Deep Bidirectional LSTM, in *Proceedings of the IEEE Workshop on Automatic*

Speech Recognition and Understanding (ASRU'13), 2013, pp. 273-278.

[15]. A. Zeyer, P. Doetsch, P. Voigtlaender, R. Schlüter, H. Ney, A comprehensive study of deep bidirectional LSTM RNNS for acoustic modeling in speech recognition, in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'17)*, 2017, pp. 2462-2466.

[16]. Z. Yu, G. Chen, D. Yu, K. Yao, S. Khudanpur, J. Glass, Highway long short-term memory RNNs for distant speech recognition, *arXiv Preprint*, 2015, arXiv:1510.08983.

[17]. Y. Bin, Y. Yang, et al., Bidirectional long-short term memory for video description, in *Proceedings of the 24th ACM international conference on Multimedia*, 2016.

[18]. J. Beerends, E. Larsen, N. Iyer, J. Vugt, Measurement of speech intelligibility based on the PESQ approach, in *Proceedings of the Workshop Measurement of Speech and Audio Quality in Networks (MESAQIN'04)*, Prague, Czech Republic, June 2004.

[19]. A. Prodeus, V. Didkovskyi, M. Didkovska, I. Kotvytskyi, D. Motorniuk, A. Khrapachevskyi, Objective and subjective assessment of the quality and intelligibility of noised speech, in *Proceedings of the International Scientific-Practical Conference Problems of Infocommunications Science and Technology (PIC S&T'18)*, 2018, pp. 71-74.

[20]. Y. Hu, P. C. Loizou, Evaluation of objective quality measures for speech enhancement, *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 16, Issue 1, Jan. 2008, pp. 229-238.

[21]. S. Quackenbush, T. Barnwell, M. Clements, Objective Measures of Speech Quality, *Prentice-Hall*, Englewood Cliffs, NJ, 1988.

[22]. J. Tribolet, P. Noll, B. McDermott, R. E. Crochiere, A study of complexity and quality of speech waveform coders, in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'78)*, 1978, pp. 586-590.

[23]. C. H. Taal, R. C. Hendriks, R. Heusdens and J. Jensen, An Algorithm for Intelligibility Prediction of Time–Frequency Weighted Noisy Speech, *IEEE Transactions on Audio, Speech, and Language Processing,* Vol. 19, No. 7, Sept. 2011, pp. 2125-2136.

[24]. C. Valentini-Botinhao, Noisy Speech Database for Training Speech Enhancement Algorithms and TTS Models, *University of Edinburgh, School of informatics. Center for Speech Technology Research (CSTR)*, Aug. 2017.

[25]. C. Valentini-Botinhao, et al., Speech enhancement for a noise-robust text-to-speech synthesis system using deep recurrent neural networks, in *Proceedings of the Conference of the International Speech Communication Association (INTERSPEECH'16)*, 2016, pp. 352-356.

[26]. J. Thiemann, N. Ito, E. Vincent, The diverse environments multi-channel acoustic noise database: A database of multichannel environmental noise recordings, *J. Acoust. Soc. Am*., Vol. 133, Issue 5, 2013, pp. 3591-3591.

[27]. D. P. Kingma, J. Ba, Adam, A method for stochastic optimization, in *Proceedings of the 3rd International Conference for Learning Representations*, 2015.

(S02)

Big Data as a Conceptual Basis for BI Systems

Zaslavskaya Veronika L. [1] and Romatsova Natalia V. [2]

[1] COO Zello Russia, Deputy Chairman of the 'Artificial Intelligence' Committee of RUSSOFT,
[2] Federal State Autonomous Educational Institution of Higher Education St. Petersburg State Electrotechnical University LETI Named after V.I. Ulyanov (Lenin)
197022, Russia, St. Petersburg, Professora Popova str., 5
Tel.: +79219494979
E-mail: veronika@zello.com

Summary: In this article, the author considers BIG data as a conceptual basis for business intelligence systems. The information obtained as a result of the development of such systems is a valuable asset and competitive advantage of the company, provided that a unique, critical data segment is identified. The author considers it extremely important to develop and distribute industry-specific BI solutions available for small and medium businesses. As an example, the author cites the development of a predictive business analytics system using the example of a network fitness club.

Keywords: Big data, BI-systems, Data visualization.

1. Introduction

The volume of the global market for big data and business intelligence (BDA) has a high growth rate (more than 10 % per year) and, according to IDC experts, amounted to $215.7 billion in 2021. The main driving force that stimulates enterprises to conduct research in the field of big data analysis and interpretation, is the competition. As a result of such research, the company can obtain information about internal processes and external factors, identify patterns, correlations, and risks, and make accurate forecasts. Thus, this information used to make management decisions becomes a significant competitive advantage, and data science, one of the key technological areas.

The Big Data market in Russia is also experiencing dynamic growth. According to the forecast of the Big Data Association, by 2024 it will reach 300 billion rubles. 62% of Russian companies are already working with Big Data solutions, many of them for more than three years. This indicates good prerequisites for the further development of Big Data in Russia: there is a steady demand for IT products and services, an understanding of the need to develop data science as a science, support for R&D, and organizations conducting their own research. A stack of technology used has been identified, and sufficient experience has been accumulated to predict the economic effect of implementing data science solutions and avoid typical implementation problems.

At the state level, the "National Strategy for the Development of Artificial Intelligence until 2030" dated October 10, 2019 was adopted. It included regulating the interaction between government organizations, non-profit associations, educational and research organizations, and commercial enterprises for the effective development of this area, as well as the popularization of data science as a science. The "Artificial Intelligence" federal project of the "Digital Economy of the Russian Federation" national program (approved by the protocol of the Presidium of the Government Commission on Digital Development, the Use of Information Technologies to Improve the Quality of Life and the Conditions for Conducting Business of August 27, 2020 No. 17) provides for the following targets:

- an increase in the number of companies developing AI solutions that received state support under the "Artificial Intelligence" federal project from 247 companies in 2021 to 1,999 in 2024;
- an increase in the number of AI specialists trained under higher and additional education programs from 1,916 people in 2021 to 4,241 in 2024;
- doubling the AI community between 2021 and 2024;
- 100 percent of federal executive authorities that have approved changes to departmental digital transformation programs and are implementing measures to introduce AI and prepare datasets.

In addition, work is currently in progress to form a registry of software related to artificial intelligence systems in accordance with GOST R 59277–2020, or containing elements related to artificial intelligence systems. Currently, work is already actively underway on national, interstate, and international standardization within the framework of the RUSSOFT non-profit partnership participating in the TK-164 "Artificial Intelligence" Technical Committee for Standardization. To date, a number of measures for state support of innovative developments in the field of big data processing have also been developed.

Big data is by far the most promising direction in the development of the IT industry. Analysts at IDC predict that by 2025, organizations, not individuals, will produce most of the data. As the volume of data increases, so do the requirements for analytical systems: instead of processing data on past events or

periods, systems are increasingly required to provide real-time results and forecasting. This creates the need not only for a complete digital transformation of the enterprise, but also for a change in the attitude to data at all levels, since the completeness and reliability of this data affects how correct the results of the BI analytics system development would be. Thus, BIG DATA as the amount of accumulated heterogeneous information that requires the use of scaling technology for efficient data storage, processing, management, and analysis can be reckoned the conceptual basis of BI systems.

In this article some aspects of BID DATA principles in BI systems are given/The main methodological approach and ideas are presented in articles of V. Zaslavskaja, for example, in [9].

2. BIG Data as a Conceptual Basis for Business Intelligence Systems

The analysis of results of measurement of linear As the volume, diversity, and the importance of data processing for the enterprise grow, it becomes the most crucial task to identify its unique, critical segment, the content of which will have a significant positive impact on the user (client) experience, become the basis for solving complex problems, and help to scale the company. Identifying and maintaining this segment holds great potential and helps to avoid wasting resources on constantly processing the entire volume of data.

The development of BI solutions based on Big Data and machine learning, as well as creating conditions for their use in the segment of small and medium-sized businesses, can have a serious impact on the Russian economy as a whole. The development of domestic software solutions based on predictive business analytics and machine learning, taking into account the specifics of a particular industry and financially accessible to small businesses, would allow many companies to avoid financial losses and save jobs.

In 2022, a significant number of small and medium-sized enterprises closed or suffered financial losses. According to FinExpertiza, from March to June of this year, more than 113.5 thousand commercial enterprises stopped working in Russia, which is 17.5 % more than in the same period last year. The vast majority of closed companies belong to the segment of small and micro business.

The head of the company needs a broad expertise and a holistic vision based on a deep understanding of the internal business processes and potential of the company, knowledge of the industry, market, trends, international situation, and other external factors. The more complete this picture, and the more factors, as well as their influence coefficients and interrelationships, the manager takes into account when making decisions, the more successful the company's strategy. In fact, the leaders of small and especially micro-companies are overloaded with operational activities.

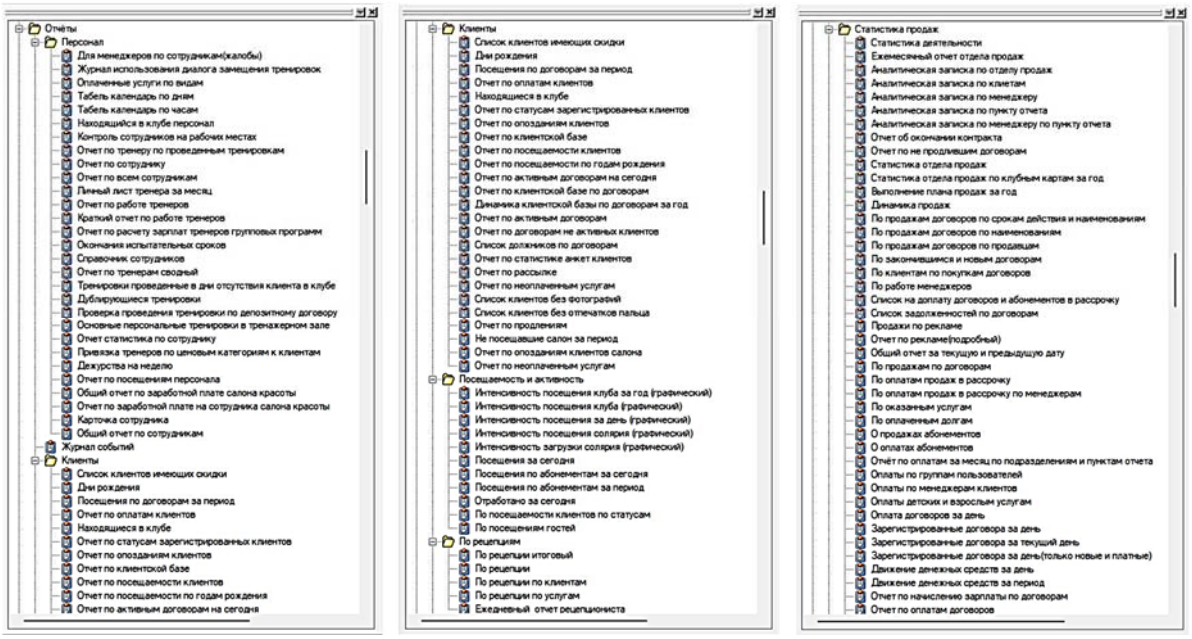


Fig. 1. Standard reports of the management accounting automation system.

Over the past 20 years, small businesses have accumulated significant amounts of data that can be effectively used in business intelligence systems, coupled with open statistics, semi-structured data from social networks, industry aggregate data, and data accumulated by various types of data loggers. State support for the development and certification of BI systems for medium, small, and micro businesses

could help overcome the main problems that become an obstacle to using BIG DATA technologies for small companies: lack of budget and competencies in the team to develop their own solutions, the difficulty of working with a variety of formats and a significant amount of unstructured and semi-structured data, all employees of the company being insufficiently aware of the value and potential of data (according to experts, 13% of respondents-heads of companies voice this problem), and necessary investments in the infrastructure and reorganization of business processes.

Let's consider the practical application of a business intelligence system using the example of one of the leading Russian chain fitness clubs, which, however, belongs to the small business segment according to the classification.

For more than 10 years, fitness clubs and many other service companies have been operating in a highly competitive environment. In order to attract and retain customers, club management is constantly working on pricing policy of memberships, marketing and advertising campaigns, analyzing the target audience, and improving the quality of customer service. The object in question is equipped with a management accounting automation system, a CRM system, and biometric equipment for accessing the club (turnstiles, video surveillance equipment with a face recognition system). To date, there is no ready-made software solution for business analysis in this area, thus the top management of the network has to be guided by expert assessments founded on the basic reports of the automation system.

Conducting in-depth interviews with the top managers of the club made it possible to form a number of questions, answering which would help to organize the work with clients more effectively, form a pricing policy more reasonably in a highly competitive environment, deeply understand business processes, the causes of various phenomena, and patterns, make deliberate management decisions, and scale the business. In order to answer these questions, it is necessary to apply new approaches and business intelligence tools, as it is impossible to answer then with an expert method.

- segmenting the target audience, forming a deep understanding of the processes and a holistic vision of the situation and prospects.

- identifying patterns in the behavior of various categories of customers, distinguishing segments of the target audience using the correlation method, as well as relationships between the time / date of visiting the club and purchasing additional services and any other actions / events recorded in the system, taking into account all the data accumulated in the management accounting automation system, CRM, data from video recorders and turnstiles, semi-structured data from social networks, statistics on the area in which the club is located, the history of weather changes, the schedule of public holidays, major district events, city events, etc.

- Personalizing communication with the client, forming recommendations based on the analysis of their preferences, psychological characteristics, and the accumulated experience of interacting with them.

What is crucial in the matter of visiting the club, and what is the main reason for visiting the club for different target groups?

- What are the common traits of clients who buy a lot of additional services and personal training?

- Which categories of clients go only to the pool or only to the aerobic room?

- Analysis and identification of patterns in customer leaving. Giving recommendations: what can help the client return to the club?

- How many people buy memberships not for themselves?

- How is the distance between the client and the club related to the frequency of visiting the club, renewing the membership, purchasing additional services, and other options (considering that some of the addresses in the database are registered, and not the actual ones).

- When a client is offered a choice of a VIP coach (more expensive) or a regular one, with which of the coaches does the client stay longer? What segment does this client belong to, what traits does he have? Are there any additional patterns?

- Forecasting visits to the club by clients for two weeks ahead, taking into account the forecast of weather conditions, season, holidays, club promotions, etc.

● Marketing and promotions. Analytics, recommendations, and forecasting.

- Recommendations on promotions for identified segments of the target audience

- Analytics of the club accounts in social networks based on generalized information about the target audience, its segments, and recommendations.

- Analysis of the potential client audience in newly built houses within walking distance from the club. Recommendations for promotions based on generalized data from open statistics.

● Optimizing and improving the performance of the sales department

- Analyzing the work of sales managers, identifying patterns and weaknesses, recommendations for improving the efficiency of each employee and the department as a whole.

- Funnel optimization: call, meeting, closing

- What is the optimal time spent by the manager from the first contact to the purchase?

● Abuse detection

● Pricing and membership renewals

- What is the critical cost, and to which it can be increased without the risk of losing a serious percentage of customers.

Identifying reasons and non-obvious connections between the client leaving or not renewing the membership and any other data?

- Predicting the renewal of memberships based on the analysis of the totality of data accumulated and obtained from open sources.

- How many visits per year are enough for a customer to be likely to renew their membership?

● Biometrics, video surveillance and face recognition

- Which areas are visited by which clients (aerobic room, gym, swimming pool, group classes). Reveal patterns between segments of the target audience and visits to certain zones.

Planned results:

- Increasing sales through a deeper understanding of the needs of target audiences, improving personal sales and advertising;
- Reducing the percentage of refusals to renew membership cards, the return of the customers who had previously left;
- Reducing the load on sales managers (including emotional, which is due to personal recommendations);
- Redistribution of the load on different zones of the club at different times and days.

Thus, a business intelligence system based on large amounts of data accumulated over the past 10-15 years by commercial enterprises, taking into account external statistics, is a unique source of information, a competitive advantage, and a valuable asset. The development and distribution of industry-specific BI solutions for companies in the medium, small and micro business segments has allowed tens of thousands of enterprises to avoid financial losses.

4. Conclusion and Discussion

The purpose of the study is to analyze the effectiveness of a predictive business intelligence system by comparing the financial performance of the club and the customer satisfaction index at the start of testing and after one, three, and six months. Criteria for evaluation are as follows:

1. Sales volumes of memberships and additional services
2. Percentage of membership renewals, percentage of the clients leaving, percentage of old customers returning
3. Efficiency of sales managers, assessment of emotional load
4. Optimizing the loading of different club zones at different times of the day and days of the week.

To solve these problems, the following mathematical models can be applied to ensure high accuracy of the result:

- Solving problems of reduced dimension, which allows us to reduce noise and increase the accuracy of the result (Figures 2 and 3)
- Solving the problem of training neural networks, which will make it possible to create a behavior model based on training and test data (Figures 4 and 5)
- Solving the problem of obtaining the most accurate results. Models built on the basis of machine learning algorithms.

Thus, the use of a business intelligence system for data processing is a unique source of information concerning internal business processes, external factors, their impact on the company, as well as a significant competitive advantage. Any such system is based on the concept of Big Data: collecting, storing, protecting, processing, analyzing, and forecasting. The development of BI solutions available for companies in the medium, small, and micro business segments would allow tens of thousands of enterprises to increase sales, optimize internal processes, predict risks, and their managers to determine the company

Funding: The work was done under financial support of the Russian Science Foundation, project No. 22-29-20123 (50% contribution) and the St. Petersburg Science Foundation in accordance with agreement No. 14/2022 dated April 14, 2022 (50 % contribution).

References

[1]. National Standard of the Russian Federation. Information technology. BIG DATA. Overview and dictionary. (ISO / IEC 20546:2019, IDT), *Standartinform,* Moscow, 2021. https://api.bigdata-msu.ru/media/uploads/2020/12/30/gost-r-iso-mek-20547-1-standart-it-eabd-struktura-i-process-primeneniya.pdf.

[2]. Federal Agency for Technical Regulation and Metrology. National Standard of the Russian Federation, GOST R 59277— 2020 Artificial intelligence systems Classification of artificial intelligence systems, *SpshjptFim*, 2021, p. 12.

[3]. Khagurova M. P., Chinazova S. K., Vodozhdokova Z. A., Gisheva S. S., Big Data as The Basic Concept of Creating a Bi-System, *The Scientific Heritage*, No. 54-7, 2020. https://cyberleninka.ru/article/n/big-data-kak-osnovnaya-kontseptsiya-sozdaniya-bi-sistemy (date of address: 09.10.2022).

[4]. Verkhovskaya O. R., Bogatyreva K. A., Dorokhina M. V., Laskova A. K., Shmeleva E., In the National Report 'Global Monitoring of Entrepreneurship' Russia 2021/2022 Graduate School of Management. *St. Petersburg State University*.

[5]. Technologies for working with Big Data: readiness for use and the main barriers, mcs.mail.ru, VK Cloud, 2022 Joint study of VK Cloud and Arenadata.

[6]. Ashinova M. K., Berezinskikh S. A., Chinazova S. K., Optimization of the production business process as a method of production management at the enterprise, *Colloquium Journal*, 19, 43, 2019, Część 6, pp. 51-54.

[7]. Data Age 2025: The Evolution of Data to Life-Critical Don't Focus on Big Data; Focus on the Data That's Big, *David Reinsel John Gantz John Rydning,* April 2017.

[8]. National strategy for the development of artificial intelligence for the period up to 2030. Approved by the Decree of the President of the Russian Federation of October 10, 2019 N 490.

[9]. Zaslavskaja V. L., Methodological aspects of Unsructured data usage in management systems, *Soft Measurements and Computing*, 10, 2022, pp. 80-89.

(S03)

Problems of Using Semi-structured and Unstructured Data in Business Intelligence Systems

Zaslavskaya Veronika L.[1] and Konovalova Vera S.[2]
[1] COO Zello Russia, Deputy Chairman of the 'Artificial Intelligence' Committee of RUSSOFT,
[2] Federal State Autonomous Educational Institution of Higher Education St. Petersburg State Electrotechnical University LETI Named after V.I. Ulyanov (Lenin)
197022, Russia, St. Petersburg, Professora Popova str., 5
Tel.: +79219494979
E-mail: veronika@zello.com

Summary: At the current stage of the development of management systems and information technologies, as well as in the context of the widespread use of software for automating business processes and the standardization of basic analytics, unstructured and semi-structured data are a valuable asset of the company. This is a resource that contains initial data for a deep understanding of both internal processes of the company and external factors influencing it, and thus the potential to improve the efficiency of the enterprise. In particular, the analysis and interpretation of unstructured data can allow the company to increase its attractiveness in a highly competitive personnel market. Another advantage would be an improvement of internal interaction through making management decisions with the use of not only standard reporting data, but also data obtained as a result of processing unstructured data. However, the effectiveness of using unstructured data directly depends on the ability to classify and systematize such data. These specified problems are analyzed by the author in the article using examples of the analysis of internal group negotiations of a commercial enterprise.

Keywords: Unstructured data, Semi-structured data, Data mining.

1. Introduction

Currently, any enterprise is constantly working with a significant amount of information. At the same time, there is an annual increase in both the internal information flow produced by the company and the global volume of data. According to IBM experts, the global volume of data will increase to 175 zettabytes by 2025. Today, a significant part of the data generated by the business processes of enterprises is structured due to the wide distribution and availability of automation systems for management and accounting, sales, financial, statistical, and production reporting. Such systems already contain standardized analytical reports on the basis of which management decisions can be made. Using the same systems and relying on the same analytical reports, companies find themselves on an equal footing, which forces them to reconsider their attitude to different types of data and approaches to their processing. This, in turn, is one of the reasons for the increase in the volume of processed data.

2. Problems of Using Semi-structured and Unstructured Data

According to various studies, unstructured data accounts for 80-85 % of a company's total information. Since the possibilities of processing the volume of internal and external data are limited by the financial budget (and the processing of unstructured data requires special attention and the use of new methods), it is extremely important to determine the level of correlation between the data used and the parameters under study, to differentiate that data into appropriate for use in the study and inappropriate.

At the same time, Business Intelligence (BI) systems mainly work with structured data. Separate systems that include an integration approach, working with semi-structured data types, are most often experimental developments. There is every reason to believe that the development of such systems that can take into account more data, identify non-obvious, hidden patterns or trends will become widespread in the near future .In this article some aspects of usage unstructured data are given/The main methodological approach and ideas are presented in articles of V. Zaslavskaja, for example , in [13].Working with semi-structured and unstructured data is justified by the wide possibilities for the subsequent application of the information received, but is characterized by a number of problems shown in the Fig. 1.

The key point when working with unstructured data is selecting (defining) ontology, i.e. "a set of a scheme for describing a subject area and rules for attributing data to this subject area"[7] . Within the

[7] Grishkovsky A., Integrated processing of unstructured data, Open Systems, DBMS, 06, 2013.
https://www.osp.ru/os/2013/06/13036849

framework of the ontology, its key points are distinguished: entities, attributes and relationships. In working with semi-structured and unstructured data, data mining, natural language processing, and text mining are actively used.

Fig. 2 shows a diagram of working with unstructured data when using this data for business intelligence. - segmenting the target audience, forming a deep understanding of the processes and a holistic vision of the situation and prospects.

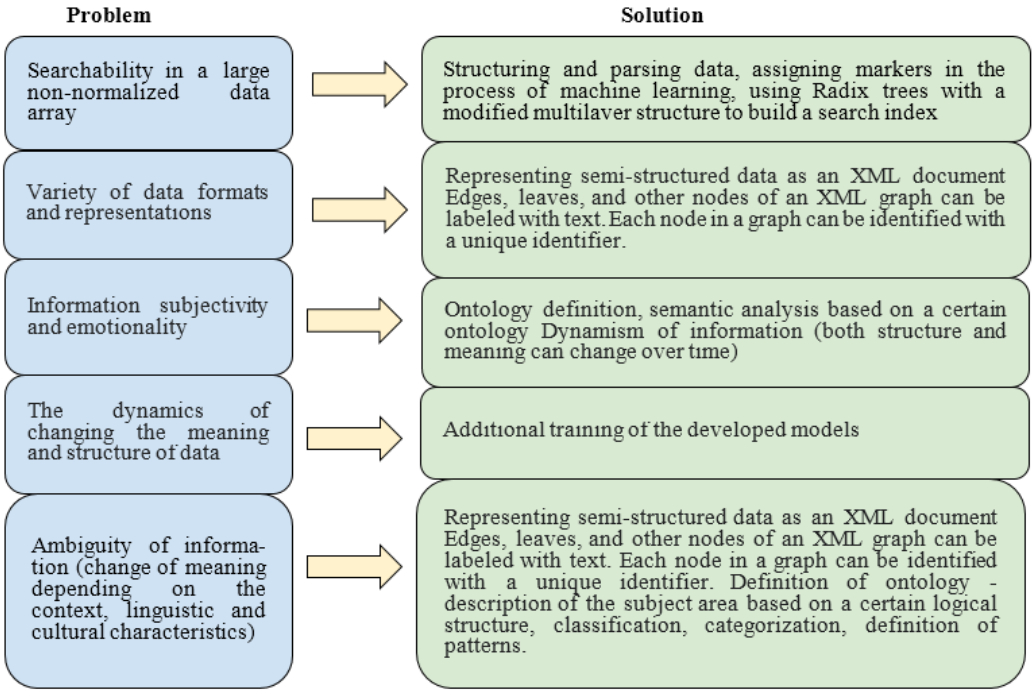


Fig. 1. Problems and possible solutions.

Fig. 2. Scheme of working with unstructured data when using this data for business intelligence.[8]

[8] Grishkovsky A., Integrated processing of unstructured data, Open Systems. DBMS, 06, 2013. https://www.osp.ru/os/2013/06/13036849

Consider the problems and potential of processing semi-structured and unstructured company data on the example of processing internal negotiations via group voice communication channels. According to statistics, the average amount of audio data produced by one user per calendar month is 100 MB. Thus, the volume of data of a company with 10,000 employees obtained only through voice calls will be 12 TB of semi-structured data per year. The structured part of the data that is stored in the database at the time of recording the message contains only information about the sender and recipient (or the channel to which the message was sent, the date and time of sending, the length of the message in milliseconds and its type: image link, text, notification etc., an example would be the audio files of negotiations).

At the same time, the amount of data contained in the message itself and available for subsequent analysis and interpretation significantly exceeds the primary structured part. Approaches and methods to the analysis of a single voice message are shown in Table 1.

The methodology, results, and application of user voice messages analysis in the context of negotiations are shown in Table 2.

Table 1. Voice message analysis.

Data	Possible approaches to obtaining information	Obtained values
The meaning of the message	Speech Recognition Applications	Unstructured text data
Emotional coloring of the message	Speech Emotion Recognition (SER)	List (time since the beginning of the message, the intended emotion)
Number of mentions of certain words	Comparison with a given dictionary (e.g. company name)	List (word, quantity)
Features of speech (loudness, speed of speech, the presence of pauses)	Analysis based on spectral-temporal, cepstral, amplitude-frequency or non-linear dynamics features	List (feature, index)
The presence of extraneous noise	Analysis based on spectral-temporal, cepstral, amplitude-frequency or non-linear dynamics features	List (time since the beginning of the message, noise index)

Table 2. Analysis of the message in the general context.

Ability to define	Possible processing method	Result of data processing	Possible application
Availability and speed of responses to inquiries	Machine learning and neural networks based on spectral-temporal features	Performance index	Evaluation of individual employees' performance
Emotional satisfaction of the employee	Machine learning and neural networks	Index of interaction (communications)	Assessment of the communication level between employees and structural units, identification of the "bottleneck"
Repeat requests and clarifications	Machine learning and neural networks based on spectral-temporal features	Response Completeness Index	Assessment of the qualification level of employees
Context of mentioning company name	Machine learning and neural networks, named entity extraction method	Employer Satisfaction Index	Assessment of employee satisfaction by the employer
Team sustainability	Machine learning and neural networks	The index of satisfaction with the organization staff	Evaluation of the level of satisfaction with the team, compatibility of employees

The next step to obtain even more substantiated results is to combine the dataset of processed (structured) internal negotiations with other datasets of the company that contain information about this user (ERP, CRM, and other systems). In general, the combined datasets will help to make a more accurate classification of company employees, identify patterns in their behavior, and much more.

Examples of practical application based on the identification of employee behavior patterns are as follows:

- Evaluation of the emotional state and satisfaction of employees, their loyalty to the company, forecasting the likelihood of their leaving;
- Forecasting the performance of an employee in another position (choosing the best candidate);
- Assessment of the team's compatibility, conflict prevention, improvement of interaction
- Fraud detection.

4. Conclusion and Discussion

Thus, semi-structured and unstructured data are difficult to process and interpret, but at the same time they are of great value to the business. Analyzing such data can help to significantly increase the objectivity of assessing current business processes and the state of the market, reveal the hidden patterns and make more accurate forecasts. When using interdisciplinary research (for example, as in the case of analyzing the emotional state of company employees based on the audio data of internal negotiations) the company can gain a unique competitive advantage.

Acknowledgements

The work was done under financial support of the Russian Science Foundation, project No. 22-29-20123 (50% contribution) and the St. Petersburg Science Foundation in accordance with agreement No. 14/2022 dated April 14, 2022 (50% contribution).

References

[1]. Burykin A. D., Separate aspects of management in social systems: cybernetic approach, *Economics and management: problems, solutions*, 2017, No. 5, pp. 79-86.

[2]. Grishkovsky A., Integrated processing of unstructured data, *Open Systems. DBMS*, 06, 2013. https://www.osp.ru/os/2013/06/13036849.

[3]. Radachinskaya A. A., Methods for analyzing unstructured data, *Actual Questions of Science*, 44, 2018, pp. 128-130.

[4]. Tomashevskaya V. S., Yakovlev D. A., Methods for processing unstructured data, *Russian Technological Journal*, V. 9., No. 1 (39), 2021, pp. 7-17.

[5]. Syasky D. Yu, Limiting the number of control models in hierarchical systems, in D.Yu. Syasky, A.D. Burykin, Questions of methodology of social sciences and humanities: modern context: collection of scientific papers (ed. E. P. Tkacheva), 2018, pp. 157-160.

[6]. Syasky D. Yu, Uncertainty in the management of complex hierarchical systems, in *Proceedings of the of the XXV Scientific-practical Conference*, Barnaul, 2021, pp. 22-36.

[7]. National standard of the Russian Federation. Information Technology. Big data. Review and dictionary. (ISO / IEC 20546:2019, IDT). *Standartinform,* Moscow, 2021.

[8]. Federal Agency for Technical Regulation and Metrology, National standard of the Russian Federation. GOST R 59277-2020 Artificial intelligence systems. Classification of artificial intelligence systems, *SpshjptFim*, 2021, p. 12.

[9]. Rakovsky, A., Moskvichev, A., and Filchenkov, A., Data augmentation method for the image sentiment analysis, in *Proceedings of the AINL FRUCT: Artificial Intelligence and Natural Language Conference,* 2016, pp. 106-109.

[10]. Federal Agency for Technical Regulation and Metrology. National standard of the Russian Federation. GOST R ISO/IEC 20546-2019 Information technology. LARGE. Overview and vocabulary (ISO/IEC 20546:2019, IDT). *Standartinform*, Moscow, 2020, p. 13.

[11]. Ignatenko A. M., Makarova I. L., Kopyrin A. S., Methods of preparing data for the analysis of semi-structured time series, *Program Systems and Computational Methods*, 4, 2019, https://nbpublish.com/library_read_article.php?id=31797.

[12]. Prokopchina S. V., Bayesian Intelligent Measurements, *Scientific Library*, Moscow, 2021.

[13]. Zaslavskaja V. L. Methodological aspects of Unsructured data usage in management systems, *Soft Measurements and Computing*, 10, 2022, pp. 80-89.

(S04)

Estimability of Effectiveness of Rehabilitation Measures Following Injuries in the Locomotor Apparatus on the Basis of Analysis of Small Samples

Alekseev Vladimir V., Korolev Pavel G., Tsareva Anna V. and Ivanova Ksenia M.
Federal State Autonomous Educational Institution of Higher Education St. Petersburg State Electrotechnical University LETI Named after V.I. Ulyanov (Lenin)
197022, Russia, St. Petersburg, Professora Popova str., 5
Tel.: +7 812 346-44-87, fax: +7 812 346-27-58
E-mail: avtsareva@etu.ru

Summary: Considered are the issues of reliability of the analysis of informative measuring signals obtained by an information-measuring system (IMS) at specified points of the human kinematic image (HKI), that are associated with the control of the state of the locomotor apparatus. We analyze the results of measurements from small samples to determine the reliability of assessment of the quality of post-trauma rehabilitation process.

Keywords: Linear accelerations, Digital twin, Informative parameter, Custom information-measuring system, Mann–Whitney U test.

1. Introduction

The parameters of the gait technique are evaluated by various means for movement recording. Recently, developments that are based on inertial sensors and video systems have gained popularity. Step phases determination is the major problem in gait analysis. However, for example, the foot plantar pressure measurement systems do not have such problem. The "Dia–Sled" hardware-software complex is used to evaluate the dynamics of the pressure distribution between the foot and the supporting surface. The complex has been developed by the St. Petersburg Scientific and Practical Center of Medical-Social Expertise, Prosthetics and Rehabilitation of the Disabled named after G. A. Albrecht Dia-Service LLC and VIT LLC. The complex consists of sensors in the form of insoles including 108 load cells. To evaluate the foot function, ordinary shoes of a subject are used that meet the following requirements: sneakers with a heel up to 1 cm, a soft top, a flexible sole. Measuring insoles should be of the same size as the shoes and the foot of the subject. The data collection and processing module is attached to the subject's waist using a belt included in the complex. The subject is examined for the absence of heavy objects in the pockets and in the hands. The study consists of several steps as follows: measurement under the static condition (3 seconds) where the heel-to-heel interval shall range from 3 to 6 cm, the head shall be positioned straight, the toes shall be on the same level; measurement under the dynamic condition (8 seconds): usual gait where the pressure of the foot on the supporting surface during walking is measured. The results include the following parameters: the plantar pressure distribution, pressure center trajectory and integral load graphs [1, 2].

The "Biokinect" training complex (R&D and manufacturing company LLC "Nevrokor") is designed for training of walking in order to restore the rhythm and time structure of the step cycle, movements in the joints of the lower extremities, proper automatism of muscles, weight bearing of lower extremities. It is possible to conduct a clinical analysis of the motion biomechanics simultaneously with rehabilitation training, using methods of functional electrical stimulation and biofeedback. However, this simulator is not a portable means of assessing rehabilitation measures or the degree of recovery of motor skills. Many companies specializing in medical equipment/training devices produce products designed to evaluate gait techniques but the reliability of the results is still questionable. While the problem has been partially solved in the case of motion video recording systems, the portable wearable systems still have this problem, therefore, it is important, especially in the field of medical and sports rehabilitation, to develop novel custom systems for evaluation of gait training that provide increased reliability of measurement results.

2. Materials and Methods

2.1. IMS Structure

Based on the analysis of the IMSs used to solve the task, a structure was determined that meets the criteria of effectiveness and requirements for the study of the kinematics of human/human limb motion, which was a penalty function-based criterion [3]. The structure combines the measuring, computing and interface components of the system that are designed for the investigation of human locomotions [4]. The structure may have a wired, wireless, accelerometric and complex implementation. An example of the implementation of a wireless system is shown in Fig. 1, where MM is a measuring module, RM is a

radio module, PC is a personal computer/laptop with a wireless radio channel, MD is a mobile device with a wireless radio channel (smartphone, tablet).

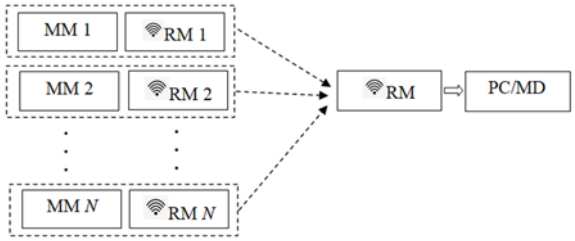


Fig. 1. Block diagram of the information-measuring system.

2.2. Algorithmic Provision

The analysis of results of measurement of linear accelerations in the informative points of the human kinematic image makes it possible to determine the values of accelerations in the gait of a conditionally healthy person. To establish the possibility of determining gait anomalies in patients with various types of pathologies. To formulate proposals for the selection of signal fragments suitable for the formation of diagnostic markers [5, 6].

The developed algorithmic software for the detection of abnormal informative parameters provides for the reliable evaluation of these parameters.

The authors have considered examples of evaluation for the neurosurgical clinical area [7]. Algorithmic software has been developed for a measuring system for monitoring the parameters of a personal kinematic image in order to implement an output system, i.e. evaluation of the condition and analysis of developmental trends for the current situation [8]. The Mann–Whitney U test has been tested for possible use, the limitations of the test capabilities has been confirmed. Rules have been developed to prevent the occurrence of type II errors when applying the Mann–Whitney U test [9, 10].

3. Results

The object of the present study is linear acceleration signals characterizing the lower extremity motion during the gait cycle.

The results (the number of gait cycles) can form both a representative sample (100 readouts) and small samples; the latter is typical, since, even in a hospital, it is rarely possible to find a straight-line section of the appropriate length, not interfered by medical personnel and patients, for the experiment.

As a result of the experiments, we obtained records of measurements of linear accelerations along three axes from sensors placed at informative points in the heel area:

• The X-axis is vertical, directed vertically upwards;

• The Y-axis is sagittal, directed in the direction of movement;

• The Z-axis is transverse, directed perpendicular to the direction of movement.

To study the problem, measurements of motion characteristics were carried out:

• In a conditionally healthy person;

• In post-trauma condition: fracture of the ankle of the fubula, fracture of the posterior edge of the tibia, deltoid ligament rupture, syndesmosis rupture;

• During treatment: sutures were applied to the deltoid ligament, syndesmosis was fixed with two screws, a retentive bandage. Following the removal of the retentive bandage, the metal structure was repositioned after a critical time; further, we observed the onset of syndesmotic ossification leading to limited mobility of the ankle joint.

The measurement results were generated in the form of protocols, where CHP refers to a conditionally healthy person; IRB refers to post-trauma condition and removed retentive bandage; IRB1 is one week of treatment following the removal of the retentive bandage; TFP2 is 2 weeks of treatment following the removal of the retentive bandage; TFP3 is 3 weeks of treatment following the removal of the retentive bandage; TFP4 is 4 weeks of treatment following the removal of the retentive bandage; TFP5 is 5 weeks of treatment following the removal of the retentive bandage; TFP6 is 6 weeks of treatment following the removal of the retentive bandage; MR is following removal of the metal structure that stabilizes the tibia and fibula; MR1 is one week following the removal of the metal structure that stabilizes the tibia and fibula; MR6 is one month and a half following the removal of the metal structure that stabilizes the tibia and fibula.

There has been no fundamental differences in the shape and amplitude of the signals, so we conducted a statistical study of the time parameters of gait, the results are shown in Table 1.

In a conditionally healthy person, two-footed support phase was about 50 milliseconds, with the spread of the duration of the support phase being 35 milliseconds. The values were the same for the right and left limbs.

In a person at the rehabilitation stage following removal of the retentive bandage, but with the screws left, the two-footed support phase in the case of stepping with the right, injured foot was on average 24 milliseconds, with a parameter spread of 40 milliseconds. When stepping with the left foot, the two-footed support phase averaged 90 milliseconds, reached 0.12 second, the spread of the phase duration was 50 milliseconds. It took in average 0.58 second from placing the right injured foot on the ground to placing the left foot on the ground; it took in average 0.53 second from placing the left foot on the ground to placing the right injured foot on the ground, which was 50 milliseconds less. Further, there was observed significant discrepancy between the parameters tpl_right – tpl_left and tpl_left – tpl_right, and complete divergence between the samples (tpl_right is the time of placing the right foot on the ground; tpl_left

is the time of placing the left foot on the ground). Following rehabilitation, the discrepancy between the parameters tpl_ – tpl_left and tpl_left – tpl_right was insignificant. Tables 1, 2 show the comparison of the time parameters of the phase structure of the subject's gait as a function of the rehabilitation period.

Table 1. Comparison of time parameters of the phase structure of gait and their spread.

No.	Parameter	Units	CHP	IRB	MR	MR_6
1.	Mean gait cycle duration	sec	1.196	1.147	1.182	1.134
2.	Spread of the parameter "gait cycle duration" (SGCD)	sec	0.105	0.165	0.125	0.025
3.	Mean support phase duration	sec	0.735	0.714	0.739	0.688
4.	Spread of the parameter "support phase duration" (SSPD)	sec	0.030	0.085	0.030	0.015
5.	Mean heel raise phase duration (MHRD)	sec	0.214	0.190	0.197	0.194
6.	Spread of the parameter "heel raise phase duration" (SHRD)	sec	0.020	0.055	0.040	0.020
7.	Mean swing phase duration (MSPD)	sec	0.458	0.436	0.444	0.451
8.	Spread of the parameter "swing phase duration" (SSPD)	sec	0.025	0.095	0.115	0.035
9.	Support phase/gait cycle ratio	%	61.4	62.2	62.5	60.7
10.	Swing phase/gait cycle ratio	%	38.3	38.0	37.5	39.8

We compared samples, which were taken in the CHP state and post-injury state with the retentive bandage removed from the ankle, characterizing the duration of the gait cycle, the support phase, the swing phase, the duration of foot lift according to the Mann–Whitney U test. This test can verify the hypothesis that two samples come from the same general population H0: F1(x) = F2(x). The hypothesis also includes equality of median values and equality of mean values $\mu 1 = \mu 2$. In the process of calculating the empirical value of the criterion, sets of values of samples X1{x1, … xn1} и X2{x1, … xn2} are put in correspondence with sets of ranks R1{R1, … Rn1} и R2{R1, … Rn2}. It was found that the samples characterizing the duration of the support phase and the foot raise phase diverged insignificantly, the discrepancy between the duration of the gait cycle and the swing phase were in the ambiguity area.

Table 2. Comparison of the spread of measured time parameters as a function of rehabilitation period

Parameter / Period of control	SGCD, sec	SSPD, sec	SHRD, sec	SSPD, sec
CHP	0.105	0.030	0.020	0.025
IRB	0.165	0.085	0.055	0.095
IRB_1	0.088	0.050	0.043	0.050
IRB_2	0.043	0.070	0.060	0.047
IRB_3	0.165	0.085	0.030	0.095
IRB_4	0.110	0.080	0.035	0.080
IRB_5	0.140	0.050	0.030	0.120
IRB_6	0.145	0.097	0.025	0.085
MR	0.125	0.030	0.040	0.115
MR_1	0.045	0.040	0.085	0.025
MR_6	0.025	0.015	0.020	0.035

Furthermore, there was observed a significant overlap of samples, but the samples included results that deviated significantly from the majority. Obviously, these were not because of failures in the measuring system but rather gait anomalies. To identify outliers in a small sample, it is practical to apply a rank comparison approach [10]. When analysing the effectiveness of rehabilitation measures, outliers should be identified not with the aim of excluding them from consideration when determining the significance of the discrepancy between samples, but rather determining disorders in the motion stereotype.

As shown in Fig. 2, small samples characterizing the value of the spread of the durations of the gait cycle and the step phases obtained at various stages of rehabilitation are characterized by the presence of overlaps and outliers. It is incorrect to use mathematical statistics methods to exclude outliers in this case; therefore, an approach based on comparing the ranks of readouts for different samples was applied. According to the test, the discrepancy between the samples in terms of the duration of the gait cycle in CHP and in patients following the removal of the retentive bandage was in the ambiguity area; after excluding the outliers, the discrepancy was insignificant. The discrepancy between the samples of the duration of the gait cycle in CHP and in a patient following MR was insignificant, the discrepancy between the samples of the durations of the gait cycle in CHP and a patient following the completion of rehabilitation was significant.

The Mann–Whitney U test is considered to be resistant to type I errors. When sampling 5-7 readouts, the type II error occurred when the values of the sample readouts diverged, where the readouts of sample 2 exceeded the values of sample 1 in a single readout that had a rank below sample 1. A rule for detecting an outlier from a sample of n readouts to reduce the probability of the type II error:

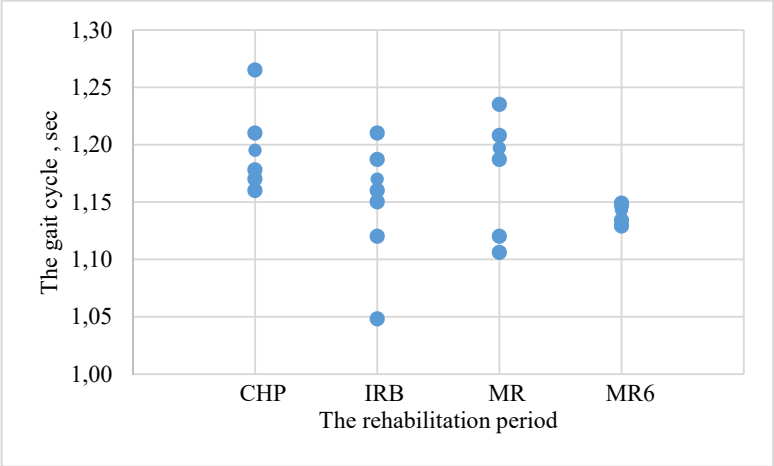


Fig. 2. Samples of the duration of the gait cycle in CHP, IRB, MR and MR_6.

If $\forall\{R_{22}, R_{23}, ..., R_{2n}\} > \{R_{12}, R_{13}, ..., R_{1n}\} \wedge \exists R_{21} < < \{R_{13}, R_{14}, ..., R_{1n}\}$, then exclude the readout $X_{21,}$ which corresponds to the rank $R_{21,}$ and reapply the test.

However, outliers still effect the occurrence of type I errors; accordingly, the test show that the empirical value was in the ambiguity area when comparing the CHP and IRB samples, with 4 readouts overlapped. In general, the problem of detecting outliers to exclude type I errors cannot be solved, but in terms of the present study, the reference values may be not only the absolute values of the parameter characterizing the CHP gait but also the parameter spread value. In this case, the rule for excluding outliers is formulated as follows: $H = R_{\max} - (R_{\max} - R_{\min})$, where H is the calculated threshold level, R_{max}, R_{min} are the maximum and minimum values of the sample of "grouped" values. Readouts that were less than the set threshold level H were excluded from the sample.

4. Conclusion and Discussion

The exclusion of a readout classified as an outlier reduces the likelihood of type I/type II errors but, to evaluate the effectiveness of rehabilitation, it is more important to identify in the sample a readout showing gait anomalies.

By using the rules for excluding outliers for small samples and criteria for evaluating the significance of their discrepancies, it is possible to determine the general trend of changes in the parameter under study; however, a detailed examination of each sample from the point of view of the appearance of anomalies is necessary to adjust the current rehabilitation measures, including the psychoemotional state of the subject.

It should be noted that processing small samples from the point of view of evaluating the parameters of human motion is quite a labour-consuming task, as there are various causes of motion anomalies, and therefore, there is an urgent need for the development of algorithmic support for the analysis of small samples based on large amounts of accumulated data using neural network technologies. These issues will be solved in next stages of the study.

The study was carried out with the financial support by the Russian Science Foundation, project No. 22-29-20123 (50% contribution) and the St. Petersburg Science Foundation in accordance with the agreement dated 14/04/2022 No. 14/2022 (50% contribution).

References

[1]. Smirnova L. M., Yuldashev Z. M., Measuring and information systems for prosthetic and orthopedic industry, *Biotechnosphere*, 2, 20, 2012, pp. 22-25.

[2]. Vedenina A. S., Expanding the functionality of plantopodography by optical scanning, *Biomedical radioelectronics*, 11, 2010, pp. 53-58. Alekseev V. V., Korolyov P. G., Olar V. O., Tsareva A. V., Systems Design for Movement Kinematics Research. Efficiency Criteria. Young Researchers in Electrical and Electronic Engineering (EIConRus), in *Proceedings of the 2017 IEEE Conference of Russian* 1-3 February 2017, pp. 251-253.

[3]. Areshko, E. O., Zabolotskaya, N. K., Korolev, P. G. Movement kinematics research systems. Architectural solutions, in *Proceedings of the XX IEEE International Conference on Soft Computing and Measurements* (SCM'2017), 2017, pp. 593-595.

[4]. V. V., Zhurbilo, P. G., Korolev, P. G., Minchev N. V., Alekseev, Movement kinematics research systems. Analysis of linear accelerations signals, in *Proceedings of the IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering,* 2018, pp. 857-860.

[5]. Korolev, P. G., Romantsova, N. V.,Galimullina, E. E., Akolzin, A. I., Movement kinematics research systems. The measuring experiment organization, in *Proceedings of the IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering,* 2019, 8656718, pp. 553-555.

[6]. Alekseev V. V., Ivanova N. E., Sokolova F. M., Korolev P. G., Measuring system for monitoring the parameters of a kinematic portrait of a person. Part 1.

Control system, *Instrumentation*, No. 9, 2019, pp. 16-24.

[7]. Alekseev V. V., Ivanova N. E., Sokolova F. M., Korolev P. G., Tsareva A. V., Measuring system for monitoring the parameters of a kinematic portrait of a person. Part 2. Output system, *Instruments,* No. 9, 2019, pp. 24-32.

[8]. Korolev P. G., Kostyrya V. A., Cook S. A., Mikus O. A., Evaluation of railway track defects. Criterion research, Collection of reports of the conference on *Soft Computing and Measurements,* May 2018, pp. 85-88.

[9]. Korolev P. G., Organization of the work of measuring instruments with metrological self-control, *Measurement. Monitoring. Management. Control.* 4, 30, 2019, pp. 51-57.

(S05)

Virtual Digital 3D Sensor. Evaluation of the Accuracy of Measuring the Parameters of the Kinematic Portrait of a Person

Vladimir V. Alekseev, Anna V. Tsareva *, Natalia V. Romantsova and Vasilisa A. Baronova
Federal State Autonomous Educational Institution of Higher Education 'Saint Petersburg Electrotechnical University LETI', Russian Federation
E-mail: avtsareva@etu.ru

Summary: The article is devoted to the issue of a formalized description of the uncertainty components of a virtual digital scale of a control system built on the basis of video recorders in order to present it as a measuring instrument. An analysis of the system structure is carried out in order to determine the hardware and methodological components, uncertainty. Here are considered the modes of calibration, certification, control, issues of creating digital twins based on the kinematic portrait of a person for use in rehabilitation systems in medicine and sports. The creation of digital twins of objects, systems, processes is widely used to improve the accuracy of monitoring their state, forecasting, improving the quality and reliability of making management decisions. At the same time, an important aspect is to reduce the uncertainty of the pattern boundaries, which increases the reliability of application of the system and the development of a control action or a prediction result. Uncertainty reduction is provided by means of measurement, with the help of which control procedures are carried out.

Keywords: Virtual 3D scale, Instrumental error, Kinematic portrait of a person, Information and measurement systems, Metrology.

1. Introduction

The rapid development of systems for registering coordinates, determining the parameters and functions of movement in space of individual informative points, objects or processes based on the analysis of video frames has led in many cases (when solving various practical problems) to the need to determine the guaranteed reliability and evaluate the uncertainty of the obtained results. Evaluation of these characteristics will make it possible to introduce developing video recording systems into the rank of measuring tools, which provide the determination of the spatial coordinates of given informative points, the dynamics of their movement, movement characteristics, movement functions, etc. It should be noted that an extremely small number of publications are devoted to the issues of metrological analysis and maintenance of coordinate and motion registration systems. Therefore, an attempt is made below to conduct a metrological analysis of the principles of video recording and determine the functions of its metrological support.

Video recording systems are widely used not only in technical fields, but also in sports and medicine. They allow to create personal systems for monitoring a person's condition at all stages of a human life activity. Especially important are the situations after the traumatic recovery of sportsmen or patients. A personal trajectory of the patient's rehabilitation can be analyzed or formulated on the basis of the generated personal kinematic portraits of a person (KPP), reflecting the various phases of the rehabilitation process: the reference KPP, the working KPP, the current KPP. At the same time, a system of digital twins based on the KPP is created. Obviously, in this case, the metrological support of registration systems becomes essentially important.

As shown by the analysis of patent and scientific and technical documentation, modern trends in the development of the concept of constructing systems for determining the coordinates of an object to improve the accuracy of measurements in the field of kinematics of motion, robotic systems, measurement of object parameters suggest: the use of satellite navigation technologies (GPS, GLONASS), radio tags and ultrasonic sensors in addition to video cameras for real-time operation [1-8]; the use of additional sensors, for example, optical or ultrasonic, which are used in marker analysis [9-15]; development and expansion of the implementation of motion video recording systems for measuring the parameters of motion objects with a given accuracy [13, 15]; development of forecasting methods and methods for calculating the trend of changes in the parameters of the movement of an object in time [5-7].

During the last 6 years there were appeared registration systems that have good adaptability to the type of positioning object. These systems use satellite navigation technologies and algorithms for simultaneous processing of data received not only from video cameras, but also from a variety of additional sensors.

Fig. 1 shows the results of the analysis of patent and scientific and technical documentation, reflecting modern trends in the development of systems for registering object coordinates.

It should be noted that the number of selected patents is increasing. This shows the growing popularity of the field of technology related to the positioning of objects, determining the parameters and

coordinates of objects. In addition, it should be noted that there are not so many selected patents. At the same time, there are practically no examples of carrying out a full-fledged metrological analysis of the tools and algorithms used, which means that the topic is relevant, but so far little has been studied.

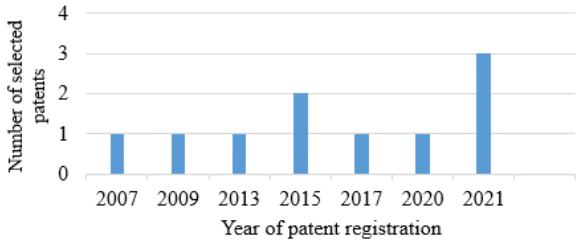


Fig. 1. Dependence of the number of selected patents on the year of their publication.

The conducted patent search allows us to state that there is a need for a classical metrological approach when describing modern systems for registering coordinates, determining parameters and functions of movement in space, implemented on the basis of video recorders, and approach the issue of analyzing such systems from the position of a measuring instrument.

From the point of view of certification of equipment as a measuring instrument, it is necessary to have information on the maximum numerical assessment of the components of the uncertainty of the measurement result, i.e. the accuracy class is determined by the maximum value of the deviation of the measurement result from the true value.

This article for the first time presents a formalized description of the error components of the virtual scale of the video system from the point of view of the measuring instrument. Systems built on the basis of video recorders are means for registering coordinates, determining parameters and functions of motion in space. They fix the coordinates of the informative points of the object at a certain point in time relative to a given reference point, which is set as the origin of the coordinate grid along the three axes. In this case, a linear three-dimensional scale (within a given accuracy) is formed in a given space (relative to the reference point), which provides the measurement of the values of certain parameters or functions in a space with a known value of uncertainty.

The formed three-dimensional space has the specified metrological characteristics and if they are defined (the value of the uncertainty of the measurement results is determined), this system can be classified as a measuring one, and let us hereinafter call the formed linear three-dimensional space **Virtual Digital Three-Dimensional Sensor (VDTDS)**.

Below we will consider the relevance of creation of systems based on the VDTDS for sports and medicine, and also analyze the instrumental and methodological components of the uncertainty of the measurement results of these systems.

1.1. The Role of a Virtual Digital 3D Sensor in the Formation of Digital Twins based on the CPP

The creation of digital twins of objects, systems, processes is widely used to improve the accuracy of monitoring their condition, forecasting, improving the quality and reliability of making management decisions. This technology has also covered the fields of medicine and sports. It is impossible to create a universal digital model of a person. Therefore, to solve specific problems in various fields, it is reasonable to create digital twins of a person [16-21]. For example, to assess the physical and psycho-emotional state of a person after an injury or operation, a kinematic portrait of a person can be used [22-24].

The kinematic portrait of a person (KPH) is a fixed set of controlled (measured) parameters and characteristics of the movement of given points of the human body in space [25]. For systems implemented on the basis of inertial sensors, for example, the Xens, Biokinect systems, measurements of linear acceleration in space along three axes and the speed of rotation around each of the three axes are carried out at each point. Thus, controlled parameters provide the possibility of obtaining a large number of measured signal values (direct measurements), as well as determining a large number of controlled characteristics determined on their basis (indirect, joint and cumulative measurements).

The system for evaluation of the state of a person according to the characteristics of his KPP is based on large amounts of practical (experimental) data for each area and requires the following tasks to be solved:

Creation of knowledge bases in the main clinical (diagnostic) areas (analysis of the characteristics of the KPP, determination of informative parameters and algorithms for their analysis, knowledge about the patterns of the rehabilitation process).

Creation of a control system (determination of points, structure of technical means, software), control and measuring system (methods of measurement, data processing and analysis).

Creation of an inference system - the formation of an assessment of the state and analysis of the development trend of the current situation.

It follows from medical practice that the same biophysical characteristics of a person can indicate different effects of his condition for different genders, age groups, weight and size parameters. Therefore, when designing a database (knowledge base), first of all, it is necessary to determine the functional and physical structure of human characteristics, with a system of informative parameters and characteristics. The operation of the inference system is based on the results of monitoring their accuracy and reliability. That is why great attention is paid to the creation of control systems.

Recently, complexes of video recorders have appeared in medical practice and in the rehabilitation practice of sportsmen, and this gives possibility to control the parameters of the movement of informative points in space [26, 27]. However, if these complexes

are used as measuring instruments, it becomes necessary to certify them as measuring instruments, and to solve this problem, it is essential to analyze the measurement method used, the sources of error and the method of certification.

Let us imagine a complex of video recorders as a measuring instrument (MI), the result of which is a virtual digital three-dimensional sensor (VDTDS).

Let us consider its physical principle, organization, technical and metrological characteristics, issues of assessing the instrumental and methodological components of the error - certification, as well as possible ways of its application for creating a database of digital twins based on the KPP.

2. Materials and Methods

The hardware and software measurement instrument VDTDS is a certain space controlled by a system of video recorders located according to a specific scheme and providing the specified metrological characteristics for determining the coordinates of an informative controlled point, which is located inside of this space, with help of software for a PC, which is the core of the virtual measurement instrument (VMI).

2.1. Physical and Mathematical Basis for VDTDS

From the above definition, it can be seen that the basis of VDTDS is a 3-dimensional space, the formation of which determines the metrological characteristics of VMI.

Controlled space and layout of recorders

The size and location of the controlled space is determined by the layout of video recorders (VR), which are installed along the perimeter relative to the center of the formed three-dimensional scale (TDS). Locations of VR installation are determined by the coordinates (see Fig. 2): VR1(X1, Y1, Z1), VR2(X2, Y2, Z2), etc.

TDS - virtual scale (VS), is formed relative to the base coordinate (X0, Y0, Z0), which can be located in any reasonable place and is specified in the TOR for the creation of the VDTDS (see Fig. 2). The size of the TDS is determined by the metrological characteristics of the scale, the values of which must satisfy the required accuracy.

The size of TDS is defined as:

$$TDS = \{X_{vs}(xmin \ldots xmax); \ldots$$
$$\ldots Y_{vs}(ymin \ldots ymax); Z_{vs}(zmin \ldots zmax)\} \quad (1)$$

System of video recorders

At the given moment of time each video recorder forms a frame that displays the entire control space (3-dimensional analysis space is represented as an information frame). At the same time, an informative point (a series of points) is displayed on each frame. The registered frames are the basis for the implementation of the main operation modes of the software and algorithms: "Calibration", "Certification", "Measurement". To implement these modes, frames from video recorders are transferred to a PC (see Fig. 3) or any other computing module, where linearization algorithms for the filmed non-linear space are applied.

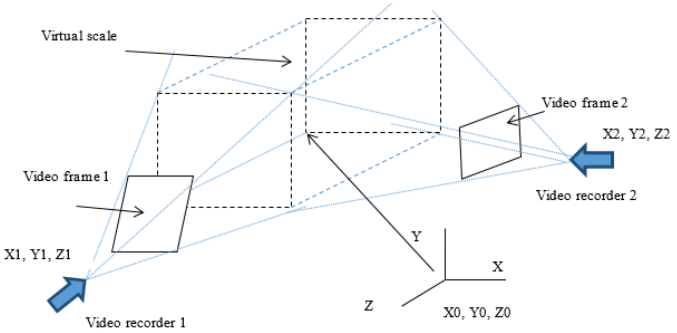


Fig. 2. Controlled space and layout of video recorders.

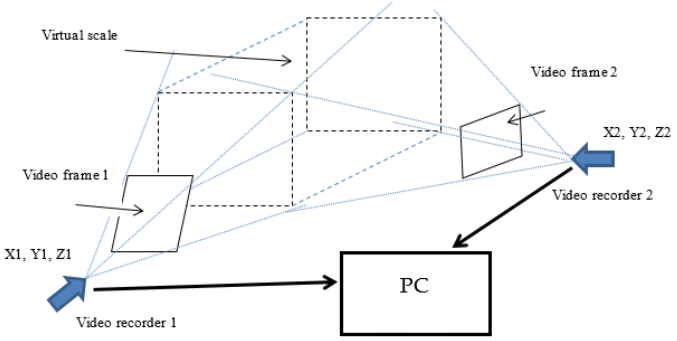


Fig. 3. Scheme of frame transmission from video recorders.

3. Results

3.1. Calibration

The purpose of calibration is to form a linear three-dimensional space for a specific layout of video recorders relative to a given base point (reference point).

Source data:

– the VR frame has a non-linear virtual scale (VR) space display function, which depends on the coordinates of the base (reference) point and the coordinates of the VR placement.

$$Vvr^{xyz} = f(X0, Y0, Z0, Xvr, Yvr, Zvr) \quad (2)$$

therefore, the error of determining the coordinates is also non-linear function

$$\Delta Vvr^{xyz}\{\Delta Xvr, \Delta Yvr, \Delta Zvr\} = \\ = f(X0, Y0, Z0, Xvr, Yvr, Zvr) \quad (3)$$

– Several frames shot by several VRs are registered at the same time.

Calibration - algorithmic transformation of the non-linear space of the VS into a linear one (conversion to a linear metric) based on the results of shooting with several video cameras.

Algorithms: a system of non-linear analytical equations for approximating the results of registration of several VRs; training of a multilayer neural network according to several VR data using exemplary tools (intellectual part of the system).

Operating modes: statics - setting the quantization step; dynamics - setting the sampling step.

Purpose: the formation of a variety of points of the reference linear 3-dimensional virtual scale.

$$Vref^{xyz} = \{X_{ref}, Y_{ref}, Z_{ref}\} = \\ = F(X0, Y0, Z0, VR1\{x^1, y^1, z^1\}, .. \\ VR2\{x^2, y^2, z^2\}, \dots VRn\{x^n, y^n, z^n\}), \quad (4)$$

where x ϵ (x_{min}, x_{max}), y ϵ (y_{min}, y_{max}), z ϵ (z_{min}, z_{max}), F is the calibration algorithm.

An example of the implementation of the calibration algorithm:

1. Resolving power of VR – 0.05 mm;
2. Calibration error – 1.0 mm;
3. Size of VS – 2.0m × 2.0m × 3.0 m;
4. Number of scale points – $2 \cdot 10^3 \times 2 \cdot 10^3 \times 3 \cdot 10^3 = 12 \cdot 10^9$.

As a result of calibration, a reference 3-dimensional set of cells (Fig. 4) of a virtual digital scale is formed and it sets the value of the coordinates of the informative point of the controlled object with a given accuracy ΔX, ΔY, ΔZ (in the example, the calibration error is 1.0 mm).

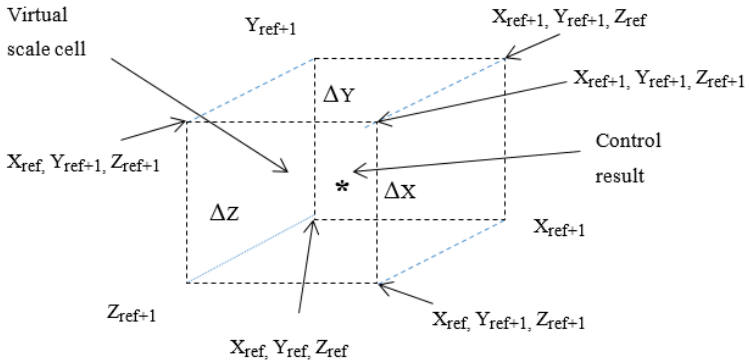


Fig. 4. Structure of the reference cell of VDTMS.

The entire VS space is divided into a set of cells that are positioned relative to the origin of the scale X0, Y0, Z0. Each cell defines a set of reference points {(X_{ref}, Y_{ref}, Z_{ref}), (X_{ref}, Y_{ref+1}, Z_{ref+1}), …., (X_{ref+1}, Y_{ref+1}, Z_{ref+1})}, and their coordinates are the values of VS. These values are determined during the calibration process. Also, these values are assigned to the measurement result during the measurement experiment.

If the result of control (the value of the coordinates of the informative point) falls within the spatial coordinates of the reference cell (in Fig. 4 that is the point marked with an asterisk - *), the measurement result is assigned the value of the coordinates of the vertex of this cell, for which the error will have a minimum value:

$$V^{*xyz} = \{X^*, Y^*, Z^*\} = \{(X_{ref}, Y_{ref}, Z_{ref}) .. \\ if(\Delta = (X^*, Y^*, Z^*) - (X_{ref}, Y_{ref}, Z_{ref}) .. \\ = min) \cup (X_{ref}, Y_{ref+1}, Z_{ref+1}) \\ if(\Delta = (X^*, Y^*, Z^*) - (X_{ref}, Y_{ref+1}, Z_{ref+1}) .. \\ = min) \dots (X_{ref+1}, Y_{ref+1}, Z_{ref+1}) .. \\ if(\Delta = (X^*, Y^*, Z^*) - (X_{ref+1}, Y_{ref+1}, Z_{ref+1}) \\ = min)\}$$

In this case, the methodological error of measuring the value of the coordinates will not exceed 0.5 mm, and the measurement error of finding the informative point in space is ΔL = 0,5 × 30,5.

Let's consider the instrumental component of the error associated with the technical characteristics of the VR.

For example, let's take the technical characteristics of the Miqus M5 camera used in the Qualisys motion capture system:

- Resolution 2048 × 2048,
- Maximum capture distance (with 16mm marker) is 15 m. This is the maximum distance at which the marker covers 2×2 pixels at 200 frames per second,
- Standard lens (field of view) – 49 × 49 °.

The scheme of shooting scale space is shown in Fig. 5.

From the given geometry of the VS space shooting scheme and the technical characteristics of the VR, it follows that recorders of this class provide the following minimum errors at various distances from the optical focus of the camera:

Δ = 6.5 mm, at a distance of 15 m;
Δ = 4, mm, at a distance of 10 m;
Δ = 2, mm, at a distance of 5,0 m;
Δ = 1, mm, at a distance of 2,5 m.

Thus, as a result of calibration, a VS is formed, and it's metrological characteristics are determined by the methodological and instrumental components of the error of the developed algorithmic software and the hardware used, the layout of the VR, the location and size of the VS. Intelligent algorithms for the linearization of three-dimensional space and algorithms for determining (measuring) the coordinates of an informative point are implemented in a PC (see Fig. 3). The results of measuring the coordinates of an informative point with a known value of uncertainty are provided to a specialist or transferred to the system. That is, as a result, a hardware-software measurement tool was created - a virtual three-dimensional digital sensor (VDTDS).

3.2. Certification

As a result of the calibration we obtain a VDTDS, which provides a measurement of the coordinates of an informative point located in the area of the scale space relative to its zero. Measurements must be carried out with a given accuracy ΔX, ΔY, ΔZ, ΔL. To confirm the value of accuracy, it is necessary to carry out certification.

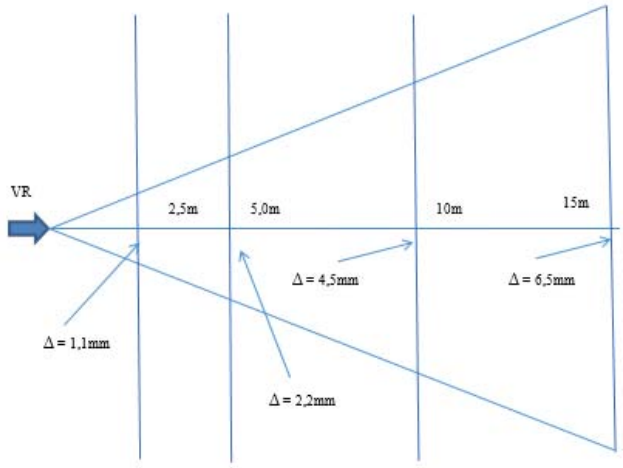


Fig. 5. Scheme of shooting scale space.

Certification is carried out using the working standard "Cane" (RET) (Fig. 6), the length of which is determined with an accuracy of an order of magnitude higher than the accuracy of the used VDTDS.

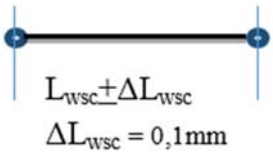


Fig. 6. The working standard "Cane".

1. Single measurements of the value of the length of the working standard in order to determine the maximum error of the VDTDS.

The result of measuring the WSC length value is obtained by implementing the algorithm for measuring the coordinates of the WSC start Lb*{Xb*, Yb*, Zb*} and the WSC end Le*{Xe*, Ye*, Ze*}. Since the VDTDS is a digital homogeneous space, the errors in measuring the coordinates of the beginning and end of the WSC are of a random nature and are uniformly distributed in the range of +0,5 × ΔX*, +0,5 × ΔY*, +0,5 × ΔZ*. In this example, the instrumental error is 1.1 mm at a distance of 5.0 m in the center of the virtual digital scale.

The result of measuring the WSC length is determined by the values of the results of measuring the coordinates of its beginning and end:

$$L^*_{wsc} = F(L^*_b\{X^*_b, Y^*_b, Z^*_b\}, L^*_e\{X^*_e, Y^*_e, Z^*_e\}) \quad (6)$$

The resulting total error in measurement of the coordinates +ΔL* is also random and it is described by the triangular probability density distribution law. In the considered example, the value of the instrumental

error of the scale varies in the range $+\Delta Ls = 2.0 \times 0.5 \times 30.5$.

During the certification process, measurements of the length of the WSC are made at the VDTDS points which reflect its structure (see Fig. 7).

For obtaining a reliable result, it is necessary, on the basis of metrological analysis, to develop a methodology that determines the measurement sites and the position of the WSC.

It is obvious that the measurement results have the minimum error if the informative point is located on the optical axis of the VR survey or in close proximity to it, and the maximum error is at the edges of the scale space or in the proximity to its tops, since the maximum nonlinear distortion of the metrics of space can be observed in these areas.

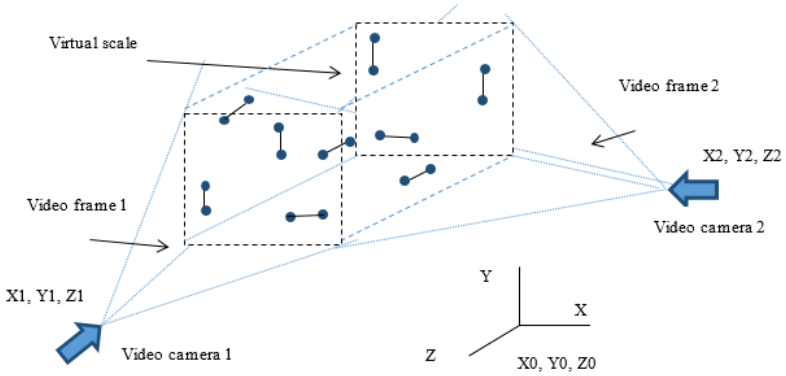


Fig. 7. Controlled space. Layout of WSC.

The WSC length measurement error can be defined as the difference between the measurement result using the VDTDS and the true value

$$\Delta L^*_{wsc} = L^*_{wsc} - L_{wsc} \qquad (7)$$

If the measurement result has an error less than the given scale error – $\Delta L_{wsc}^* \leq \Delta L_s$, the scale meets the requirements for accuracy and it can be assigned the appropriate accuracy class.

If not – $\Delta L_{wsc}^* > \Delta L_s$, it is necessary to calibrate it and confirm the requirements. If this is not possible, it is necessary to change the structure (hardware scheme) and algorithmic software function of the VDTDS in order to achieve the desired result.

2. Multiple measurements. Determination of the systematic component of the error, whether it is of methodological or instrumental nature, or resulting from the impact of influencing factors.

If multiple tests show that the mathematical expectation of the WSC length measurement error ΔLwsc* tends to “0”, that means that there is no systematic component.

Otherwise, it is necessary to specify that component and include it into the assessment of the uncertainty of the results of measurement with the VDTDS. At the same time, the total error must meet the requirements of the accuracy class assigned to the certified scale.

3.3 Measurements

We carried out experiments to measure the reference object and calculate the errors and uncertainties of the measurement result. The reference object is a working standard "Cane" of known length, and on its edges markers are fixed. The measurement is carried out using Miqus M5 cameras and specialized software. The cameras cover the entire measurement space, and each marker is visible to at least two cameras.

The location of the cameras is shown in Fig. 8. Also Fig. 8 shows the distances from a certain point of the boundary of the working space to two cameras that simultaneously record the marker located at this point.

The results of calculating the error in the length of the working standard, depending on the coordinates of one of the markers are given below. Fig. 9 shows the dependence of the error value of the working standard length on the values of the coordinates along the X and Y axes. Fig. 10 shows the dependence of the error on the coordinates along the X, Y, Z axes, the color shows the number of values of the error calculation results for a given coordinate.

In the experiments carried out in the working space shown in Fig. 8, there was a person who moved holding the working standard "Cane" at the level of the knees, hips and chest. From Figs. 9-10 it can be seen that the maximum error in the length of the working standard is achieved at the edges of the space along the X axis, which is explained by the location of the cameras and the boundaries of the working space. Out of the location of cameras (Fig. 8), when a person turns at the edge of the working space along the X axis, the view of several cameras to the markers of the working standard is obstructed. When the view of the nearest cameras is blocked, the markers are fixed by two cameras VR1 and VR2, installed at a distance of 7271 mm and 6943 mm, which is shown in Fig. 8. VR1 and VR2 are selected, because the total distance to the fixed marker is minimal, while VR3 will not affect the

error of the measurement result, since it is located farther from the fixed marker. The distances from VR1 and VR2 are in the range of 5 ÷ 10 m, which gives the measurement error of the marker coordinates Δ from 2.2 mm to 4.5 mm according to Fig. 5. Since the definition of the length of the working standard is based on measuring the coordinates of two markers, then in the worst case, the length measurement error will be 9mm. According to the figures below and Table 1, the error in measuring the length of the working standard is not more than 6 mm, which confirms the calculations.

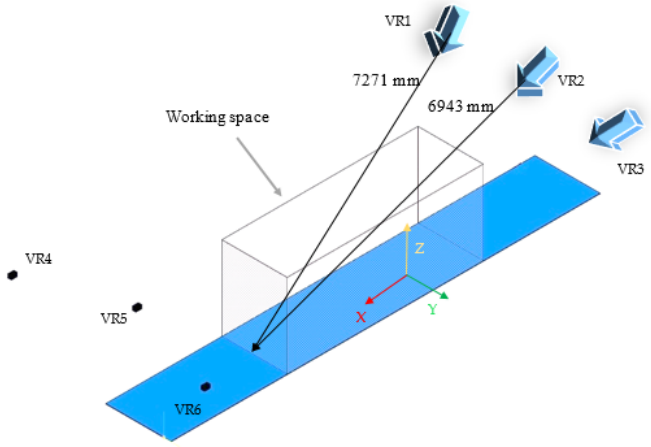


Fig. 8. Location of cameras relative to the workspace.

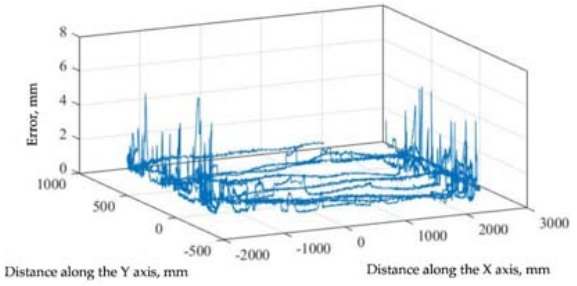


Fig. 9. Error distribution relating to the working space.

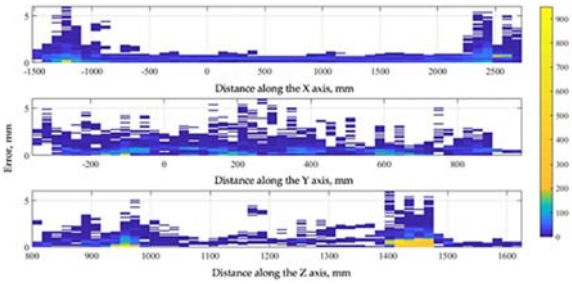


Fig. 10. Error distribution along each axis.

Figs. 11-12 show the results of the distribution of the error in the length of the working standard when measuring the coordinates of the markers in the corners of the working space, while view of the cameras do not overlap. In this case, the value of the error in the length of the working standard was experimentally determined, which is not more than 3.5 mm. The distance from the corners of the working space to the nearest camera is about 3 m, and according to Fig. 5, in the range from 2.5 m to 5 m, the error in determining the marker coordinate is from 1.1 to 2.2 mm. Thus, the error in measuring the length of the working standard in the worst case will be 4.4 mm, which does not contradict the experimental data.

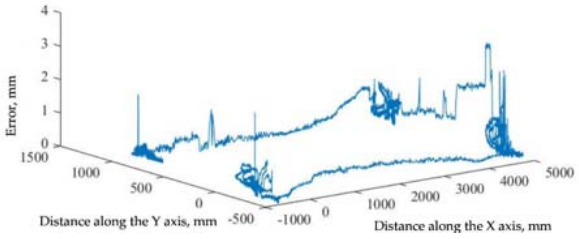


Fig. 11. Error distribution relating to the working space.

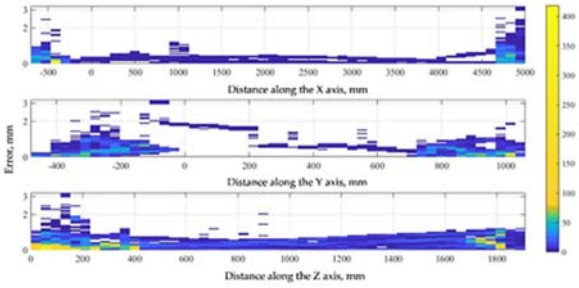


Fig. 12. Error distribution along each axis.

Figs. 13-14 show the results of the experiment for determining the error in measuring the length of the working standard during rotational movement in the entire working space, while the markers of the working

standard were in the field of view of camera lens. As can be seen from Figs. 13-14, the value of the error in the length of the working standard is distributed rather evenly in the space and does not exceed 2.5 mm, because the markers are fixed by cameras, the distance to which is less than 5 m. according to Fig. 5, in the range from 2.5 m to 5 m, the error in determining the marker coordinate is from 1.1 to 2.2 mm. Thus, the error in measuring the length of the working standard in the worst case will be 4.4 mm, which does not contradict the experimental data.

Table 1 shows the results of calculating the errors in measuring the length of the working standard for the experiments described above. The calculation of errors was carried out according to the expressions presented below. Calculation of the maximum absolute error:

$$\Delta L = \left| L - L_{spl} \right|, \quad (8)$$

where ΔL is the error of the result of measuring the length of the working standard, L_{spl} is the length of the working standard.

Average error calculation:

$$\Delta L_{avg} = \frac{\sum_{i=1}^{N} |L - L_{spl}|}{N}, \quad (9)$$

where N is the number of measurements.

The uncertainty estimate was calculated according to the expression:

$$S_L = \sqrt{\frac{\sum_{i=1}^{N} \left(L - L_{\text{обр}}\right)^2}{N}} \quad (10)$$

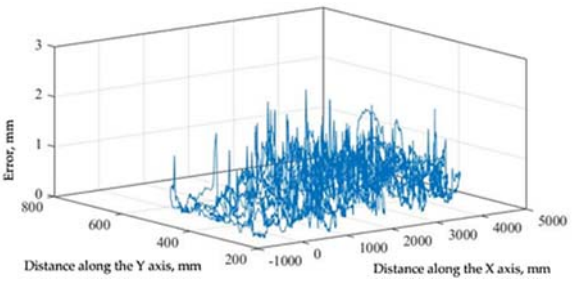


Fig. 13. Error distribution relating to the working space.

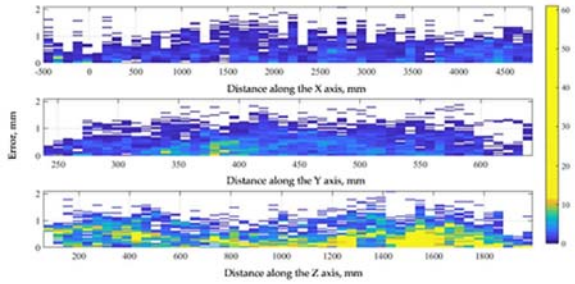


Fig. 14. Error distribution along each axis.

Table 1. Results of calculation of errors and uncertainties when measuring the length of the working standard

Experiment number	ΔL, mm	$\Delta L_{avg.}$, mm	S_L, mm	Notes
1.	6.0	0.40	0.3	Calculation of the error and uncertainty of the length of the working standard when measuring the coordinates of markers during the movement of a person
2.	3.2	-0.20	0.5	Calculation of the error and uncertainty of the length of the working standard when measuring the coordinates of markers in the corners of the working space
3.	2.1	0.13	0.5	Calculation of the error and uncertainty of measurements of the length of the working standard during rotational movement in the entire working space

4. Discussion and Conclusions

According to the results of the experiments, it can be noted that the calculated values of the errors in measuring the length of the working standard do not exceed the value of the error estimate in terms of the analysis of the instrumental component. In order to assess the methodological component of the error, it is customary to carry out simulation modeling. This procedure can be planned for the next stage of the work.

When creating digital twins of objects, processes and systems, an important aspect is to reduce the uncertainty of pattern boundaries, which increases the reliability of the system application and the generation of a control action or a prediction result. It is advisable to reduce uncertainty on the side of the measuring instrument used for caring out measurement procedures.

In this paper, the system for controlling the coordinates of an informative point (a set of points) and motion functions for creation of digital twins, is presented as a means of measuring. The article analyzes the main stages of the creation and application of the VDTDS in order to determine the instrumental and methodological errors and the uncertainty of the control results. These modes are: certification, calibration, measurements. The obtained results are important, because more and more researchers and engineers in this field are trying to

increase the accuracy of measuring system data, while the confirmation of the measurement result error is carried out only according to experimental data, which evidently is a necessary condition, but insufficient from the point of view of the subsequent procedure for metrological certification of equipment. Equipment certification involves a set of tests, and the result of these tests serves as confirmation of the accuracy class of the measuring instrument. At the same time, developers, engineers and researchers need to evaluate the components of the errors and the uncertainty of the measurement results in order to correctly assess the declared accuracy class.

Acknowledgements

The work was done under financial support of the Russian Science Foundation, project No. 22-29-20123 (50% contribution) and the St. Petersburg Science Foundation in accordance with agreement No. 14/2022 dated April 14, 2022 (50 % contribution).

References

[1]. RealTrac. Available online: https://real-trac.com/ru/company/ (accessed on 13 January 2022).
[2]. RTLS. Available online: http://www.rtlsnet.ru/technology/ (accessed on 13 January 2022).
[3]. MERUSOFT. Available online: https://merusoft.ru/en/rtls-positioning-technologies/ (accessed on 13 January 2022).
[4]. Quuppa. Available online: https://www.quuppa.com/understanding-real-time-location-technology/ (accessed on 13 January 2022).
[5]. Blyankinshtein I. M. et al., The method of registering the dynamics of the movement of an object, *Patent RU 0002551765 C1*, Application №2005134294/28, registration date 10.04.2008.
[6]. Ma Zhimeng, Wan An et al., Binocular camera and mechanical arm coordinate system registration method, *Patent CN112815848* (A). Application № CN202110140362, registration date 18.05.2021.
[7]. Shan Zhongde, Wang Jun, *et al.,* Method for dynamically measuring deformation of large-size rotating body mold, *Patent CN112991270* (A). Application № CN202110181635, registration date 18.06.2021.
[8]. Johnson, Philip M. (Durham, N. H.). Non-contact coordinate measuring machine using hybrid cyclic binary code structured light, *Patent US 10429183*, Application № 15/711,238 September 21, 2017, registration date 22.03.2018.
[9]. Faro Technologies. Available online: https://www.faro.com/ (accessed on 15 January 2022).
[10]. Markerless Motion Capture. Available online: https://metamotion.com/software/iPiSoft/Markerless-Motion-Capture_iPiSoft-iPi_Studio.html (accessed on 15 January 2022).
[11]. QUALISYS Available online: https://www.qualisys.com/ (accessed on 15 January 2022).
[12]. Heidemann Rolf, Three-dimensional coordinate measuring instrument coupled to a camera having a diffractive optical element, *Patent US2021180940,* Application US202017025553 20200918, US201962948590P 20191216, registration date 17.06.2021.
[13]. Lankalapall, Kishore, Steffey et al., Coordinate Measurement System with Auxiliary Axis, *Patent US10969760* (A1). Application №16/364,474 March 26, 2019, registration date 17.10.2019.
[14]. Podgornov V.A. et al., Method of determining position of object in space and panoramic apparatus for implementing said method, *Patent RU 2009 103683 A.*, Application 2009103683/28, 04.02.2009, registration date 10.08.2010.
[15]. Gerasimov S. I., Totyshev K. V. Method for determining the spatial coordinates of a moving test object in the form of a body of revolution with known geometric parameters, *Patent RU 2013 125 514 A,* Application 2013125514/28, 31.05.2013, registration date 10.12.2014.
[16]. Lee, A., Kim, J., Jang, I. Movable Dynamic Data Detection and Visualization for Digital Twin City, in *Proceedings of the 2020 IEEE International Conference on Consumer Electronics - Asia (ICCE-Asia),* Seoul, Korea (South), 09 December 2020.
[17]. Bochkovskiy, A., Wang, C. Y., Liao, H. Y. M., YOLOv4: Optimal Speed and Accuracy of Object Detection, arXiv 2020, abs/2004.10934.
[8]. Ssin, S., Cho, H., Woo, W., GeoVCM: Virtual Urban Digital Twin System Augmenting Virtual and Real Geo-spacial Data, in *Proceedings of the 2021 IEEE International Conference on Consumer Electronics (ICCE),* Las Vegas, NV, USA, 10-12 January 2021.
[19]. Gomerova, A., Volkov, A., Muratchaev, S., Lukmanova O., Afonin, I. Digital Twins for Students: Approaches, Advantages and Novelty, in *Proceedings of the 2021 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus),* St. Petersburg, Moscow, Russia, 26 January 2021.
[20]. Mendi, A. F., Erol T., Dogan, D. Digital Twin in the Military Field, *IEEE Internet Computing,* Vol. 26 Issue 5, 2022, pp. 33-40.
[21]. Alam, K. M., Saddik, A. El. C2PS: A digital twin architecture reference model for the cloud-based cyber-physical systems, *IEEE Access,* 5, 2017, pp. 2050–2062.
[22]. Donghan, W., Wei, T., Xiangyu, D., Yan, L. Applications and Analysis of Digital Twin in Prognostic and Health Management, *in Proceedings of the 2021 IEEE 11th International Conference on Electronics Information and Emergency Communication (ICEIEC)*, Beijing, China, 18-20 June 2021, pp. 200-203.

[23]. Alekseev, V.V., Ivanova, N.E., Ivanov, A.Y. et al., Assessment of the Time Characteristics of Walking Techniques Using a Mobile Information and Measurement System. *Biomed Eng,* 2021, 55, pp. 189–193.

[24]. Sokolova, F.M., Tsareva, A.V., Yakovleva, O.A. Problems of evaluation of the effectiveness of walking technique recovery of neurosurgical natients, *Uchenye zapiski universiteta imeni P.F. Lesgafta,* 9, 187, 2020, pp. 354-361.

[25]. Alekseev, V. V., Ivanova, N. E., Korolev, P. G. Use of micromechanical sensors to control the parameters of a kinematic portrait of a person. *Devises,* 2017, 7, pp. 6-15.

[26]. Vicon. Available online: https://www.vicon.com/hardware/ (accessed on 15 March 2022).

(S01)

Intelligent Measurements in the Era of Digitalization

Prokopchina S. V.
Financial University under Government of Russian Federation,
Russia, Moscow, Leningradskij pr., 49
Tel.: + 7 (915)-007-04-89
E-mail: svprokopchina@mail.ru

Summary: The broad digitalization of the company's activities is focused on solving complex tasks of monitoring and managing production and socio-economic systems. The functioning of such systems takes place in conditions of significant uncertainty of information, active and unpredictable influence of the external environment, dynamism of situations. Therefore, for the effective implementation of digitalization systems, special methodologies, technologies and systems focused on such situations are required. First of all, this concerns methods and means of obtaining measurement information.

The article discusses a new direction of modern measurement theory, which has been called intelligent measurements. A comparison of classical and intellectual measurement methodologies is given. It is shown that classical measurements are a special case of intelligent measurements. The advantages of using this type of measurement are indicated both for the theory and practice of measurements, and for solving problems of artificial intelligence. Examples of the use of Bayesian measurement methodology, technology and practice for solving various applied problems are given

Keywords, Intelligent measurements, Bayesian approach, digitalization.

1. Introduction

For a sufficiently long period of time, measurement was understood as the process of obtaining a numerical value of a physical quantity by comparing it with some model quantity. At the same time, all components of the measuring process (measurement model, model standard, comparison scheme or measurement technique) were clearly defined a priori. The measurement conditions were considered unchanged at any time during the implementation of measurements.

The processes of digitalization of various spheres of human activity in industry, energy, ecology, social and economic spheres have defined new measurement tasks, such as measurement of non-quantitative quantities, measurement with uncertainty (fuzziness, significant inaccuracy, incompleteness of a priori information), measurement in changing conditions and the active influence of environmental factors on the object of measurement. It is appropriate to note that, despite the fact that the term "digitalization" has become popular in the last decade, the processes of digitalization have begun to intensify since the mid-80s of the last century, which was associated with the development of computer systems.

At that time, in order to solve new measurement problems, the development of a new type of measurement – intelligent measurements - began. Moreover, this term was originally introduced in 1986 in the works of D. Hoffmann, A. Karaya, L. Finkelstein [1, 2] to define self-testing and self-adjusting measuring systems. However, such intelligent measurements and tools based on them could not solve the above tasks, since their intellectualization was focused only on improving the structure of the measuring systems themselves, and not on developing the methodology of the measuring process.

In the mid-90s of the last century, a new direction of measurement intellectualization was developed [3-5], in which intelligent measurements were primarily understood as involving artificial intelligence methods in the measurement process. New types of measuring scales have been proposed, focused on application in conditions of significant uncertainty and integration of data and knowledge to obtain a measuring solution. The use of such measurements was especially effective for complex measuring objects that actively interact with the external environment. Measurement techniques have become much more complicated and have become quite complex information technologies implemented by means of integrated sensor and processor systems.

Currently, there is an intensive development of this type of measurement, which covers the development of intelligent sensor systems, in particular intelligent IoT, control systems for complex industrial and socio-economic complexes, big data processing systems, business sustainability management systems, image processing systems, messages, weakly structured and unstructured information [6, 7].

An important stage in the development of intelligent measurements is the development of new terminology standards, the release of which is scheduled for 2023.

However, despite a number of numerous monographs and articles, defended dissertations devoted to intellectual measurements, despite a sufficient number of projects carried out on this methodological basis, the systematization of knowledge and experience in the use of intellectual measurements remains very relevant for a wide range of specialists.

In this regard, this article attempts to show the new fundamental possibilities of intellectual measurements and provide guidelines for obtaining information about them to solve the problems of digitalization in various spheres of society.

2. Intelligent Measurements: what are the Fundamental Differences from Classical Measurements

Speaking about the differences between the principles of intelligent measurements and measurements according to the classical scheme, we can note a number of positions on which there are significant differences.

In classical measurements of physical quantities, three main positions are methodologically determined:

1. What to measure.

That is, the object of measurement is determined – the value of a physical quantity.

2. What type of measuring device.

The type of measuring instrument is determined.

3. How to measure.

The measurement method is determined.

4. Measurement metrology.

As an assessment of the quality of the measurement result, the first two central moments of a random variable are used: estimates of mathematical expectation and variance.

5. Interpretation of measurements.

The interpretation of the received solution is made by the meter itself.

In intelligent measurements, there are, in addition to the above, additional positions:

6. Object type.

The type of object is characterized by three features:

-simple and complex measurement objects;

- An object of a quantitative or qualitative type;

- An object of static or dynamic type.

A simple object has no additional measurable properties, like a measurement object in classical measurements.

A complex object is a system with a number of properties that will be measured. A complex object has a hierarchical structure determined by the presence of these additionally measurable properties and their relationships.

The result of measuring an object of a quantitative type is determined by numerical values.

The result of measuring an object of a qualitative type is represented by linguistic values.

An object of static type has a constant measurement model.

A dynamic type object has a time-varying measurement model.

7. Measurement conditions.

In intelligent measurements, it is assumed that the object of measurement has relationships with its surrounding environment. In classical measurement schemes, this relationship is absent, and it is assumed that the measurement conditions are unchanged and do not affect the object of measurement.

Measurement conditions in intelligent measurements are represented by a set of sets of a priori knowledge, metrological requirements, restrictions. They are explicitly included in the measurement equation.

8. Implementation of the measurement process and measurement equations.

The technologies for implementing the measuring process are fundamentally different. In classical measurements, under conditions of complete a priori certainty on the components of the measuring process, such as the model of the measuring object, measurement methods, model objects, the measurement process was a simple transformation of the sensory value into the required physical quantity.

Equation of classical measurements:

$$h = f\{X\}, \qquad (1)$$

where h is the result of measuring a physical quantity; f{X} is a measurement function that converts the sensory value of X into a measurement result.

Intelligent measurements are designed for measurements under conditions of considerable uncertainty of the components of the measuring process listed above and therefore require special technologies for their implementation.

In the intelligent measurements proposed in [3-5], the measurement process is based on the principle of organizing perceptron pattern recognition, which are the reference points of special scales, called scales with dynamic constraints (DBS) because of their ability to change the measurement range during measurements. Therefore, the measurement process in measurements of this type is implemented on the basis of the logical conclusion of solutions. In conditions of uncertainty, with such implementation of measurements, measurement solutions are obtained in a fuzzy form in the form of a set of reference points of the SDO scale with their corresponding probability values for numerical values or possibilities for linguistic values of SDO. Therefore, the equation of intelligent measurements is written in the form [3-5].

The equation of intelligent measurements:

$$\{hkt(Q) \mid \{MX\}kt(Q)\} \mid (Yt(Q); \{Xit\}) = \{\text{argexstr } C [\varphi jt(fit\{Xit\} \mid (Yt(Q) * Yt(OE) * Gt(OE))]\}, \qquad (2)$$

where hkt(Q) is the measuring solution with a set of metrological characteristics {MX}kt(Q)}, obtained under conditions of Yt (Q) and the external environment model Gt(OE)); obtained under conditions of Yt (OE); C is the criterion for choosing a solution (scale reference); φjt is the optimizing transformation of the measurement result into a measurement solution, fit{Xit} is the functional transformation of the initial information presented in numerical or linguistic form.

If criterion C takes the form of a Bayesian decision rule, then such intelligent measurements are called Bayesian.

By the form of equation (2), it can be concluded that the measurement result in intelligent measurement systems is subject to additional processing in order to obtain an interpreted measurement solution in the form of estimates or recommendations. However, this processing is based on full metrological support of each stage of its implementation. Thus, the implementation of equation (2) is an integral measurement method for obtaining and intelligent processing of information.

A comparison of the formulas of the measurement equations (1) and (2) shows that classical measurements are a special case of intelligent measurements, provided that the measurement result is obtained by calculating functional dependencies, the selection criterion is the proximity of the distance of the obtained experimental value to the reference point, the measurement conditions are not taken into account and the result is obtained only in the form of numerical values.

If, in equation (2), some other criterion is used as a criterion for choosing solutions, rather than Bayesian, then it is possible to obtain intelligent measurement equations of other types, other than Bayesian.

Some of them are given in [3-5], where the classification of measurements, including intellectual ones, is also given.

For example, if we take the criteria of Zadeh's fuzzy logic, then equation (2) will be the equation of soft measurements. In the works of V. B. Tarasov [13] the possibility of using other logics, for example, logic to create systems of cognitive measurements, is being considered.

9. Why measure – the purpose of measurements.

In intelligent measurements, the answer to the question "why measure" is very important, since the measurement result is not a value of a physical quantity, but a measurement solution that can be expressed in the form of an assessment, a comparative (audit) assessment, a conclusion, a recommendation with explanations of the reasons. That is, this type of measurement result corresponds to the purpose of measurements. If the purpose of the measurements is to check for compliance with regulations, then the result of the measurement will be an audit decision. If a measuring solution is necessary for the management of an object, then the result of the measurements will be a management recommendation with explanations of its application.

10. Metrology of knowledge.

In intelligent measurements, not only the metrology of the values obtained (data metrology) is used, but also the metrology of obtaining a quality solution (knowledge metrology).

All of the above allows you to significantly expand the functionality of measurements.

As will be shown later, these capabilities are realized on the basis of integration of measurement theory and artificial intelligence methods.

For artificial intelligence methods, new capabilities of intelligent measurements are also very important, allowing you to answer and solve problems: obtained and manage the quality of measurement solutions.

In the theory of intelligent measurements, the direction of metrological justification of decisions has been developed. Each measurement solution is accompanied by a set of metrological characteristics, including not only accuracy indicators, as in classical measurements, but also reliability indicators, consisting of error levels of the 1st and 2nd kind, reliability, risk of decisions, the amount of entropy removed during measurements, the amount of information received. This is very important in conditions of uncertainty and the construction of complex measurement technologies. In addition, an important function implemented in intelligent measurements is the possibility of metrological synthesis of solutions, which allows you to build highly reliable and efficient measuring systems

2. How to integrate data and knowledge

In intelligent measurements, the principles of modern trends in DATA SCIENCE, BIG DATA, BI are implemented, and therefore a very significant opportunity is the possibility of integrating data and knowledge, which is realized through the use of scales such as SDO.

3. How to make decisions interpretable and understandable

Users (Ambient AI) is the main issue of artificial intelligence at the moment.

In intelligent measurement and their systems, this problem has been solved. Each decision has an explanation in the form of the reasons that influenced the state of the factor, and the factors that need to be adjusted to normalize the state of the measured indicator are indicated.

3. Intelligent Measurements: new Opportunities in the Theory and Practice of Measurements and in Artificial Intelligence

Let's define new opportunities of information technologies based on intelligent technologies.

They can be represented by the following list:

1. Using different types of data. This is possible with the use of coupled scales with dynamic constraints;
2. The use of knowledge based on linguistic SDOs;
3. Integration of data and knowledge based on the USDO and BIT;
4. Accelerated processing of big data, and dimension reduction based on modified Bayesian convolution;
5. Processing of small samples and data from unique experiments based on expert assessments and SDO;
6. Quality management of solutions and metrology of data and knowledge based on metrological support;

7. Dynamic modeling in conditions of uncertainty and the creation of developing digital counterparts of complex technical and socio-economic systems;

8. Measuring the influence of environmental factors and factors of mutual influence of the object and the environment;

9. Self-learning and self-development of intelligent measuring systems;

10. Building intelligent sensor networks based on the regularizing Bayesian approach;

11. Development of hybrid measurement networks based on the integration of instrument measurement systems, IIoT and neural networks under uncertainty;

12. Creation of intellectual workplaces (IRM) of specialists – self-training of users;

13. Integration of technical and socio-economic systems based on a holistic system approach as a way to sustainable development of a digital society.

3. Practical examples of the use of Bayesian intelligent measurements.

Since the 90s of the last century, over several decades of development of the theory and practice of intelligent measurements, in particular, Bayesian intelligent measurements, considerable experience has been accumulated in their application in various fields of activity, scientific projects, in the field of education and professional training of specialists.

Practical examples of such experience are given below with links to relevant sources where detailed information can be found.

1. Intelligent sensor networks in power engineering [3-6]:

1.1. Power generation;

1.2. Power supply;

1.3. Renewable energy sources;

2. Intelligent sensor networks in ecology and environmental protection [3, 6, 12];

2.1. Industrial ecology;

2.2. Protection of wild animals;

2.3. Digitalization of nature reserves and protected areas;

3. Digitalization systems in the socio-economic sphere [3, 4, 6];

3.1. Digitalization of the regional economy. Management of sustainable development of territories;

3.2. Housing and communal services and water supply [7];

3.3. Digitalization of cultural and scientific objects for the spiritual development of society [4, 6, 7];

3.4. Digitalization of environmental management. [3, 4, 11];

3.4.1. Fishery activities. [3, 4, 5, 11];

3.4.2. Agricultural activity. [3, 4, 5, 11];

3.4.3. Management of the use of natural resources;

3.5. Digitalization of municipal administration. [3, 4, 5, 11];

3.6. Investment design and development of the region based on intelligent measurement technologies [3-5];

4. Digitalization of industrial production and enterprise management. [3-5, 11, 12];

4.1. Enterprise management systems in conditions of uncertainty based on risk accounting and potential development [3, 4, 12];

4.2. Digitalization of personnel management based on Bayesian intelligent measurements for the use of unstructured data [3, 11];

5. Digitalization and forecasting of geopolitical processes based on global measurement technologies. [3, 10].

Conclusion

Thus, the new direction in measurement theory is a bridge between the classical measurement theory and artificial intelligence, which makes it possible to effectively use the advantages of these theories based on the deep integration of their methodologies. Intelligent measurement technologies are particularly effective for implementing digitalization processes in conditions of significant flows of diverse information, the predominance of the importance of processing and interpreting information over "raw" data and the special relevance of obtaining analytical solutions in conditions of uncertainty. We hope that the information provided on the experience and effectiveness of using intelligent measurement technologies will be useful to a wide range of both researchers and practitioners.

References

[1]. Finkelstein L., Hofmann D., Intelligent measurement. A look at the current state and modern trends, *Measurement*, Vol. 5, No. 4, 1987, pp. 151-153.

[2.] Hofmann D., Karaya K., Intelligent measurements for obtaining objective information in science and technology, in *Proceedings of the X IMECO International Congress,* Prague, Vol. 1, 1985, pp. 19-34.

[3]. Prokopchina S. V., Bayesian intellectual measurements, *Scientific Library*, Moscow, 2021.

[4]. Prokopchina S. V., Modeling and management of digitalization processes in conditions of uncertainty, *Scientific Library*, Moscow, 2021.

[5]. Prokopchina S. V., Bayesian intelligent technologies as a methodological basis for big data processing in conditions of uncertainty, *Economics and Management: Problems, Solutions*, 2019, Vol. 11, No. 3, pp. 105-109.

[6]. Prokopchina S. V., Shcherbakov G. A., Efimov Yu. V., Modeling of economic systems in conditions of uncertainty, Study guide - workshop, 2nd edition, edited by G. A. Shcherbakov, *Scientific Library*, Moscow, 2018.

[7]. Prokopchina S. V., Methods for implementing the concept of 'Smart City' based on Bayesian Intelligent Technologies, *Journal of Physics: Conference Series*, 1703, 1, 2020, 012018.

[8]. Prokopchina S. V., Methodological foundations of scaling in modern Measurement Theory. Classification

of measurement scales and their application under uncertainty based on Bayesian Intelligent Technologies, *Journal of Physics: Conference Series*, 1703, 1, 2020, 012003.

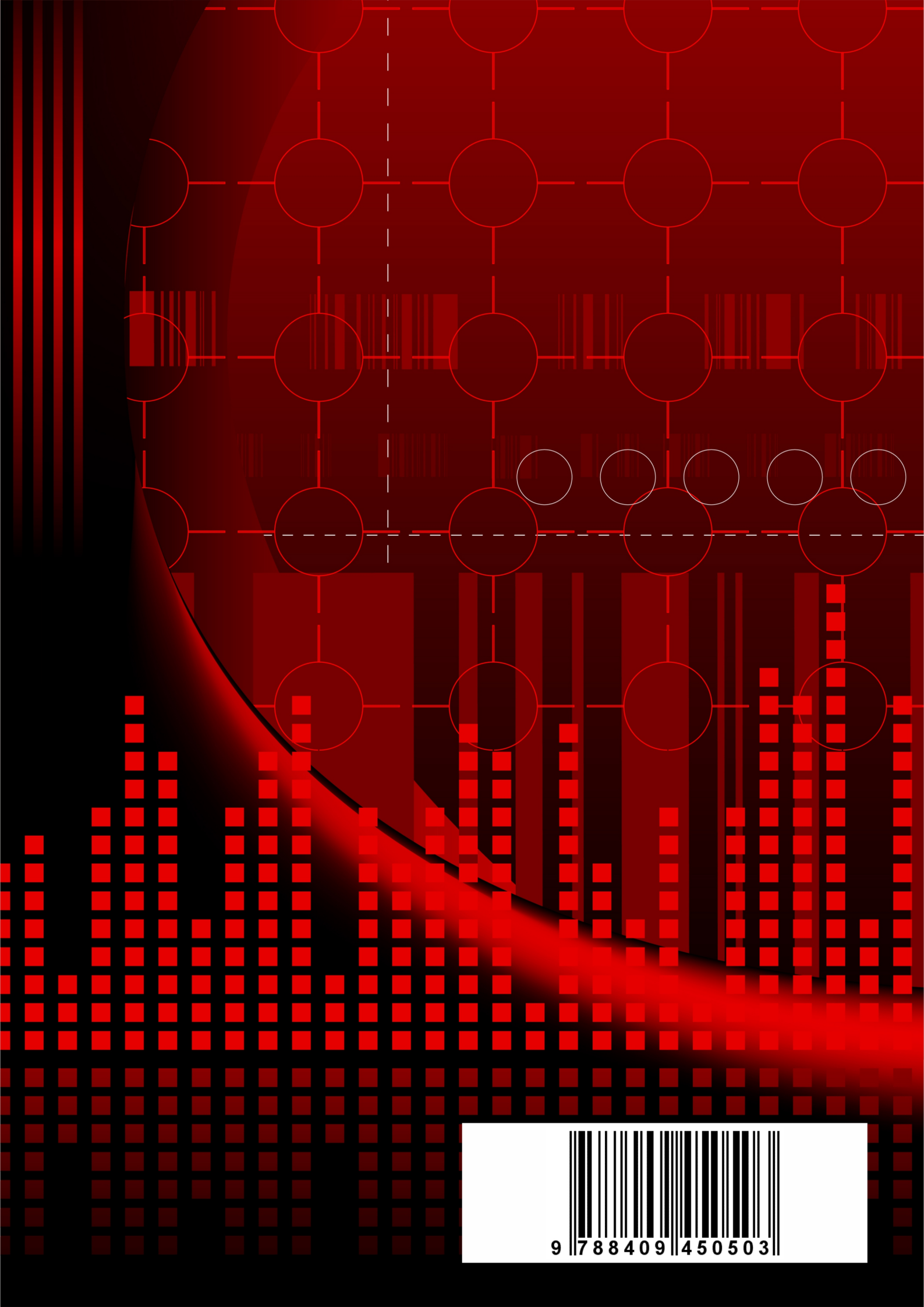
[9]. Prokopchina S. V., A new type of neural networks: Bayesian measuring neural networks (BIN) based on the methodology of the regularizing Bayesian approach, *Soft Measurements and Calculations*, Vol. 27, 2020, pp. 17-24.

[10]. Prokopchina, S. V., Methodological Aspects of the Concept of Global Measurements, in *Proceedings of the 23rd International Conference on Soft Computing and Measurements (SCM' 2020),* 2020, 9198796, pp. 3–6.

[11]. Prokopchina S., New trends in measurement theory: Bayesian intelligent measurement and its application in the digital economy, *CEUR Workshop Proceedings*, 2782, 2020, pp. 80–88.

[12]. Prokopchina S. V., Mishchenko S. N., Bayesian Intelligent Technologies for Managing Complex Production Systems, in *Proceedings of the 23rd International Conference on Soft Computing and Measurements (SCM' 2020),* 2020, 9198794, pp. 97–100.

[13]. Tarassov V. B., Svjatkina M. N., Using Zadeh's Granulation Concep: Granular logics and Their Applications in Sensor Data Analisis, in Recent Development and New Directions in Soft Computing, *Springer*, New York, 2014, pp. 453-466.


9 788409 450503