Sensors & Transducers, Vol. 170, Issue 5, May 2014, pp. 148-152

Sensors & Transducers

http://www.sensorsportal.com

Research for Result Classification Method in Semantic Network (SN) Using Iterative Classification Algorithm

Chen Jian
Computer Engineering Technical College, Guangdong Institute of Science and Technology, Zhuhai 519090, P. R. China
E-mail: zhuhaicj@163.com

Received: /Accepted: 30 April 2014 /Published: 31 May 2014

Abstract: In semantic network component architecture is useful because of its internal availability for information integration. In this paper we represent a new approach called Multi-Factor based Classification (MFC) embedded with the traditional Iterative Classification Algorithm which takes different factors (Node information level, Range coordinate from the communication node, SI, Implicit factor of data to reach communication node) into consideration to form components. Then the result of the components is estimated based on the Equal of the node distribution, Node range per component, Web component range and required information level of each centre. Our result shows that by changing Multi-Factors we can generate components with more equally distributed nodes, change component range and Choose low information using centre.

Keywords: Iterative classification algorithm, Information utilization rate, Equal division, Implicit factor.

1. Introduction

Semantic networks (SN) are highly distributed networks of autonomous small, lightweight sensors (nodes) in large numbers to monitor physical or environmental conditions by the measurement of, temperature, sound, vibration, pressure, motion or pollutants and to cooperatively pass their data through the network to a main location (often called a sink).

It has already made its way in military object, habitat monitoring [1] and object tracking because of the characteristics such as feasibility of rapid deployment, self-organization (different from Infrastructure Mode or ad hoc network [2]) and fault tolerance. But limited computation capability, limited information and small memory size has made designing the SNs is very difficult [3]. The information consumption is the most important factor among these three factors, because the battery is not changeable if once the sensor nodes are deployed. The information is also the major consideration in designing the routing of the SNs. Hierarchical protocols reduce information consumption in the networks by Classification. Classification algorithms partition data objects (patterns, entities, instances, observances, units) into a certain number of components (groups, subsets, or categories).

Several available operational definitions [4] summarized by R. Xu, I. Donald are as follows:

"A component is a set of entities which are alike, and entities from different components are not alike."

A component is "an aggregate of points in the test space such that the range between any two points in the component is less than the range between any point in the component and any point not in it."

"Components may be described as continuous regions of this space (d-dimensional feature space)

Article number P_2049

containing a relatively high density of points, separated from other such regions by regions containing a relatively low density of points."

In these protocols, nodes are divided into some components and some nodes based on some factor are the selected as component heads (CH). These component heads exchange data with the Communication Node (BS) which costs the most information of the nodes. Even though this concept has similarity with the Basic Service set (BSS) of Infrastructure mode where there is an Access point (AP) (here the CH) and few cells communicate via this access point, the method in SN is much more dynamic and information Utilization rate. Due to these advantages, sensor nodes can remarkably save their own information.

In this paper by we have proposed a new method called MFC (Multi-Factor based Classification) using Iterative Classification algorithm and variation of factor and thus made four contributions:

1) We can have control over the random node distribution by considering various factor combinations. Thus we can avoid components with poor distribution of node or highly dense components. The communication node can centrally design the network with Good components by our given criteria.

2) A good component can be defined in a new way with the following property:

a. Node distribution is approximate uniform

b. Inter-component range is high

c. The intra-component range is low and thus

d. The ratio of Web component range called validity is low

3) Minimum and maximum range of nodes in each component varies with factors.

4) By adding valid factor we can choose low information using centre and thus make a low information consumed network.

The remainder of the paper is organized as follows. In the following section, related work is discussed. We, then briefly describe Iterative Classification Algorithm in Section 3. Implementation of scheme is described in Section 5 followed by the experimental results in Section 6. Section 7 presents our conclusion and direction for future works.

2. Related Work

For Classification, various factors have been taken into consideration. The most popular Classification mechanism LEACH [6, 7] along with LEACH-C has taken residual information level of the nodes for component head selection for creating components. It has been achieved by setting the probability $P_i(t)$ of a node, becoming a component-head as a function of nodes information level relative to the aggregate information remaining in the network rather than purely as a function of the number of times the node has been component-head:

$$P_i(t) = \frac{E_i(t)}{E_{total}(t)}(k + 0.5),$$

where $E_i(t)$ is the current information of node i, and $E_{total}(t) = \sum_{t=1}^{N}(E_i(t)+1)$.

The main drawback of LEACH is that the number of component heads is uncertain and there is a chance of poor Classification while LEACH-C requires the position of the entire sensors to avoid this problem. On the Other hand instead of considering node information level ACE [8] has considered node degree for Classification by using an exponentially decreasing function for f_{min}:

$$f_{min} = \left(e^{-k_1 \frac{t}{cI}} - k_2 + 1 \right) d^2$$

In this formula, t is the time passed since the protocol began and cI is the duration of the protocol as described earlier. d is the estimated average degree (number of neighbors) of a node in the network, and is pre-calculated prior to deployment. k1 and k2 are chosen constants that determine the shape of the exponential graph.

Classification using Genetic Algorithm range between nodes and number of component heads has been taken into account While Gupta [10] in their method selected residual information of the nodes, the number of neighbor nodes and centrality for Classification.

A component head selection algorithm is proposed by Scott that uses RF signal strength for head selection which is widely implemented in many commercial radios. A component head is selected based on its perceived RF signal strength of its neighbors. This approach has several advantages. First, unlike relying on Euclidean range which requires localization or network knowledge, using RF signal strength compensates for the network properties. Nodes closed by range may not be closed by signal requiring more information for transmissions. This also compensates for dead spots, uneven propagation, and changing RF propagation characteristics. Additionally, component heads located in areas of higher node density are expected to have a lower average range between end nodes and the component head. So, it is clear that various factors are chosen for different Classification algorithm.

Based on above studies we wanted to investigate the effect of taking all these factors into consideration together while Classification.

3. A Brief Description of Iterative Classification Algorithm

Iterative Classification is an algorithm to classify or to group given objects based on attributes or

factors, into K number of groups. K is a positive integer number. The Iterative Classification algorithm was developed by J. Mac-Queen (1967). The grouping is done by minimizing the sum of squares of ranges between data and the corresponding component centre

Given k, the Iterative Classification Algorithm is implemented in 4 steps:

1. Partition objects into k nonempty subsets and Compute seed points as the centre z1 (1), z2 (1), z3 (1),…., zk (1) of the components of the current partition.

2. At the k-th iterative step, distribute the object {x} among the K components using the relation,

$$x \in C_j(k) \quad if \quad \left\| x - z_j(k)^2 \right\| < \left\| x - z_i(k)^2 \right\|$$

For all i = 1, 2… k; i ≠ j; where $C_j(k)$ denotes the set of samples whose centre is $z_j(k)$.

3. Compute the new component centers $z_j(k+1)$, j =1, 2… k such that the sum of the squared ranges from all points in $C_j(k)$ to the new component centre is changed.

The measure which changes this is simply the sample mean of $C_j(k)$. Therefore, the new component centre is given by

$$Z_j(k+1)^2 = \frac{1}{N_j}\left(\sum_{x \in C_{j(k)}} x + 1 \right)$$

where j= 1, 2, 3, ..., k, and N_j is the number of samples in $C_{j(k)}$.

4. Go back to Step 2, stop if $z_j(k+1) = z_j(k)$ for j = 1, 2, …, k then the algorithm has converged and the procedure is terminated. Otherwise go to Step 2.

4. Factors

Selecting appropriate factor for any given SN is another challenge. But here we wanted to put forward those factors which are common for almost all SN. For that most of the factors we have chosen are related to basic hardware level. They are residual information level of the nodes, Transmission range (often characterized by SI), Modulation type, Range (among the nodes, from the BS etc), Node density, Node angel, Centrality, bit rate, Turn on & wake up time, Processing information, Implicit factor. We have selected Nodes information. Range from the BS, SI and Implicit factor for our experiment which we think most of the real life scenarios and application.

One of the most significant factors for SN is range. The position of the communication node in the semantic network has great impact on the transmission and the receiving signal strength.

For our experiment we have considered the Range coordinate from the communication node (BS) to each node. It varied from 85 m to 175 m. The higher the range more information is required to data interchange

$$Euclidian \quad distance = \sqrt{(x_i - x_c)^2 + (y_i - y_c)^2}$$

The relationship between SI (Signal Intensity) values and range is the foundation and the key of ranging and positioning technologies in wireless sensor networks.

As the use of SI ranging need less communication overhead, lower implementation complexity, and lower cost, so it is very suitable for the nodes in semantic network which have limited information.

The principle of SI ranging describes the relationship between transmitted information and received information of wireless signals and the range among nodes. This relationship is:

$$P_r = P_t * (1/d)^{n+1}, \tag{1}$$

where P_r is the received information of wireless signals, P_t is the transmitted information of wireless signal, d is the range between the sending nodes and n is the transmission factor whose value depends on the propagation environment.

Take 10 times the logarithm of both sides on (1) then Equation (1) is transformed to Equation (2).

$$10 lg P_r = 10 lg P_t - 10 n log d, \tag{2}$$

Pr, the transmitted information of nodes, is given 10logP is the expression of the information converted to dBm. Equation (2) can be directly written as Equation (3).

$$P_R(dBm) = A - 10 n log d, \tag{3}$$

By Equation (3), we can see that the values of factor A and factor n determine the relationship between the strength of received signals and the range of signal transmission.

For our experiment we have taken the SI value from 0 to 100 dB converting the dBm value collected by the nodes by the BS.

$$RSSI(dB) = 10 log\left(\frac{P_t}{P_r} + 1\right)$$

Classification in SN should take into account the implicit factor. Implicit factor is an important factor for system reliability such as the case of emergency response, and accuracy of data reporting in case of high frequency periodical data updates. On the other hand, information consumption is essential to ensure survivability of sensor nodes and hence the lifetime of the system. In a Multi-hop network, significant delay occurs at each hop due to contention for the wireless channel, packets processing and queuing

delay. The implicit factor is therefore a function of the number of communication hops between the source and the gateway (CH). Implicit factor is also a dependent on range. For our experiment our networks implicit factor is a function of range. As the range increases the implicit factor increases and vice versa.

5. Implementation of Scheme

Assumptions:

1) All nodes have same and adequate amount initial information.

2) Each node or sink has ability to transmit message to any other node and sink directly.

3) Each sensor node has radio information control; node can tune the magnitude according to the transmission range.

4) Each sensor node has location information and fixed after deployment.

Then the result of the components is estimated based on the Equal of the node distribution, Node range per component, Web component range and required information level of each centre.

Assume that there are N nodes distributed equally in an M × M region. If there are k components there are on average N/k nodes per component (one component head and (N/k)—1 non component head nodes). As we have tested with 22 nodes, Table 1 shows the required nodes per component below.

Table 1. Average nodes per component for our experiment.

Number of components	Average nodes/component
2	9.8
3	7
4	5.2
5	4.1

This is the range between a point and its component centre [9] to determine whether the components are compact. We take the average of all of these ranges, defined as

$$int\, ra = \frac{1}{N}\left(\sum_{i=1}^{K}\sum_{x\in C_i}\left\|x - z_i\right\|^2 + 1\right),$$

where N is the number of nodes in the network, K is the number of components, and z_i is the component centre of component C_i. We obviously want to change this measure.

This is the range between components. We calculate this as the range between component centers, and take the minimum of this value, defined as

$$\text{int}\, er = \min\left(\left\|z_i - z_j\right\|\right)^2$$

i=1,2, …, k-1 and j=i+1, …, k.

we take only the minimum of this value as we want the smallest of this range to be changed, and the other larger values will automatically be bigger than this value.

For the use in simulation the generated random numbers must be transformed to random variables using suitable distribution method.

***Uniform Process Generator*:**

Let X be a equally distributed random variable with probability density function

$$f(x) = \frac{1}{b-a}, a < x < b$$

$$f(x) = \int_a^x \frac{dy}{b-a} = \frac{x-a}{b-a}, \qquad a > x < b$$

$$X = a + (b-a)r$$

Is the required process generator

***Exponential Process Generator*:**

$$f(x) = ae^{-ax} \qquad a > 0, x > 0, = 0 \qquad otherwise.$$

$$f(x) = \int_0^x ae^{-t}dt = 1 - e^{-ax}$$

$$E(x) = \frac{1}{a}, \quad Var(x) = \frac{1}{a^2} = \left[E(x)\right]^2$$

By inverse technique, R = 1–e–ax or 1–R = 1–e–ax X = –(1/α) ln R, since R is likely to occur as 1–R = –E(x). ln R where R is IID (Independent Identically Distributed). For our experiment we have chosen Uniform Process Generator but future work can investigate more with negative Exponential Process Generator.

6. Experimental Results

In our work we carry out our simulation with 22 sensors deployed randomly assuming that they are deployed in components with inter-components communication will happen only through component heads of the respective components. Table 2 shows the simulation factors.

Table 2. Factors used in simulation study.

Factor	Value
Node information (B)	1J-3J
Range from the Communication node(D)	85 m-175 m
SI (R)	0 to 100 dBm
Implicit factor (L)	Depends on the range
Number of nodes	22

When we compare the effect of Multi-Factors in two component network, we know that about 83 % of the components based on the various combinations of two factors are equally distributed to the theoretical value while 50 % of components are uniform for one factor.

When we observe the node distribution in the three component network, we find that a combination of two factors (Node information level & Range coordinate from BS) results 100 % uniform node distribution in each component which reveals a significant intersection point of three lines on the theoretical value base line .

We see in the four component network that changes in the factor combination results in an alternation in the range (maximum to minimum) of nodes per component. Our experiments show by increasing or decreasing we can alter node range in any specific component.

One of the significant observations of our experiment is that the goal of minimizing the intra component range can be achieved by choosing the right combination of the factors. For example, in two component network combination of three factors results in minimum intra component range (0.252 m) while for five component network two factors results the least (0.360 m).

We can know that minimizing the inter component range results maximization in the intra component range. High Inter component range reduces the cross talk or interference and high information consumption.

If the components that we choose have centre with high information that will results in a less information Utilization rate. From our experiment we know that the increase in the number of factor in five component network results in centre with low information requirement for the same component.

7. Conclusions and Future Works

For our experiment we have chosen 4 factors which leads to a variation of $2^4 = 16$. Now if the number of factors increases the number of combination will increase exponentially, which will eventually lead to NP complete problem. Various techniques can be applied to solve this kind of problems. For instance In LEACH-C [6] the BS finds components using the simulated annealing algorithm to solve the NP-hard problem of finding k optimal components.

Different SN is made up for different purposes. So the effect of factors will vary from network to network. In out experiment we tried to work with those factors which are common for almost all the sensors. But finding the right combination of factors and their optimum value for specific SN is a great challenge.

For our experiment we have worked with uniform distributed data type. Further experiment can be done by choosing exponential, Negative exponential distribution.

We have used python for our experiment because of its robustness and Utilization rate to deal with huge number of data. But while conducting this experiment we were not aware of any tools that can perform the same functionality. Experiments our concept using different programming language and compare our data with the output can pave the way to further research field.

Acknowledgement

This work has been supported by the Science and Technology Project of Zhuhai City (Grant No. 2011A050101006).

References

[1]. F. Akyldiz, W. Su, Y. Sankarasubramaniam and E. Cayirci, Integrated information network: a survey, *Computer Networks*, Vol. 38, No. 4, 2002, pp. 393-422.

[2]. B. Elbhiri, S. El Fkihi, R. Saadane and D. Aboutajdine, Classification in wireless integrated information network based on near optimal bi-partitions, in *Proceedings of the EURO-NF Conference on Next Generation Internet (NGI)*, 2-4 June 2010, pp. 1-6.

[3]. Y. Wang and G. H. Cao, On full-view coverage in camera integrated information networks, in *Proceedings of the IEEE International Conference on Computer Communications (INFOCOM)*, 10-15 April 2011, pp. 1781-1789.

[4]. I. F. Akyildiz, T. Melodia and K. R. Chowdhury, A survey on integrated information network, *Computer Networks*, Vol. 51, No. 4, 2007, pp. 921-960.

[5]. Ruo Hu, Channel access controlling in wireless integrated information network using smart grid system, *Applied Mathematics & Information Sciences*, No. 11, 2012-11, pp. 813-820.

[6]. Ruo Hu, Stability analysis of wireless integrated information network service via data stream methods, *Applied Mathematics & Information Sciences*, No. 11, 2012, pp. 793-798.

[7]. Hu Ruo, New network access control method using intelligence agent technology, *Applied Mathematics & Information Sciences*, No. 2, 2013, pp. 44-48.

[8]. S. Banerjee and S. Khuller, A classification scheme for hierarchical control in multi-hop wireless networks, in *Proceedings of the IEEE International Conference on Computer Communications (INFOCOM)*, 16-21 July 2001, pp. 1028-1037.

[9]. A. Garg and M. Hanmandlu, An resource-aware adaptive classification protocol for integrated information networks, in *Proceedings of the International Conference on Intelligent Sensing and Information Processing (ICISIP)*, 2006, pp. 13-30.

[10]. D. Lu, N. Jie and X. X. Huang, Classification based spectrum allocation scheme in mobile ad hoc networks, *Bulletin of Advanced Technology Research (BATR)*, Vol. 5, No. 12, 2011, pp. 37-41.