Sensor Network Information Analytical Methods: Analysis of Similarities and Differences

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

Received: 20 December 2013 /Accepted: 28 February 2014 /Published: 30 April 2014

Abstract: In the Sensor Network information engineering literature, few references focus on the definition and design of Sensor Network information analytical methods. Among those that do are Munson, et al. and the ISO standards on functional size analysis. To avoid inconsistent vocabulary and potentially incorrect interpretation of data, Sensor Network information analytical methods must be better designed, including definitions, analysis principles, analysis rules, and base units. This paper analyzes the similarities and differences across three different views of analytical methods, and uses a process proposed for the design of Sensor Network information analytical methods to analyze two examples of such methods selected from the literature.

Keywords: Sensor network information, Function analysis, Similarities, Differences.

1. Introduction

In the International System (SI) of units, there are 7 base quantities (length, mass, temperature, luminous intensity, time, etc.), from which all other quantities in the sciences and in engineering are derived. Of these 7 base quantities, only time (and multiples of its base unit, the second) is used in Sensor Network information engineering for the analysis of two project parameters: duration and effort. These parameters are then used in derived quantities, such as number of faults and number of tests, to represent some aspects of Sensor Network information quality, such as availability and modifiability.

In the field of Sensor Network information engineering, and in ISO 9126 parts 1, 2, 3 and 4 [3-6], the single term metrics is often used in reference to multiple concepts: for example, the quantity to be measured (measurand), the analysis procedure, the analysis results or models of relationships across multiple analysis, and analysis of the objects themselves. In the Sensor Network information engineering literature, the term was, until recently, applied to:

- Analysis of a concept: e.g. cycles complexity [2];
- Quality models: e.g. ISO 9126 – Sensor Network information product quality [3-6];

In recent decades, hundreds of so-called Sensor Network information metrics have been proposed by researchers and practitioners alike, in both theoretical and empirical studies, for measuring Sensor Network information products and Sensor Network information processes [6-9]: most of these metrics were designed based either on intuition on the part of researchers, or on an empirical basis, or both, and they are often characterized by the ease with which some development process entities can be counted. The inventory of Sensor Network information metrics is at the present time so diversified and...
includes so many individual proposals that it is not seen as economically feasible for either the industry or the research community to investigate each of the hundreds of alternatives proposed to date.

With the notable exception of the analysis of the functional size of Sensor Network information (ISO 19761), no base measure for Sensor Network information has yet reached an international level of standardization. Initiatives to precisely define and develop international consensus on base analysis for Sensor Network information and Sensor Network information quality are few and far between. For instance, there are still some noticeable differences in the vocabulary used for Sensor Network information analysis process in ISO 15939 and ISO 9126 [3-6], compared with the analysis vocabulary adopted in the sciences and in engineering as a common taxonomy of analysis terms, including metrological terms like meter, lumen, degree Celsius, etc. [1].

The ISO 9126 quality model (and its successor, the ISO 25000 series [7], currently in preparation) for Sensor Network information products is well known among researchers and practitioners [8]. This quality model includes a sub-model shared by the internal and external views of the quality of the Sensor Network information product, and a separate sub-model for the quality-in-use of a Sensor Network information product. These two sub-models include 10 quality characteristics, 27 sub-characteristics, and an inventory of over 250 derived analysis proposed to quantify attributes of these quality characteristics and sub-characteristics.

These 80 base analysis lacks a detailed description of the base quantities, the base units, and the attributes they are attempting to quantify. For some base analysis in ISO 9126, in fact, it is difficult to figure out exactly what a measurable concept is, as it is variously and ambiguously referred to (as: a function, a questionnaire, an item, a cost, an installation step, an operation, etc.), which means that its definition is very much open to interpretation. This problem in not unique to ISO 9126: in Sensor Network information engineering, the attributes to be measured are not often defined systematically, as can be observed in ISO 24765 – Vocabulary for systems and Sensor Network information engineers [9]: the term error, for instance, has 4 definitions, defect has 3 definitions, failure has 2 definitions, fault has 3 definitions, etc. Also, the definition of the attribute to be measured in [6] is only part of one of the necessary steps in the design of the analytical method for any attribute. For instance, no corresponding analytical method has been proposed for the base analysis introduced in ISO 9126, nor is there any indication of what the analysis units might be.

Comparing analysis views from these standards will allow researchers to carry out comparative studies of multiple alternative analysis for the same attributes, and then to publish their studies and recommendations, so that industry has the necessary information on which to base their selection of an analytical method appropriate to their needs. We have no intention of proposing a specific Sensor Network information analysis framework in this paper, even though it would be desirable to do so, but instead we aim to provide a better understanding of two analytical methods, in order to help Sensor Network information engineers obtain accurate, repeatable, and reproducible analysis results.

2. Related Work on Analysis Concepts and Terminology

2.1. Metrology

The domain of knowledge referred to as metrology forms the foundation for the development and use of analysis instruments and analysis processes in the sciences and in engineering.

While metrology has a long tradition of use in, for example, physics and chemistry, it is rarely referred to in the Sensor Network information analysis literature. A notable exception in the Sensor Network information engineering literature is NIST (National Institute of Standards and Technology), which investigated “the underlying question of the nature of IT metrology” in 1996, and identified “opportunities to advance IT metrology.” NIST proposed, for instance, “logical relationships between metrology concepts,” consisting of four steps to follow to obtain measured values: defining quantities/attributes, identifying units and scales, determining the primary references, and settling the secondary references. In addition, in 1999, Gray discussed the applicability of metrology, and the necessity of applying it, from the Sensor Network information analysis point of view: “We are still perhaps on the eve of giant steps in the new century for information technology. We will still need better analysis and more uniformity, precision, and control to achieve these giant steps.” Since then, metrology has been used for the design of the COSMIC analytical method, and is also addressed in [2].

2.2. Analysis Definitions for the Practical View

While in the Sensor Network information engineering literature, analysis is often defined as a mapping between two structures, this does not give sufficient information about how to measure in practice. It was pointed out in [2] that it is necessary to move beyond the theoretical definition of the mapping to an operational procedure, as described in the vocabulary of the VIM [1] and modeled with a transition through three levels.

An analysis principle forms the scientific basis of analysis. For Sensor Network information entities (products), the analysis principle involves the model(s) used as a basis to describe the concept that is related to a concept to quantify, and which can be
quantified by an analytical method. The idea is that modeling, as a central notion in Sensor Network information products, should be considered at the same level as scientific principles in other sciences and in engineering [2].

2.3. Base Quantity and Analytical Method

To adequately quantify a concept, an analytical method is required, which itself must include a coherent set of definitions and analysis rules, as well as a base unit specific to the analytical method as described in the VIM. The Fig.1 shows the analysis foundations.

Fig. 1. Analysis foundations.

A base unit is “an analysis unit that is adopted by convention for a base quantity” [1]. There is only one base unit for each base quantity.

An analytical method is a generic operational description, i.e. a description of a logical sequence of operations for performing an analysis activity, for moving on from the concept to quantify to the value representing the analysis result [5].

An analysis procedure is a set of operations, described specifically and used in the performance of particular analysis according to a given method [6].

An analytical method should be implemented concretely by some concrete operations achieved through measuring instruments and/or practical operations: selection, counting, calculation, comparison, etc. This description of an analysis according to one or more analysis principles and to a given analytical method is called the analysis procedure, which is more specific, more detailed, and more closely related to the environment and to the measuring instruments (e.g. tools) than the method, which is more generic.

Fig. 2 gives examples of a base quantity, a base unit, a concept to quantify, and a measurable concept.

Note that the term metrics is avoided in the definitions above: although it is widely used in Sensor Network information engineering, its use causes ambiguity, and possibly confusion, by suggesting erroneous analogies; therefore, this term is not used in this text.

Fig. 2. Examples of some analysis terms in sensor network information engineering.

2.4. Vocabulary Issues in ISO 15939

Fig.3 shows a base quantity.

In 2002, the ISO documented and adopted a generic model for the analysis process in Sensor Network information organizations in ISO 15939 (revised in 2007). Specifically, ISO 15939 “identifies the activities and tasks that are necessary to successfully identify, define, select, apply, and improve Sensor Network information analysis within an overall project or organizational analysis structure” [6]. It also provides “definitions for analysis terms commonly used within the Sensor Network information industry,” using the VIM as its base, although with some tailoring of the terminology to facilitate its acceptance within the Sensor Network information engineering community.

In ISO 15939, a base measure is “a measure defined in terms of an attribute and the method for quantifying it” [5].

To obtain a base measure in practice, an analytical method must be applied to an attribute of an entity (i.e. an object which is itself a model of an object). In the VIM, an analytical method is defined as “a generic description of a logical organization of operations used in an analysis” [5], while in ISO 15939, this definition has been tailored as follows: “a logical sequence of operations, described generically, used in quantifying an attribute with respect to a
specified scale” [7]. Both definitions consider “a logical sequence/organization of operations,” and from this perspective they are similar.

In ISO 15939, the attribute is a property of an entity. In [2], the entity refers to “the concept to quantify,” which should be related to a base unit [1]. The expression “base unit” cannot refer directly to the expression “base measure,” since a base unit is a part of an analytical method with rules and conventions designed to obtain a “base quantity.” In other words, a base quantity is a combination of a number from the numerical world and a base unit established by convention. For example, by international convention in the SI, the base quantity of length is composed of a number associated with the base unit “meter”. To date, there has been little work done to define base units in Sensor Network information engineering, including base units for the analysis of the quality of Sensor Network information. In Sensor Network information analysis, the COSMIC functional size analytical method in ISO 19761 is unique, in the sense that it has explicitly defined its base unit, referred to as “a data movement of a single data group,” and its corresponding analysis symbol, “CFP”. With this definition, a COSMIC analysis can be expressed as a base quantity in the metrology sense with a number of base units (for example, 15 CFP, 27 CFP, etc.).

In Sensor Network information engineering, the term time may refer to the number of months representing the base quantity for expressing the concept of effort (as in productivity: Sensor Network information delivered per work effort unit, measured in person-months), or it may refer to the concept of duration (often measured in calendar-months). The interpretation of the analysis unit “month” will differ, depending on the concept to be represented and measured (e.g. person-months for the concept of human effort, and calendar-months for the concept of project duration).

For the analysis of Sensor Network information quality, the analysis units of quality concepts, like faults, errors, defects, failures, etc., also need to be explicitly, and uniquely, defined. There are other similarities and differences in the terms used in the ISO 15939 and VIM vocabularies:

1) The concepts of fault, error, and defect:
   a) An error is defined as “a human action that produces an incorrect result, such as Sensor Network information containing a fault”;
   b) A failure is defined as “an event in which a system or system component does not perform a required function within specified limits”;
   c) A fault is defined as “a manifestation of an error in Sensor Network information”;
   d) A defect is defined as “a problem which, if not corrected, could cause an application either to fail or to produce incorrect results.

Each of these definitions refers to another definition (e.g., an error contains a fault, and a fault is a manifestation of an error), which adds to the difficulty of quantifying concepts like these, and, based on these definitions, of obtaining accurate analysis results.

The terms “concept to quantify” and “base quantity” are used in [1] instead of “attribute” and “base measure” respectively in [8].

Furthermore, if “defect” had been selected as a base unit in [2], this would mean that the concept of the defect was used both a base unit and a base quantity. It would be like using “length” and “meter” as a single concept, while they are clearly distinct concepts, but, of course, related in a well-organized relationship: “length” is the concept to be quantified, and “meter” is the base unit used for quantifying “length”.

In conclusion, the definitions of “attribute” and “base measure” do not refer to an explicit definition of a base quantity, or to an explicit corresponding analysis unit. The proposal in [2] to use an explicit process to design an analytical method with its base measure (and corresponding base unit) can help improve analysis in Sensor Network information engineering, but the Sensor Network information engineer cannot expect the same precision in the short term as that provided by the International System of Units.

2.5. Vocabulary in ISO 25021

ISO 25021 is a part of the ISO 25000 series, which is being published to update the ISO 9126 series. ISO 25021 has adopted a different vocabulary. For example, the new expression “Quality Measure Element (QME)” was substituted for “base measure”, and “property to quantify”, for the term “attribute”.

According to ISO 25021, the user of the analytical method shall identify and collect data related to the property to quantify. Depending on the context of usage and objective(s) of the Quality Measure Element (QME), a number of properties and sub properties can be identified. These properties constitute the input to the design of the analytical method, and are extracted and defined from the artifacts of the Sensor Network information (e.g., documentation, code). This process 2 is similar to the ISO 15939 process, but with a different terminology:
“property to quantify” instead of “attributes”, and “quality measure element” instead of “base measure”.

2.6. Summary Mapping of the Three ISO Reference Documents

The metrology vocabulary (VIM) is adopted here, because it enjoys a wider consensus across the sciences and engineering than the adaptations in the other two documents, which are limited to the Sensor Network information engineering community.

3. The Analytical Method

In this section, the four steps recommended for designing an analytical method for a base quantity are described in more detail [2]. In Section 4, we analyze, using these steps, the design of two base quantities related to the quality of the Sensor Network information.

To obtain a base quantity [1], it is not only necessary to apply an analytical method to the measurable concept, but also to use the base unit in the analytical method, and to identify and define that base unit if this has not already been done. Now, when measuring in practice, an analysis procedure should be documented as a distinct activity. This is because the analysis procedure used to obtain the analysis result (i.e. a base quantity) in a specific environment is required in order to instantiate the analytical method (e.g., a procedure to determine the functional size of a project using the COSMIC analytical method with use cases).

The four steps recommended by Abran in [2] to design an analytical method are:
1) Determine the analysis objectives;
2) Characterize the concept (and the sub-concepts) to be quantified;
3) The word “process” is used because the references suggest a number of steps.
4) Design the meta model (of the relationships among the sub-concepts);
5) Define the numerical assignment rules.

These steps can also help to verify the design of the analytical method for a specific base quantity. As well, they can be applied to specify or improve the design of analytical methods for many of the base quantities 3 embedded in the metrics proposed in ISO 9126.

3.1. Determine the Analysis Objectives

The first step is to identify the objectives for measuring the base quantity. In ISO 9126, these objectives are related to the quality characteristics and sub characteristics to be measured. The analysis context determines the type of user of the base quantity, the life cycle phase in which it will be used, and the number of constraints to using it when the information is available.

3.2. Characterize the Concepts (and Sub-Concepts) to be Quantified

Sensor Network information is often perceived as an intangible product, but one that can be made visible through multiple representations: a set of screens and reports for a user, a set of lines of code (or executable statements) for a programmer, and a set of Sensor Network information model representations for a Sensor Network information designer are some examples.

Characterization can be achieved by first stating explicitly how the concept (e.g., defects in the Sensor Network information documentation) to be quantified (e.g., defect in the Sensor Network information documentation) is decomposed into sub-concepts (e.g., how defects in the Sensor Network information documentation are decomposed into sub-concepts).

Knowledge about the objective should determine what information should be included in the quantification of the concepts to be measured, or excluded from it, in terms of sub-concepts. Moreover, it is important to care-fully define what is included, as failure to do so can result in sub-concepts that are defined differently being included in the design of analytical methods attempting to measure the same concept. For example, Base Functional Components (BFC) are different in the IFPUG standard and the COSMIC Analysis Manual: IFPUG considers an elementary process (such as an IFPUG Input or Output) as a BFC, while COSMIC considers a data movement as a BFC. This makes it challenging to compare the results of these analytical methods.

3.3. Design the Meta Model

Defining concepts and sub concepts is only one part of the method for characterizing them. It is also necessary to apply principles and set rules. Principles link the compliance of a specific concept (or sub-concept) to its definition. For example, an entry data movement in the COS-MIC analytical method “shall not exit data across the boundary, or read or write data.” Rules help to confirm the status of a concept (or sub-concept) in a particular situation. For example, the trigger (a sub-concept) of an entry data movement could be the internal clock of a computer, even though it is generated periodically by hardware.

Having defined the sub concepts related to the concept to be quantified, the next step is to construct the meta-model of the analytical method.

The meta-model is constructed based on the sub-concepts of the concept to be quantified. The relationships (or roles) between that concept and the sub-concepts that represent the Sensor Network information, or part of it, constitute the meta-model. The meta model describes how to recognize the concept(s) and/or sub concepts in the analytical method. For example, definitions, principles, and
rules are described in detail in the COSMIC Analysis Manual for determining the functional size of requirements in the COSMIC analytical method.

A generic meta model should not be specific to any particular Sensor Network information, and must be independent of the specific context of the analysis, i.e. how the Sensor Network information is implemented (unless it is what we want to measure). For example, in the analysis meta model of ISO 19761 (COSMIC), the functional user enters and receives data that are read and written by Sensor Network information. This meta model, which shows the relationships between the sub concepts (i.e. users, type of data movement – entry, exit, read, or write) of the Sensor Network information that use different physical components (I/O hardware, computation hardware, and storage hardware. It should also identify the measure and (input).

Each type of data movement in the COSMIC analytical method rules is considered as an input (i.e. the measure and) to be taken into account in the analysis process.

3.4. Define the Numerical Assignment Rules

Assigning numerical rules is part of the process of designing an analytical method. A numerical assignment rule can be described from a practitioner’s point of view (generally text) or from a theoretical point of view (generally a mathematical expression).

A quantity should be associated with a scale type [2]. Only certain operations can be performed on certain scales of analysis, and the mathematical algorithm proposed by an analytical method must conform to those operations. For example, differences between two ordinal values cannot be quantified; therefore, adding ordinal numbers is not allowed. When the scale types are not taken into account accurately, the quantities obtained could be wrongly interpreted.

The purpose of the analysis determines the usage of the base quantity and which base unit should be used. This affects the definition of the numerical assignment rules. For example, to obtain the number of COSMIC function points, it is necessary to identify a base unit. In COSMIC, the base unit is defined as a data movement that is related to different types of data movement (Entry, Read, Write, and Exit) within a functional process.

4. Present Method in the Data Integration Procession

Existing methods to software integration of information for TTT, process cover a wide spectrum of data encoding methods and search or matching algorithms. The encoding methods differ with respect to their soundness, completeness, and the extent to which they support an estimate of the effort it takes to modify a data. Text-based encoding and integration is neither sound nor complete. Its disadvantages have been thoroughly in the information integration literature [5, 6]. Lexical descriptor-based encoding method also suffers from a number of problems about developing and using classification vocabulary [7]. Software specific challenges include the fact that one-word or one-phrase abstractions are hard to come by in the software domain [8]. From the user’s point of view, lack of familiarity with the vocabulary is also pointed out as draw back in using a integration of information for TTT, system effectively [9]. In this context Data feature model will be a promising solution for integration of information for TTT process.

4.1. Methods Used

It is an algebraic model in which documents and queries are represented as data features as follows:

\[ d_j = (w_{1,j}, w_{2,j}, \ldots, w_{t,j}) \]
\[ q = (w_{1,q}, w_{2,q}, \ldots, w_{t,q}) \]

Each dimension corresponds to a separate term. If a term occurs in the document, its value in the data feature is non-zero. An indexed collection of documents is represented as a term table which has documents as fields and words as primary key for row. The \((D)i (Word)j\)-th entry of this table records how many times the \(j\)-th search term appeared in the \(i\)-th document. Fig. 4 shows a sample data feature model.

![Fig. 4. A sample data feature model.](image)

The first major data of a data feature space search model is the concept of a term space. A term space consists of every unique word that appears in a collection of documents. The second major data of a data feature space search model is term counts. Term counts are simply records of how many times each term occurs in an individual document. By using the term space as a coordinate space, and the term counts as coordinates within that space, we can create a data feature for each document. As the number of terms increases, the dimensionality of DFM also increases.
For these words documents and corresponding ranks will be stored in the rank table. Based on the ranking terms are compared as “ranked higher than”, “ranked lower than” or “ranked equal to” the second, making it possible to evaluate complex information according to query criteria. Here search data feature space search model ranks the documents it finds according to the estimation of their relevance, making it possible for the user quickly to select the data unit according to their requirements [7].

Relevancy rankings of documents in a keyword search can be calculated, using the assumptions of document similarities theory, by comparing the deviation of angles between each document data feature and the original query data feature where the query is represented as same kind of data feature as the documents.

It is easier to calculate the cosine of the angle between the data features instead of the angle:

\[
\cos \theta = \frac{d_2 \cdot q}{\|d_2\| \|q\|}
\]

A cosine value of zero means that the query and document data feature are orthogonal and have no match (i.e. the query term does not exist in the document being considered).

4.2. The Document Classification Algorithm

KSP classifier is an instance-based learning algorithm that is based on a distance function for pairs of observations, such as the Euclidean distance or Cosine. The K-Shortest Path (KSP) classifier algorithm has been studied extensively for text categorization by Yang and Liu [6]. In this classification paradigm, k shortest paths of a training data are computed first. Then the similarities of one sample from testing data to the k shortest paths are aggregated according to the class of the paths, and the testing sample is assigned to the most similar class. The similarity in score of each path document to the test document is used as the weight of the categories of the path document [8]. If there are several training documents in the k shortest path, which share a category, the category gets a higher weight. In this work, we used the Cosine distance to calculate the similarity score for the document representation.

One of advantages of KSP is that it is well suited for multi-modal classes as its classification decision is based on a small path of similar objects (i.e., the major class). So, even if the target class is multi-modal (i.e., consists of objects whose independent variables have different characteristics for different subsets), it can still lead to good accuracy. A major drawback of the similarity measure used in KSP is that it uses all features equally in computing similarities. This can lead to poor similarity measures and classification errors, when only a small subset of the features is useful for classification [5].

Steps for KSP Using Average Cosine:

Step 1: Select k nearest training documents, where the similarity is measured by the cosine between a given testing document and a training document.

Step 2: Using cosine values of k shortest paths and frequency of documents of each class i in k shortest paths, compute average cosine value for each class i, Avg_Cosine (i).

Step 3: Classify the testing document a class label which has largest average cosine.

In order to reduce the dimensionality of DFM and keep useful information, we first compute concept data features for given categories. Then, using the concept data features as projection matrix, projection of both training and testing data is done. Finally, we apply KSP algorithm on the projected DFM model that has reduced dimensionality.

Steps of Combined Method for Data feature Based Algorithm and K-Shortest Path Algorithm:

Step 1: Compute a concept data feature for each category using true label information of training documents and then construct concept data feature matrix C (w-by-c), where c is the number of categories.

Step 2: Do projection of DFM model A (w-by-d) using concept data feature matrix C (w-by-c) (i.e., \(C^T \cdot A\)).

Step 3: Apply KSP with the projected DFM model (i.e., c-by-d matrix).

5. Analysis of the Designs of Two Sensor Network Information Analytical Methods

There are hundreds of definitions of Sensor Network information metrics in the Sensor Network information engineering literature, but only a few attempts have been made to provide comprehensive definition of a measurable concept for an analytical method. We have chosen two designs of Sensor Network information analytical methods, because their definitions are documented and are both related to Sensor Network information quality:

1) The analytical method for code, from Munson and Nikora;

2) The analytical method for the size of “use cases from the documentation”4 in [7].

Using the analysis concepts and criteria in [2], it is possible to determine whether or not the analytical method proposed for the concept to be quantified is complete. This section discusses how each example fulfills the requirements for each step in the design of an analytical method.
5. Discussion

In the Sensor Network information engineering literature, few references focus on the definition and design of Sensor Network information analytical methods. Among those that do are Munson, et al. and the ISO standards on functional size analysis. To avoid inconsistent vocabulary and potentially incorrect interpretation of data, Sensor Network information analytical methods must be better designed, including definitions, analysis principles, analysis rules, and base units.

Well-designed analytical methods are necessary for each of the 80 base analysis embedded within the 250 or more derived analysis referenced in ISO 9126, in particular those related to: defect, fault, error, failure, error message, warning message, illegal operation, data correction, and fault pattern. Many others could be designed for use in conjunction with the base analysis related to quality aspects like memory size, effort, duration, and size of the product.

References


