Hindawi
Mathematical Problems in Engineering
Volume 2018, Article ID 8058670, 6 pages
https://doi.org/10.1155/2018/8058670

Research Article

Big Data Validity Evaluation Based on MMTD

Ningning Zhou,[1] **Guofang Huang**,[2] **and Suyang Zhong**[1]

[1]*School of Computer, Nanjing University of Posts and Telecommunications, Nanjing 210023, China*
[2]*State Key Laboratory of Smart Grid Protection and Control, Nanjing 211106, China*

Correspondence should be addressed to Ningning Zhou; zhounn@njupt.edu.cn

Received 7 November 2017; Revised 23 March 2018; Accepted 10 April 2018; Published 10 June 2018

Academic Editor: Ester Zumpano

Big data has been studied extensively in recent years. With the increase in data size, data quality becomes a priority. Evaluation of data quality is important for data management, which influences data analysis and decision making. Data validity is an important aspect of data quality evaluation. Based on 3V properties of big data, dimensions that have a major influence on data validity in a big data environment are analyzed. Each data validity dimension is analyzed qualitatively using medium logic. The measuring of medium truth degree is used to propose models to measure single and multiple dimensions of big data validity. The validity evaluation method based on medium logic is more reasonable and scientific than general methods.

1. Introduction

Big data has been studied extensively in recent years and several investigations have focused on the big data phenomenon[1–7]. The top international journals 'Nature'[8] and 'Science'[9], respectively, in 2008 and 2011, took 'big data' and 'dealing with data' as the topic, which made people explore the enthusiasm of big data. However, there is no universal definition of big data in academia. In a literal sense, the most fundamental nature of big data lies in the large data size, but it also involves a high degree of complexity associated with data collection, management, and processing. The "big" of big data is mainly reflected in three aspects [10–12]: (1) data volume is large (Volume); (2) the complexity of data type is high (Variety); (3) data flow, especially the generation of information flow in Internet, is fast (Velocity). The 3V properties have now been widely accepted to describe big data. Some people will also express the potential huge value of Value into it, so that 3V is extended to 4V.

Although big data is valuable, it is a challenge to unlock the potential from the large amount of data [13]. High quality is a prerequisite for unlocking big data potential since only a high-quality big data environment yields implicit, accurate, and useful information that helps make correct decisions. Even state-of-the-art data analysis tools cannot extract useful information from an environment fraught with "rubbish" [14, 15]. However, it is difficult to maintain high quality because big data is varied, complicated, and dynamic. This highlights a need for the analysis and evaluation of big data quality while constructing a high-quality big data environment.

Data quality involves many dimensions that include data validity, timeliness, fuzziness, objectivity, usefulness, availability, user satisfaction, ease of use, and understandability. Data validity is particularly important in the evaluation of data quality. It is a priority due to the massive data size, increased demand for data processing, and broad variety of data types. However, few studies have been done on the evaluation of data validity[16, 17]. Wei Meng proposed to measure data validity using the update frequency [18]. Update frequency of date is a dimension of the quality of data. However, this dimension reflects the novelty of the data rather than the validity. Qingyun et al. proposed to evaluate data validity by formulating a constraint in the dataset [19]. The constraint of evaluating data validity is whether it is in a range compliant with the truth or not. This constraint is one of the dimensions of data validity, but it is not comprehensive. In [20], Jie et al. proposed to devise constraints using three rules (i.e., static, transaction, and dynamic) and they evaluated data validity by measuring the degree to which the rules were satisfied. The method for data validity evaluation varies with the application. It focused on the restricting rules on GIS, but it is too special and it is not general. Moreover, due to the

special attributes of big data, these methods are not entirely suitable for big data. To the best of our knowledge, there is no method for qualitative and quantitative analysis of big data validity.

In this paper, first, we comprehensively analyze dimensions that have a major influence on data validity based on the 3V properties of big data. Data validity refers to the level of need that users or enterprise have for data. Completeness, correctness, and compatibility are particularly serious in a big data environment and become the primary factors that affect data validity. Hence, big data validity is measured in this paper from the perspectives of completeness, correctness, and compatibility. It is used to indicate whether data meets the user-defined condition or falls within a user-defined range. Next, a qualitative analysis of each dimension of data validity is performed using medium logic. Finally, the measure of medium truth degree (MMTD) is used to propose models to measure single and multiple dimensions of big data validity. Our Model for measuring one dimension of big data validity is based on medium logic. Logical correctness ensures that the evaluation results are more reasonable and scientific.

2. Overview of Medium Mathematics Systems

Medium principle was established by Wujia Zhu and Xi'an Xiao in 1980s who devised medium logic tools [21] to build the medium mathematics system, the corner stone of which is medium axiomatic sets [22].

2.1. Notations for Medium Mathematics Systems. In medium mathematics system [21], predicate (concept or property) is represented by P; any variable is denoted as x, with x completely possessing property P being described as P(x). The "¬ "symbol stands for inverse opposite negative and it is termed as "opposite to". The inverse opposite of predicate is denoted as ¬ P. Then the concept of a pair of inverse opposite is represented by both P and ¬ P. Symbol "~"denotes fuzzy negative which reflects the medium state of "either or" or "both this and that" in opposite transition process. The fuzzy negative profoundly reflects fuzziness; "$\prec$" is a truth-value degree connective which describes the difference between two propositions.

2.2. Measuring of Medium Truth Degree

2.2.1. Measuring of Individual Medium Truth Degree. According to the concept of super state[23], the numerical value area of generally applicable quantification is divided into five areas corresponding to the predicted truth scale, namely ¬ ^{+}P, ¬ P, ~P, P, and ^{+}P. In "true" numerical value area T, α_T is ε_T standard scale of predication P; In "false" numerical value area F, $\alpha_F$is ε_F standard scale of predicate ¬ P. $f(x)$ is an arbitrary numeric function of variable x. According to the numeric interval of $f(x)$, the distance ratio function h_T(or h_F) which can scale the individual truth degree is defined. Adopting the concept of distance and using length of numerical value interval to different predicate truth as norm, the distance ratio function is defined, and from this the individual truth degree function is established as follows [23].

For $f(X)\to R$ and $y= f(x) \in f(X)$, the distance ratio $h_T(y)$ which relates to P is

$$h_T(y) = \begin{cases} \dfrac{-d(y, \alpha_F - \varepsilon_F)}{d(\alpha_T - \varepsilon_T, \alpha_F - \varepsilon_F)} & y < \alpha_F - \varepsilon_F \\ 0 & \alpha_F - \varepsilon_F \le y \le \alpha_F + \varepsilon_F \\ \dfrac{d(y, \alpha_F + \varepsilon_F)}{d(\alpha_T - \varepsilon_T, \alpha_F + \varepsilon_F)} & \alpha_F + \varepsilon_F < y < \alpha_T - \varepsilon_T \\ 1 & \alpha_T - \varepsilon_T \le y \le \alpha_T + \varepsilon_T \\ \dfrac{d(y, \alpha_F + \varepsilon_F)}{d(\alpha_T + \varepsilon_T, \alpha_F + \varepsilon_F)} & y > \alpha_T + \varepsilon_T. \end{cases} \tag{1}$$

For $f(X)\to R$ and $y= f(x) \in f(X)$, the distance ratio $h_F(y)$ which relates to ¬ P is

$$h_F(y) = \begin{cases} \dfrac{-d(y, \alpha_T + \varepsilon_T)}{d(\alpha_T + \varepsilon_T, \alpha_F + \varepsilon_F)} & y > \alpha_T + \varepsilon_T \\ 0 & \alpha_T - \varepsilon_T \le y \le \alpha_T + \varepsilon_T \\ \dfrac{d(y, \alpha_T - \varepsilon_T)}{d(\alpha_T - \varepsilon_T, \alpha_F + \varepsilon_F)} & \alpha_F + \varepsilon_F < y < \alpha_T - \varepsilon_T \\ 1 & \alpha_F - \varepsilon_F \le y \le \alpha_F + \varepsilon_F \\ \dfrac{d(y, \alpha_T - \varepsilon_T)}{d(\alpha_T - \varepsilon_T, \alpha_F - \varepsilon_F)} & y > \alpha_F - \varepsilon_F \end{cases} \tag{2}$$

where $d(a,b)$ is the Euclidean distance.

The bigger the value of $h_T(y)$ is, the higher the individual truth degree related to P is. The bigger the value of $h_F(y)$ is, the higher the individual truth degree related to ¬ P is.

2.2.2. Measuring of Set Medium Truth Degree. $f{:}X\to R^n$ is the n-dimensional numerical mapping of the set X. The measuring of truth scale of disperse set X which relates to P (or ¬ P) can be scaled by the additivity of the truth scale [23, 24]$h_{nT-S}(y_i)$ (or $h_{nF-S}(y_i)$) and the average additivity of the truth scale [23, 24]$h_{nT-M}(y_i)$ (or$h_{nF-M}(y_i)$) of set which relates to P (or ¬ P).

When $y_i = (f_1(x_i), f_2(x_i), \ldots, f_n(x_i)) = (y_{i1}, y_{i2}, \ldots, y_{in}) \in f(X)$, the additivity of the truth degree of disperse set X which relates to P is

$$h_{nT-S}(y_i) = \sum_{k=1}^{n} (h_T(y_{ik})). \tag{3}$$

The average additivity of the truth degree of disperse set X which relates to P is

$$h_{nT-M}(y_i) = \frac{1}{n}\sum_{k=1}^{n} (h_T(y_{ik})). \tag{4}$$

The additivity of the truth degree of disperse set X which relates to ¬ P is

$$h_{nF-S}(y_i) = \sum_{k=1}^{n} (h_F(y_{ik})). \quad (5)$$

The average additivity of the truth degree of disperse set X which relates to ¬ P is

$$h_{nF-M}(y_i) = \frac{1}{n}\sum_{k=1}^{n} (h_F(y_{ik})). \quad (6)$$

3. Qualitative Analysis of Big Data Validity

Data validity refers to the degree of data demand for users or enterprises. It is used to describe whether data satisfies user-defined conditions or falls within a user-defined range.

3.1. Selection of Dimension for Big Data Validity Evaluation. A large amount of incompatible data is generated due to the 3V properties of big data. Furthermore, data correctness and completeness can be compromised during generation, transmission, and processing. These problems are particularly serious in a big data environment and become the primary factors that affect data validity. Hence, big data validity is measured in this paper from the perspectives of completeness, correctness, and compatibility.

3.2. Dimensions of Big Data Validity

3.2.1. Data Completeness. In Cihai (an encyclopedia of the Chinese language), completeness refers to the state where components or parts are maintained without being damaged. In the Collins English Dictionary and Oxford Dictionary, completeness is defined as the state including all the parts, etc., that are necessary: whole. In the 21st Century Unabridged English-Chinese Dictionary, completeness means including all parts, details, facts, etc. and with nothing missing.

A universal definition of big data completeness is lacking. In the context of a specific application, big data completeness can be defined as follows.

Definition 1. If data has n properties and each property has all necessary parts, it is regarded as complete. Otherwise, it is incomplete.

Definition 2. Completeness refers to the degree to which data is complete. It is denoted by *C1*.

Let $R_1, R_2, \ldots R_n$ denote the n data properties and $V(R_i)$ denote the completeness of property R_i. Note that R_i has different forms for different applications. For example, the completeness of a property is zero if the property value is missing for some data, and 1 otherwise. Hence, R_i can be defined as

$$V(R_i) = \begin{cases} 0, & R_i \text{ missing} \\ 1, & R_i \text{ exists.} \end{cases} \quad (7)$$

The importance of each data property varies with the application. Let $w_1, w_2 \ldots w_n$ denote the weights for n properties in an application, where

$$\sum_{i=1}^{n} w_i = 1. \quad (8)$$

Consider data with n properties; its completeness is computed as the weighted sum of the completeness of all its properties.

$$C1 = \sum_{i=1}^{n} V(R_i) \times w_i. \quad (9)$$

3.2.2. Data Correctness. In Cihai, correctness refers to compliance with truth, law, convention, and standard, contrary to "wrongness". In the Collins English Dictionary and Oxford Dictionary, correctness is defined as accurate or true, without any mistakes. In the 21st Century Unabridged English-Chinese Dictionary, completeness means accurate, compliant with truth, and having no mistakes.

Currently, there is no universal definition for data correctness in the field of big data. Whether data is correct and the degree to which data is correct are defined as follows from the perspective of the application.

Definition 3. Consider data with n properties. If each property is compliant with a recognized standard or truth, it is regarded as correct. Otherwise, it is incorrect.

Definition 4. Correctness refers to the degree to which data is correct. It is denoted by C2.

Let $R_1, R_2, \ldots R_n$ denote the n data properties and $Z(R_i)$ denote the correctness of property R_i. If the value of R_i is in a range compliant with the truth, the correctness of this property is 1. Otherwise, it is 0. The correctness of the property, $Z(R_i)$, is defined as

$$Z(R_i) = \begin{cases} 1, & R_i \in \mathrm{dom}(R_i) \\ 0, & R_i \notin \mathrm{dom}(R_i) \end{cases} \quad (10)$$

where $\mathrm{dom}(R_i)$ denotes the range of R_i.

Data correctness C2 is computed as the weighted sum of each property:

$$C2 = \sum_{i=1}^{n} Z(R_i) \times w_i \quad (11)$$

where w_i denotes the weight of each property in the application and satisfies (8).

3.2.3. Data Compatibility. In Cihai, compatibility refers to coexistence without causing problems. In the 21st Century Unabridged English-Chinese Dictionary, compatibility means that ideas, methods, or things can be used together. In the case of big data, data compatibility is defined as follows.

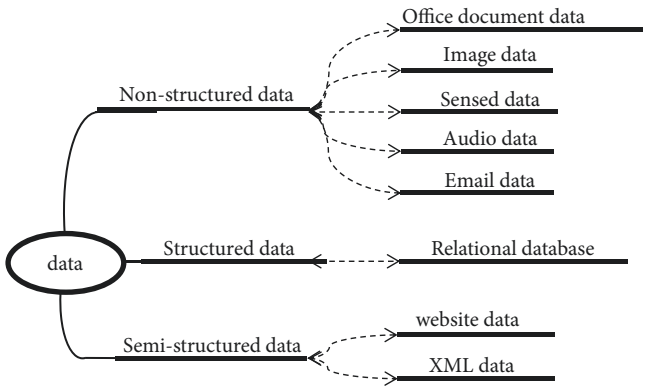

Figure 1: Data types in a big data environment.

Definition 5. If a group of data is of the same type and describes the same object consistently, the data is regarded as compatible with one another; otherwise, it is mutually exclusive.

Definition 6. Compatibility C3 refers to the degree to which a group of data is compatible with one another. Compatibility C3 is defined as

$$C3 = 1 - \frac{d_t}{d_a} \tag{12}$$

where d_a denotes the total amount of data in the group and d_t denotes the amount of incompatible data in the group.

4. Medium Truth Degree-Based Model for Measuring Big Data Validity

4.1. Data Normalization. Data variety is a significant aspect of big data. In addition to traditional structured data, a large amount of nonstructured and semistructured data has been generated by advances in the Internet and the Internet of Things (IoT). Examples include website data, sensed data, audio data, image data, and signal data, as shown in Figure 1. While this enriches content, it is more challenging to store, analyze, and evaluate data. Data needs to be normalized before appropriately evaluating big data validity.

Structured and nonstructured data in a big data environment have different content, forms, and structures, so they cannot be managed uniformly. Hence, a data model needs to be developed to provide a uniform description of both structured and nonstructured data.

Based on [25], a tetrahedron data model is proposed for nonstructured data. The proposed model consists of four parts: basic property, semantic feature, bottom-layer feature, and original document. In order to process structured and nonstructured data uniformly, a new part of data type is introduced to describe document type. Consider an audio document as an example of nonstructured data. Its document type belongs to audio document. Its basic property includes document name and intuitive information on document size and creation time. Its semantic feature is the information in the document. The bottom-layer feature is audio frequency and bandwidth. As for structured data, it does not have a basic property, semantic feature, or bottom-layer feature. It is thus directly stored in the original document. Semistructured data like an XML document has some structured data, which is dynamic. Hence, it is difficult to store these data by constructing a mapping table. Fortunately, these data can be extracted to form a string, enabling them to be stored in the database like structured data.

In this manner, structured and nonstructured data can be stored in the database uniformly. For nonstructured data like an image, the content can be analyzed using a description of the image in terms of the basic property, semantic feature, and bottom-layer feature. Structured and semistructured data can be analyzed directly.

4.2. Determination of Logical Predicate and True-Value Range

4.2.1. Determination of Logical Predicate. In order to evaluate data completeness, correctness, and compatibility, let the predicate W denote the high degree, ¬ W low degree, and transition ~W. The correspondence between numerical range and predicates is shown in Figure 2.

4.2.2. Determination of Logical Interval. Weights need to be allocated to the completeness and correctness of data in an application. Data usefulness will not be compromised as long as the major property exists, even if the subordinate property is missing. Based on the proportions of major and subordinate properties, values A and B are computed as follows:

$$B = \sum_{i=1}^{m} w_i$$
$$A = 1 - B \tag{13}$$

where w_i denotes weight and m denotes the largest weight of subordinate properties. Assume that the weights of n properties are sorted in descending order as follows: $w_1 \leq w_2 \leq \cdots \leq w_m \leq w_{m+1} \cdots \leq w_n$, where $w_1,\ldots,w_m$ denote weights of subordinate properties and $w_{m+1},\ldots,w_n$ denote weights of major properties. The value of m is determined as follows. Sort all weights and compute the sum of weights starting with the smallest weight w_1 until the sum of weights is no larger than the weight w_{m+1}, as shown in

$$\sum_{i=1}^{m} w_i \leq w_{m+1}. \tag{14}$$

4.3. Model for Measuring One Dimension of Big Data Validity. The weight of each property in each dimension of the data is first determined to obtain the correspondence between the numerical range of one dimension and the logical predicates: high degree, low degree, and transition, as shown in Figure 2. The distance ratio function $h_T(C)$ with respect to W is selected as the model to measure completeness:

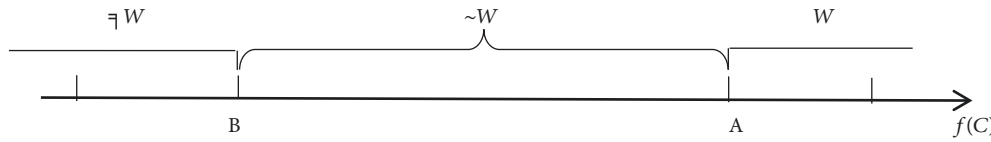

Figure 2: Correspondence between numerical range and predicates.

$$h_T(C) = \begin{cases} 0 & f(C) \le B \\ \dfrac{f(C) - B}{A - B} & B < f(C) < A \\ 1 & f(C) \ge A \end{cases} \quad (15)$$

where f(C) is defined as in (9), (11), and (12). Use the completeness measuring model as an example for the analysis. f(C) in (15) is C1 in (9) and the completeness measuring model is $h_T(C1)$. If the value of data completeness is in the false range (low degree of logic truth¬ W), the value of data completeness is 0 and means that data is missing. If the value of data completeness is in the true range (high degree of logic truth W), the value of data completeness is 1 and means that data is complete. If the value of data completeness is in the transition range (medium degree of logic truth ~W), the value of data completeness is between 0 and 1; closer to 1 means more complete data, and closer to 0 means more missing data.

The model for measuring data correctness or compatibility is similar to the model for completeness. The model measures data correctness $h_T(C2)$ when f(C) in (15) is C2 in (11) and measures data compatibility $h_T(C3)$ when f(C) in (15) is C3 in (12).

4.4. Multidimension Model for Measuring the Integrated Value of Big Data Validity. For a set of K data, completeness and correctness can be measured by the average additive truth scales $h_{kT-M}(C1)$ and $h_{kT-M}(C2)$ which are defined as

$$h_{kT-M}(C1) = \frac{\sum_{i=1}^{K} h_T(C1(i))}{K}$$
$$h_{kT-M}(C2) = \frac{\sum_{i=1}^{K} h_T(C2(i))}{K} \quad (16)$$

where C1(i) and C2(i) denote completeness and correctness for each element in the data set, as defined in (9) and (11).

For a data set in a big data application, the integrated value of data validity can be measured by the weighted sum of metric values for each dimension. Hence, an integrated multidimension model H for measuring data validity in a big data application is

$$H = h_{kT-M}(C1) \times W(C1) + h_{kT-M}(C2) \times W(C2) + h_T(C3) \times W(C3) \quad (17)$$

where $h_{kT-M}(C1)$, $h_{kT-M}(C2)$, and $h_T(C3)$ denote completeness, correctness, and compatibility, respectively, and $W(C1), W(C2), W(C3)$ denote the weights of completeness, correctness, and compatibility, respectively, according to certain application. Thus, we have

$$W(C1) + W(C2) + W(C3) = 1. \quad (18)$$

Compared with the tetrahedron evaluation models, the two models have both similarities and differences. The idea of the multidimension model for measuring data validity in a big data application in this paper (17) is similar to the tetrahedron evaluation models, but the difference between these two models lies in the measuring of each dimension. Our model for measuring one dimension of big data validity is based on medium logic. Logical correctness ensures that the evaluation results are more reasonable and scientific.

5. Conclusions

Medium mathematics systems are introduced for the evaluation of big data validity. A medium logic-based data validity evaluation method is proposed. The contributions of this paper are as follows: (1) Based on the 3V properties of big data, dimensions that have a major influence on data validity are determined. (2) Data completeness, correctness, and compatibility are defined. (3) A medium truth degree-based model is proposed to measure each dimension of data validity. (4) A medium truth degree-based multidimension model is proposed to measure the integrated value of data validity. In the future, other factors that influence big data quality will be studied and corresponding measurement models will be developed.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the State Key Laboratory of Smart Grid Protection and Control of China (2016, no. 10) and the National Natural Science Foundation of China no. 61170322, no. 61373065, and no. 61302157.

References

[1] N. Ramakrishnan and R. Kumar, "Big Data," *The Computer Journal*, vol. 49, no. 4, pp. 20–22, 2016.

[2] V. Marx, "Biology: the big challenges of big data," *Nature*, vol. 498, no. 7453, pp. 255–260, 2013.

[3] A. I. Naimi and D. J. Westreich, "Big Data: A Revolution That Will Transform How We Live, Work, and Think," *American Journal of Epidemiology*, vol. 179, no. 9, pp. 1143-1144, 2014.

[4] W. Pan, Q. Yang, C. Aggarwal, and C. Koch, "Big Data," *IEEE Intelligent Systems*, vol. 32, no. 2, pp. 7-8, 2017.

[5] C. L. P. Chen and C. Y. Zhang, "Data-intensive applications, challenges, techniques and technologies: A survey on Big Data," *Information Sciences*, vol. 275, pp. 314–347, 2014.

[6] H. V. Jagadish, J. Gehrke, A. Labrinidis et al., "Big data and its technical challenges," *Communications of the ACM*, vol. 57, no. 7, pp. 86–94, 2014.

[7] I. Anagnostopoulos, S. Zeadally, and E. Exposito, "Handling big data: research challenges and future directions," *The Journal of Supercomputing*, vol. 72, no. 4, pp. 1494–1516, 2016.

[8] F. Frankel and R. Reid, "Big data: Distilling meaning from data," *Nature*, vol. 455, no. 7209, p. 30, 2008.

[9] S. Staff, "Dealing with data. Challenges and opportunities. Introduction," *Science*, vol. 331, no. 6018, pp. 692-693, 2011.

[10] J. Manyika, M. Chui, and B. Brown, *Big data: The next frontier for innovation, competition, and productivity[J]. Analytics*, Big data, The next frontier for innovation, 2011.

[11] M. S. Viktor, *Big data : a revolution that will transform how we live, work, and think*, John Murray, 2013.

[12] A. I. Naimi and D. J. Westreich, "Big Data: A Revolution That Will Transform How We Live, Work, and Think," *Mathematics & Computer Education*, vol. 17, pp. 181–183, 2013.

[13] D. B. Lindenmayer and G. E. Likens, "Analysis: don't do big-data science backwards," *Nature*, vol. 499, no. 7458, article 284, 2013.

[14] S. Bryson, D. Kenwright, M. Cox, D. Ellsworth, and R. Haimes, "Visually exploring gigabyte data sets in real time," *Communications of the ACM*, vol. 42, no. 8, pp. 82–90, 1999.

[15] N. R. Gough and M. B. Yaffe, "Focus issue: Conquering the data mountain," *Science Signaling*, vol. 4, no. 160, pp. 2-3, 2011.

[16] R. H. Moe, A. Garratt, B. Slatkowsky-Christensen et al., "Concurrent evaluation of data quality, reliability and validity of the Australian/Canadian Osteoarthritis Hand Index and the Functional Index for Hand Osteoarthritis," *Rheumatology*, vol. 49, no. 12, Article ID keq219, pp. 2327–2336, 2010.

[17] F. Bray and D. M. Parkin, "Evaluation of data quality in the cancer registry: Principles and methods. Part I: Comparability, validity and timeliness," *European Journal of Cancer*, vol. 45, no. 5, pp. 747–755, 2009.

[18] W. Meng, *Research and application of data quality evaluation in data warehouse*, Hebei University of Technology, Tianjin, China, 2004.

[19] Q. Yang, P. Zhao, and D. Yang, "Research on Data Qulity Assesment Methodology," *Computer Engineering and Applications*, vol. 40, no. 9, pp. 3-4, 2004.

[20] Jie. Liang, "The Designing Method of Data Validity Restricting Rule Based on GIS," *Computer Engineering and Applications*, vol. 7, pp. 215–217, 2005.

[21] W. J. Zhu and X. A. Xiao, "Propositional calculus system of medium logic," *Nature*, vol. 8, pp. 315-316, 1985.

[22] X. A. Xiao and W. J. Zhu, "A system of medium axiomatic set theory," *Science in China (A)*, vol. 31, no. 11, pp. 1320–1335, 1988.

[23] L. Hong, X.-A. Xiao, and W.-J. Zhu, "Measure of medium truth scale and its application," *Journal of Computer*, vol. 29, no. 12, pp. 2186–2193, 2006.

[24] L. Hong, X.-A. Xiao, and W.-J. Zhu, "Measure of medium truth scale and its application," *Journal of Computer*, vol. 30, no. 9, pp. 1551–1558, 2007.

[25] B. Lang and B. Zhang, "Key Techniques for Building big-data-oriented Unstructured Data Management platform," *Information technology and Standardization*, vol. 10, pp. 53–57, 2013.