ATLANTIS PRESS

International Journal of Computational Intelligence Systems
Vol. 14(1), 2021, *pp.* 1831–1841
DOI: https://doi.org/10.2991/ijcis.d.210622.001; ISSN: 1875-6891; eISSN: 1875-6883
https://www.atlantis-press.com/journals/ijcis/

Research Article

Research on Comprehensive Evaluation of Data Source Quality in Big Data Environment

Wenquan Li[*], Suping Xu, Xindong Peng

School of Information Engineering, Shaoguan University, Shaoguan, Guangdong, China

ARTICLE INFO

Article History

Received 06 May 2021
Accepted 16 Jun 2021

Keywords

Big data quality
Analytic hierarchy process (AHP)
Entropy method
Comprehensive evaluation
Fuzzy set

ABSTRACT

Data quality is the prerequisite of big data research and the basis of all data analysis, mining, and decision support. Therefore, a comprehensive fuzzy evaluation method for big data quality evaluation is proposed. Through the analysis of big data quality characteristics, a big data quality evaluation system for the whole process of data processing is constructed. The subjective weight and objective weight of each indicator are calculated through the analytic hierarchy process and entropy method. In order to overcome the subjective and one-sided shortcomings of the single weight determination method, the subjective weight and the objective weight are organically integrated through the distance function method to determine the combined weight of each indicator. The quantified result of big data quality is obtained through fuzzy calculation of membership degree. Finally the ranking results of the proposed method are compared with those of some existing multi-attribute decision-making (MADM) methods. The obtained results indicate that the proposed method is reasonable and efficient to deal with MADM problems. It can comprehensively measure the level of big data quality, and provide users with accurate and efficient quality evaluation results.

1. INTRODUCTION

Since the birth of big data, people have gradually realized the social and economic value of big data and paid great attention to it [1–3]. Many developed countries have successively issued relevant policies to promote the development of big data, and have promoted big data as a national strategy [4–6]. Through big data analysis, it is possible to summarize experience, discover laws, predict trends, assist decision-making, release and utilize the huge value contained in data resources, provide convenience for people's lives, provide value for business activities, and provide strategic opportunities for national development [7–10]. However, the value of big data is based on accurate, comprehensive, and high-quality data. The quality of the data directly determines the validity and reliability of the big data analysis results. Wang *et al.* [11] more clearly stated that the quality of big data is the prerequisite for big data research and the basis of all data analysis, mining, and decision support. Therefore, evaluating the quality of big data is a necessary prerequisite for big data applications.

Big data quality evaluation research has attracted widespread attention. Some scholars have proposed different evaluation models from the perspectives of big data characteristics, quality measurement, application scenarios, and life cycle. For example, Desai [12] used Monte Carlo and neural network methods to calculate composite data quality, and built a big data quality evaluation model based on eight core measurement parameters. Ardagna *et al.* [13] constructed big data quality evaluation and expectation models for different scenarios based on the relationship between confidence, execution time, and budget. Batini *et al.* [14] presented a big data quality evaluation model with data types, data sources, and application domains as the core structural features. Mo [15] developed a quality evaluation model consisting of original quality, process quality, and result quality, and proposed a data quality evaluation standard based on the concept of grassroots stratification. In addition, some scholars have constructed different evaluation index systems of big data quality based on the evaluation dimensions, quality characteristics, data types, and other indicators of big data quality. For example, Cai *et al.* [16] summarized the basis for dividing data quality indicators based on the characteristics of tobacco, military, medical, meteorological communications, and other industries, and established a big data quality evaluation system for recognized indicators such as accuracy, applicability, and timeliness. Aggarwal [17] established the index system of big data quality evaluation from three dimensions of quantity scale, changed speed, and variety type, and proposed the evaluation basis of each index, respectively. Kulkarni [18] established a big data quality evaluation system from three aspects: content-based measurement, context-based measurement, and rating-based measurement, and gave the calculation method of the weight of each indicator. Zhao *et al.* [19] divided big data quality into objective and subjective standards, constructed a unified data model and a data quality standard model for multi-source internet platforms, and gave a standard measurement method for big data quality.

[*] *Corresponding author. Email: 78192128@qq.com*

Although scholars have conducted a lot of research on the evaluation of big data quality, with the deepening understanding of the connotation of big data quality and the increasing complexity of big data structure [20,21], data quality evaluation under big data still faces some challenges: (1) lack of a general big data quality evaluation model and (2) the single weight determination method is generally used to determine the weight of the evaluation index, which is relatively subjective and one sided.

The essence of big data quality evaluation is a multi-attribute decision-making (MADM) problem. MADM is the decision-making problem of choosing the best alternative or sorting the alternatives, with the consideration of multiple attributes. In this paper, we introduce a new method that can be used for comprehensive fuzzy evaluation method (CFEM) in a MADM problem for big data quality. It carries out in-depth analysis of the original attributes, usage attributes, and result attributes of big data quality, and constructs a big data quality evaluation system based on content quality, use quality, and result quality. CFEM uses the analytic hierarchy process (AHP) and entropy method to calculate the subjective weight and objective weight of each indicator, and uses the distance function method to organically integrate the weights required by the two methods to determine the comprehensive weight of each indicator. Then the final quantitative evaluation result is calculated by the determined index weight and the fuzzy operation of membership degree. An example under real data environment is used in this study for validation of the proposed method. Based on this example, the effectiveness and practicability of the method are verified, and ranking results are analyzed. Moreover, a comparison is made between the ranking results of the proposed method with the results of some latest methods, such as CoCoSo, C-TOPSIS, A-TOPSIS, EDAS, and SECA.

The rest of the paper is listed as follows: In Section 2, big data quality fuzzy evaluation model is presented. In Section 3, an illustrate example is provided. In Section 4, a comparison and analysis is given. Some conclusions are derived in Section 5.

2. BIG DATA QUALITY FUZZY EVALUATION MODEL

The purpose of big data quality evaluation is to evaluate the true situation of data through scientific and objective methods, so that data users can clearly and definitely understand the current state of the data [22], so as to make judgments and decisions on the data, and take appropriate measures to improve the quality of data. The evaluation process of big data quality can be divided into three stages. The first is to build a big data quality evaluation index system; the second is to determine the combined weight of each index; the last is a fuzzy comprehensive evaluation.

2.1. Construction of Big Data Quality Evaluation Index System

2.1.1. The content of big data quality evaluation

The big data processing process mainly includes data acquisition, data processing, data storage, data analysis, data visualization, data application, and other links [23]. Data quality runs through the entire big data processing process. Therefore, the factors that determine the quality of big data are the content attributes, usage attributes, and result attributes of the data source.

(i) Content attributes: Content attributes refer to the attributes of the big data source itself, and represent the original quality of the big data. It includes data completeness, timeliness, accuracy, and reliability. Completeness refers to the lack of data, the lack of data may be the lack of structure, the lack of content, or the lack of information in a certain field in the data; timeliness refers to the time interval between data generation and utilization, which reflects the degree of the newness of the data, the more timely the data is used, the stronger the timeliness is; accuracy refers to the data accurately and faithfully reflect the conditions of the original world; reliability refers to the stability and safety of data in a certain period. The evaluation of content attributes is the key content of big data quality evaluation.

(ii) Usage attributes: Usage attributes refer to the external attributes of big data sources, reflect the availability of data, and represent the quality of big data usage. It includes data availability, compatibility, and operability. Availability refers to the availability, integration, and comprehensibility of data; compatibility refers to the coordination and use of data in different modes, reflecting the degree of convertibility and adaptability between data; operability refers to the degree of data access, acquisition, merging, access, and other operations. The evaluation of using attributes is the main content of big data quality evaluation.

(iii) Result attributes: The result attribute refers to the application attribute of big data, which reflects the final application result of data, and represents the result quality of big data. It includes the applicability, consistency, and validity of the data. Applicability refers to the ability of data to meet the intended use requirements, reflecting the degree of relevance to the target data; consistency refers to whether the semantics, format, and structure of data are consistent, and it reflects the degree of similarity between data contexts; validity refers to the application effect of the data, which reflects the value and application degree of the data. The evaluation of result attributes is an important content of big data quality evaluation.

2.1.2. Construction of the index system

The selection of evaluation indexes is directly related to the objectivity and impartiality of the evaluation results. The author referred to many research results at home and abroad, and on the basis of following the principles of science, system and maneuverability, conducted in-depth analysis of the original attributes, usage attributes, and result attributes of big data, and a big data quality evaluation system based on content quality, use quality, and result quality was constructed, as shown in Table 1.

2.1.3. Index evaluation

The big data quality evaluation index system includes quantitative indicators and qualitative indicators. Qualitative indicators are

Table 1 Evaluation index system of big data quality.

Goal	First Indicators	Secondary Indicators	Evaluation Basis	Index Characteristics
Evaluation system of big data quality (U)	Content quality (A1)	Completeness (B1)	Evaluation of data completeness and missing rate	Quantitative
		Timeliness (B2)	Evaluation of degree of data new or old and timeliness of use	Quantitative
		Accuracy (B3)	Evaluation of data error rate and abnormal rate	Quantitative
		Reliability (B4)	Evaluation of data of stability, standardization and security	Qualitative
	Use quality (A2)	Availability (C1)	Evaluation of data of availability, integration, and comprehensibility	Qualitative
		Compatibility (C2)	Evaluation of data of convertibility, adaptability, and verifiability	Qualitative
		Operability (C3)	Evaluation of data of accessibility and availability	Qualitative
	Result quality (A3)	Applicability (D1)	Evaluation of data of relevance, credibility, and coherence	Qualitative
		Consistency (D2)	Evaluation of the consistency of semantics, format, and structure	Quantitative
		Effectiveness (D3)	Evaluation of data of usability, value added, and traceability	Qualitative

fuzzy evaluated by the experience of domain experts; quantitative indicators are calculated and determined by the characteristics of big data. Assuming the data source $\mathrm{S} = \left\{\mathrm{D_1}, \mathrm{D_2}, \cdots, \mathrm{D_n}\right\}$, it means that the data source has n object instances, $\mathrm{D_i}$ is the ith object instance. Each object has a total of $\mathrm{m_i}$ features, which can be expressed as $\mathrm{C_i} = \left\{\langle \mathrm{K_{i1}}, \mathrm{V_{i1}}\rangle, \langle \mathrm{K_{i2}}, \mathrm{V_{i2}}\rangle, \cdots, \left\langle \mathrm{K_{im_i}}, \mathrm{V_{im_i}}\right\rangle\right\}$, where $\mathrm{K_i}$ represents the attribute name of the ith feature, $\mathrm{V_i}$ represents the corresponding feature value. Different data sources and object data have different feature numbers $\mathrm{m_i}$. According to the research progress in recent years, the quantitative indicators are determined according to the following methods:

(i) Completeness. It refers to the completeness of the characteristic attributes of each object entity in the data source and the nonempty value of the existing attributes, and the complete data space mapped according to the fusion of multiple data sources [24]. The complete feature space of the entity described by the object $\mathrm{D_i}$ is $\hat{\mathrm{R}}_\mathrm{i} = \left\{\mathrm{E_1}, \mathrm{E_2}, \cdots, \mathrm{E_{m_i}}\right\}$, then the completeness of the data source can be expressed by $\mathrm{Q_1}$:

$$\mathrm{Q_1} = \frac{\sum_{j=1}^{n} t_{1j}}{n} \tag{1}$$

Among them, $t_{1j} = \frac{\sum_{i=1}^{m_i} F_1\left(V_{ij}\right)}{m_i}$, F_1 is a function to judge whether the characteristic value is nonempty. If it is not empty, it is 1; if it is empty, it is 0.

(ii) Timeliness. It refers to the new and old degree of object instance data in the data source. If there are more obsolete and invalid data in the data source, the timeliness will be worse, otherwise, the timeliness will be better. The formula for measuring the obsolescence of data sources is $t_{2j} = \frac{\sum_{i=1}^{m_i} F_2\left(V_{ij}\right)}{m_i}$, where $F_2\left(V_{ij}\right)$ is to judge whether the j record under the ith data source is old or new.

$$F_2\left(V_{ij}\right) = \begin{cases} \dfrac{T_{ij} - T_{publish}}{T_{expire} - T_{publish}}, T_{expire} < T_{ij} \leq T_{publish} \\ 1 + \ln\left(1 + \displaystyle\int_{T_{expire}}^{T_{ij}} \left(T_{expire}/T_{occur} w T_{ij} e^{T_{ij}}\right) d_{T_{ij}}\right), \\ T_{ij} > T_{expire} \end{cases} \tag{2}$$

Among them, $\mathrm{T_{occur}}, \mathrm{T_{publish}}, \mathrm{T_{expire}}$ are the 3-dimensional time vector about the data source and its description information [25]. Further, the timeliness $\mathrm{Q_2}$ of the data source can be expressed as

$$\mathrm{Q_2} = \frac{\sum_{j=1}^{n} \left(1 - t_{2j}\right)}{n} \tag{3}$$

(iii) Accuracy. It refers to the correctness and accuracy of the description of the object attribute value in the data source. The accuracy of the data source is represented by $\mathrm{Q_3}$.

$$\mathrm{Q_3} = \frac{\sum_{j=1}^{n} t_{3j}}{n} \tag{4}$$

Among them, $t_{3j} = \frac{\sum_{i=1}^{m_i} F_3\left(V_{ij}\right)}{m_i}$, F_3 is a function for judging whether the feature value is correct or not. If it is correct, it is 1, otherwise it is 0; $t_{3\mathrm{j}}$ represents the correctness of the jth data.

(iv) Consistency. It refers to the consistency of the semantics, format, and structure of data between different entities with the same reference in the data source, including the consistency between the same source and the different source. The consistency of the data source can be represented by $\mathrm{Q_4}$.

$$Q_4 = \frac{\sum_{j=1}^{n} t_{4j}}{n} \tag{5}$$

Among them, $t_{4j} = \frac{\sqrt{\sum_{j=1}^{m_i} \sum_{k=1,k\neq j}^{m_i} Re\left(k_j, k_k\right) \bullet Dis\left(V_{ij}, V_{ik}\right)}}{m_i}$, *Re* is the correlation between two features; *Dis* is the logical distance between the two values; t_{4j} represents the consistency of the *j*th data.

2.2. The Weighting Model of Indicator Combination

The most critical issue of big data quality evaluation is to determine the weight of each indicator. Due to the diversity and extensiveness of big data sources, there are different methods for determining the weights of different quality standards. Determined based on the subjective sensitivity and importance of each indicator by experts, called subjective weight. According to the actual characteristics of each indicator and determined by mathematical calculations, it is called objective weight. Comprehensive subjective and objective confirmation weights are called combined weights.

2.2.1. Determination of the subjective weight of indicators based on AHP

AHP is one of the most versatile decision-making techniques which reflect the natural behavior and human thought. This technique is the study of complex problems based on their mutual effects and converts them to a simple form and solves them. AHP can be used when faced with competing multiple choice or multicriteria decision-making problems [26]. The decision maker begins providing a hierarchical tree. The decision hierarchical tree shows the comparable factors and evaluates the competing alternatives in a decision, and then a series of paired comparisons are carried out. The comparisons show the weight of each factor toward competing alternatives to evaluate in a decision. Finally, the hierarchical analysis process logic integrates matrices of pairwise comparisons together in a way to make a better decision [27].

AHP is a combination of qualitative and quantitative system analysis method, which is applied to deal with a condition of uncertainty and subjective judgment [28,29]. The steps include constructing judgment matrix, hierarchical ordering, and consistency testing.

(i) Construct a judgment matrix. The judgment matrix indicates that it is directed against the indicators of the previous level, comparison of relative importance between indicators at this level [31]. In order to reduce the difficulties caused by factors of different nature, relative scales are used when comparing pairwise. The 1–9 scale method is usually used to construct the judgment matrix [30].

Assuming that the index *B* of a certain layer in the big data quality evaluation system has *n* secondary indicators $B = \left\{B_1, B_2, \cdots, B_n\right\}$, the relative importance judgment matrix of each indicator is

$$B = \left(b_{ij}\right)_{n\times n} = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \ddots & b_{2n} \\ \vdots & \vdots & & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nn} \end{bmatrix} \tag{6}$$

Among them, $b_{ij} * b_{ji} = 1$.

(ii) Calculate the eigenvector and the maximum eigenvalue. The sum-product method is used to solve the eigenvector approximate solution and the largest eigenvalue of the judgment matrix B. It mainly includes the following steps:

Step 1: Normalize the judgment matrix B by column to obtain the normalized matrix $A = \left(a_{ij}\right)_{nxn}$, The formula is given by Equation (7).

$$a_{ij} = \frac{b_{ij}}{\sum_{k=1}^{n} b_{kj}} \tag{7}$$

Step 2: Add the matrix *A* row by row to get the corresponding feature vector $R = \left(R_i\right)_{n\times 1}$. The formula is given by Equation (8).

$$R_i = \sum_{j=1}^{n} a_{ij} \tag{8}$$

Step 3: Normalize the feature vector *R* to obtain the weight vector $\text{ш} = \left(\text{ш}_i\right)_{n\times 1}$ of the hierarchical single sorting. The formula is given by Equation (9).

$$\text{ш}_i = \frac{R_i}{\sum_{i=1}^{n} R_i} \tag{9}$$

Step 4: Calculate the maximum eigenvalue $\lambda_{\max}$ according to the following formula. The formula is given by Equation (10).

$$\lambda_{max} = \sum_{i=1}^{n} \frac{(B\text{ш})_i}{(n\text{ш})_i}. \tag{10}$$

(iii) Consistency inspection. Due to the complexity and uncertainty of actual big data evaluation problems, especially when there are many evaluation factors involved, it is easy to cause judgment errors. Therefore, it is necessary to check the consistency of the judgment matrix to reduce subjective deviation and ensure that the judgment result is consistent with the actual situation. The test formula is given by Equation (11).

$$\mathrm{CR} = \frac{\lambda_{\max} - \mathrm{n}}{\mathrm{RI} * (\mathrm{n} - 1)}. \tag{11}$$

In Equation (11), RI is the average random consistency index, which is related to *n* [32]. When CR < 0.1, *B* is passed the consistency test and considered to be able to objectively reflect the relative importance of the indicators, therefore, it can be adopted. Otherwise, *B* needs to be adjusted until it passes. The final ш_i is the weight value corresponding to the *i*th indicator.

(iv) Calculate subjective weight value. Multiply the weight value of the first-level index and the weight value of the second-level index to obtain the weight of each index relative to the target. Therefore, the subjective weight is $\overline{w} = \left(\overline{w_1}, \overline{w_2}, \cdots, \overline{w_n}\right)$.

2.2.2. Determination of objective weight based on entropy method

Objective weight is the weight obtained through data operation based on the decision matrix formed by the characteristic information of the data source on each index. In order to reduce the influence of subjective factors and fully consider the original data information, this paper uses the entropy method to determine the objective weight of the indicator. The entropy method is a method to determine the index weight based on the information characteristics of the sample data itself [33,34]. It draws on the idea of information entropy, by calculating the information entropy of the index, and determining the weight of the index according to the influence of the discrete degree of the index on the whole. The higher entropy means the smaller the degree of dispersion, the smaller the utility value and the smaller the weight; the smaller the entropy means the greater the degree of dispersion, the greater the utility value, and the greater the weight [35]. According to the characteristics of the entropy method, assuming that the number of evaluation indicators is n and the sample size is m, the initial decision matrix can be obtained as follows:

$$X = \left(x_{ij}\right)_{m\times n} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \ddots & x_{2n} \\ \vdots & \vdots & & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix}. \tag{12}$$

Among them, x_{ij} represents the value of the jth evaluation index in the ith sample. In the evaluation process, quantitative indicators are directly or indirectly obtained according to the characteristics of the data; qualitative indicators are scored based on the experience of experts, the indicators are divided into multiple evaluation levels, and the measurement scale value is determined according to the level. The calculation steps of the entropy method are as follows:

Step 1: Calculate the proportion p_{ij} of each sample index value under each index:

$$p_{ij} = \frac{x_{ij}}{\sum_{i=1}^{m} x_{ij}}. \tag{13}$$

Step 2: Calculate the information entropy value e_{j} of each indicator:

$$e_j = -K \sum_{i=1}^{m} p_{ij} ln\left(p_{ij}\right). \tag{14}$$

Among them, K is a constant, $K = 1/\ln(n)$; e_{j} is the information entropy value of the jth index, and the value range is $0 \le e_j \le 1$.

Step 3: Calculate the difference coefficient d_{j} of each index, the difference coefficient indicates the effect of the index on the research object, and it directly affects the weight of the evaluation index:

$$d_j = 1 - e_j. \tag{15}$$

Step 4: Calculate the weight $\underline{w} = \left(\underline{w_j}\right)_{1\times n}$ of each indicator relative to the comprehensive evaluation:

$$\underline{w_j} = \frac{d_j}{\sum_{j=1}^{n} d_j}. \tag{16}$$

According to the nature of the entropy value: $0 \le \underline{w_j} \le 1$, $\sum_{j=1}^{n} \underline{w_j} = 1$, the objective weight of the evaluation index can be finally obtained.

2.2.3. Determination of combination weight based on distance method

In order to ensure the reliability and credibility of the weights taken by each evaluation index and overcome the shortcomings of a single method for weight determination, this paper uses the distance function method to organically combine the two weights for comprehensive weight determination. The distance function method uses the concept of distance function to align the degree of difference between the subjective and objective weights with the degree of difference in the corresponding distribution coefficient [36–38], taking into account the subjective experience of the evaluator on the actual situation, and has important statistical significance. The calculation formula is:

$$d\left(\overline{w_i}, \underline{w_i}\right) = \left[\frac{1}{2} \sum_{i=1}^{n} \left(\overline{w_i} - \underline{w_i}\right)^2\right]^{1/2} \tag{17}$$

$$w_i = \alpha \overline{w_i} + \beta \underline{w_i} \tag{18}$$

Among them, $d\left(\overline{w_i}, \underline{w_i}\right)$ is the distance function; $\overline{w_i}$ is the subjective weight; $\underline{w_i}$ is the objective weight; $\mathrm{w_i}$ is the combined weight value; α, β are the distribution coefficients, satisfying $\alpha + \beta = 1, 0 < \alpha, \beta < 1$.

In order to make the degree of difference between the subjective and objective weights consistent with the degree of the distribution coefficient, the distance function and the distribution coefficient are equalized, and the expression is given as follows:

$$d\left(\overline{w_i}, \underline{w_i}\right)^2 = (\alpha - \beta)^2. \tag{19}$$

$$\alpha + \beta = 1. \tag{20}$$

Combine formulas (19) and (20) to obtain the distribution coefficient of the combined weight:

$$\mathrm{\alpha} = \left[\frac{\sum_{i=1}^{n} \left(\overline{w_i} - \underline{w_i}\right)^2}{8}\right]^{1/2} + \frac{1}{2}. \tag{21}$$

Substituting α and β into formula (18), the combined weight vector $w = \left(w_1, w_2, \cdots, w_n\right)$ of each evaluation index can be obtained.

3. COMPREHENSIVE EVALUATION OF BIG DATA QUALITY

3.1. Source of Experimental Data

In order to verify the validity and rationality of the method in this paper for data quality evaluation under big data, experiments are carried out under real data environment. The test data mainly comes from the collected objective data and the subjective data of experts.

3.1.1. Objective data

In order to fully reflect the effect of the algorithm, objective data comes from real data sets. Collect commodity information on 6 mainstream integrated B2C e-commerce platforms in China through custom crawlers. The data source platforms are defined as S1, S2, S3, S4, S5, and S6, which correspond to Tmall, JD.com, Suning.com, Gome Online, Yihaodian, Vipshop. The collected commodity data include 3,562,570 commodities in 10 categories. The amount of commodity data on different platforms is shown in Table 2.

3.1.2. Subjective data

Twenty industry experts and scholars were invited to evaluate the two parts of the data by anonymous questionnaire. First, the relative importance of evaluation indicators was scored to obtain the relative importance of each indicator. Second, according to the characteristics of the collected data, the qualitative indicators in the evaluation index system were fuzzy evaluated to obtain the fuzzy membership data of each indicator [39,40].

3.2. Evaluation Index Weight Calculation

3.2.1. Subjective weight calculation

According to the actual impact of each index on the quality of big data, a judgment matrix is constructed according to the AHP, and an expert group familiar with the quality of big data will compare and score the importance of each evaluation index. For experts with large differences in opinions, multiple rounds of scoring are used until the judgment data is basically consistent and the consensus test is passed. The details are shown in Tables 3–6.

According to the Table 3, the maximum eigenvalue ($\lambda_{\max}$)is 4.118 and CR is 0.044369. The judgment matrix (A1-B) passed the consistency test. Among the influences of evaluation content quality on overall data quality, the importance and influence of completeness are the highest (0.558), followed by accuracy (0.263), reliability (0.122), and timeliness (0.057).

According to the Table 4, the maximum eigenvalue ($\lambda_{\max}$) is 3.054202 and CR is 0.046726. The judgment matrix (A2-C) passed the consistency test. Usability is the most important and influential factor (0.703), followed by compatibility (0.182) and operability (0.115) in the evaluation of the impact of usage quality on the overall data quality.

According to the Table 5, the maximum eigenvalue ($\lambda_{\max}$) is 3.03871 and CR is 0.033375. The judgment matrix (A2-D) passed the consistency test. Applicability has the highest importance and influence (0.633), followed by validity (0.261) and consistency (0.106) among the influences of evaluation result quality on overall data quality.

According to the Table 6, the maximum eigenvalue ($\lambda_{\max}$) is 3.03871 and CR is 0.033375. The judgment matrix (U-A) passed the consistency test. In the evaluation process of big data quality, the importance and influence of content quality are the largest (0.633), followed by use quality (0.261) and result quality (0.106).

3.2.2. Objective weight calculation

According to the entropy method proposed above, the objective weights of the indicators are calculated, and the quantitative indicators are directly calculated based on the collected data; the qualitative indicators are scored based on expert experience, and according to the scoring results, the membership degree of each index is calculated. The degree of membership refers to the degree of affiliation of each evaluation index corresponding to the comment set and each level [41,42]. The indicators are divided into ten levels. Different levels correspond to different scores. The tenth level corresponds to the highest score of 10, and the first level corresponds to the lowest score of 1. Using r_{ij} to represent the probability that the ith index is rated as the jth level. The value of r_{ij} is determined by the calculation of $r_{ij} = n/m$, where n is the number of

Table 2 Distribution of data sets on various platforms.

Category	S1	S2	S3	S4	S5	S6
Mobile digital	96540	85246	82498	78512	67548	57524
Office appliances	74328	72016	69521	65842	57426	76521
Costume jewelry	85212	65482	62547	75345	41625	65234
Beauty care	65239	67265	65482	56538	54682	70258
Food and drink	72456	70215	76254	70892	65478	56842
Books audiovisual	49521	42156	65241	46521	42683	56147
Toy musical instrument	39852	42103	45218	21354	24587	36587
Daily necessities	84526	65482	75423	65324	61248	74523
Outdoor sport	69524	53248	62157	36518	45294	45216
Home textiles	39564	35465	42653	45214	32972	45681
Total	676762	598678	646994	562060	493543	584533

Table 3 The judgment matrix of content quality (A1-B).

A1-B	B1	B2	B3	B4	Sort Weight
B1	1	7	3	5	0.558
B2	1/7	1	1/5	1/3	0.057
B3	1/3	5	1	3	0.263
B4	1/5	3	1/3	1	0.122

Table 4 The judgment matrix of use quality (A2-C).

A2-C	C1	C2	C3	Sort Weight
C1	1	5	5	0.703
C2	1/5	1	2	0.182
C3	1/5	1/2	1	0.115

Table 5 The judgment matrix of result quality (A3-D).

A3-D	D1	D2	D3	Sort Weight
D1	1	5	3	0.633
D2	1/5	1	1/3	0.106
D3	1/3	3	1	0.261

Table 6 The judgment matrix of target layer (U-A).

U-A	A1	A2	A3	Sort Weight
A1	1	3	5	0.633
A2	1/3	1	3	0.261
A3	1/5	1/3	1	0.106

people rated as the jth level, and m is the total number of experts. The entropy method is used to calculate the objective weight of the big data quality standard, and the results are shown in Table 7.

The above evaluation results show that the entropy value of the integrity evaluation index is the smallest, the degree of dispersion is the largest, the utility value is the largest, and the weight is the largest; and the consistency evaluation index has the largest entropy value, the degree of dispersion is the largest, the utility value is the smallest, and the weight is the smallest.

3.2.3. Combination weight calculation

Substituting the abovementioned subjective and objective weight values into the distance function calculation formulas (20) and (21), the distribution coefficient can be obtained:

$$\alpha = \left[\frac{\sum_{i=1}^{n} \left(\overline{w_i} - \underline{w_i} \right)^2}{8} \right]^{1/2} + \frac{1}{2} = 0.531,$$

$$\beta = 1 - \alpha = 0.469.$$

Substituting α and β into Equation (18), the combined weight value of each evaluation index can be obtained, as shown in Table 8.

From the combined weights in Table 8, it can be seen that completeness (0.350), availability (0.156), and accuracy (0.152) are the three largest and most important indicators. This result is consistent with the reality. First of all, data integrity plays a decisive role. It is the most basic guarantee of data and determines whether data has value to a certain extent. Second, the availability of data is a key indicator for evaluating the quality of big data. Unavailable data will lead to serious errors in knowledge and decision-making derived from the data, therefore, the quality of data with low availability is also poor. Finally, accuracy reflects the authenticity and error-freeness of the original big data. Data that is not accurate enough cannot truthfully reflect the actual situation. If the data with poor accuracy is analyzed and mined, the results obtained will be quite different from the actual situation, which will seriously affect the decision-making effect. In addition, operability (0.032), consistency (0.034), and timeliness (0.038) are the three indicators with the smallest weights, indicating that they are also the smallest in importance and influence on the quality of big data.

Finally, using the combined weights in Table 8, the final scores and ranking order of the 6 data sources can be obtained, as shown in Table 9.

4. COMPARISON AND ANALYSIS

4.1. Comparison with Existing Ranking Methods

To verify the acceptability of the CFEM method, our combined weights are used in CoCoSo [43], C-TOPSIS [44], A-TOPSIS [45], EDAS [46], and SECA [47] to solve the aforementioned problem. Table 10 illustrates the obtained rankings of different methods.

The Kendall's coefficient of concordance for the above results is equal to 0.91111, indicating a great amount of concordance among different ranking. Considering CFEM, its correlation with other methods seems acceptable. Afterward, the Spearman's correlation among different methods with this aggregated one is presented in Table 11.

In view of the results of Table 11, the least correlation of the CFEM method points 0.828571 (with EDAS), while this method exhibits a correlation more than 94% with other methods. So it implies that the results of new CFEM method have a great concordance with previous methods. All in all, it can be concluded that the CFEM method results can be considered acceptable.

4.2. Comparison with Fuzzy Complementary Judgment Matrix Method

To verify the reasonable and effective of the CFEM method in determine the weight of each evaluation index, we compared it with the fuzzy complementary judgment matrix method. The fuzzy complementary judgment matrix method is a method to determine the index weight and ranking according to the fuzzy complementary judgment matrix [48]. It first constructs the fuzzy judgment matrix, applies mathematical transformation to the matrix, and finally uses the sorting formula to calculate the weight of each indicator. First, the nine scales in this paper are transformed into corresponding scales, which are (absolutely strong, extremely strong, strong, slightly strong, same, slightly weak, weak, extremely weak,

Table 7 Objective weight of evaluation index.

Data No.	Completeness	Timeliness	Accuracy	Reliability	Availability	Compatibility	Operability	Applicability	Consistency	Effectiveness
s1	0.78	0.72	0.83	0.8	0.75	0.75	0.9	0.85	0.73	0.65
s2	0.62	0.89	0.94	0.7	0.9	0.75	0.75	0.75	0.65	0.65
s3	0.43	0.75	0.74	0.75	0.65	0.9	0.9	0.75	0.73	0.55
s4	0.54	0.84	0.66	0.85	0.65	0.85	0.85	0.75	0.64	0.6
s5	0.76	0.82	0.68	0.65	0.85	0.75	0.8	0.6	0.73	0.55
s6	0.75	0.82	0.86	0.85	0.8	0.85	0.85	0.8	0.58	0.65
Weights	0.346	0.041	0.134	0.079	0.125	0.046	0.034	0.089	0.061	0.045

Table 8 Combined weight of evaluation index.

Goal	First Indicators	Secondary Indicators	Subjective Weight	Objective Weight	Combined Weight
Evaluation system of big data quality (U)	Content quality (A1)	Completeness (B1)	0.353	0.346	0.350
		Timeliness (B2)	0.036	0.041	0.038
		Accuracy (B3)	0.167	0.134	0.152
		Reliability (B4)	0.077	0.079	0.078
	Use quality (A2)	Availability (C1)	0.183	0.125	0.156
		Compatibility (C2)	0.048	0.046	0.047
		Operability (C3)	0.03	0.034	0.032
	Result quality (A3)	Applicability (D1)	0.067	0.089	0.077
		Consistency (D2)	0.011	0.061	0.034
		Effectiveness (D3)	0.028	0.045	0.036

Table 9 Evaluation result of data source.

Data No.	Score	Rank
s1	0.77824	2
s2	0.75800	3
s3	0.62485	6
s4	0.65720	5
s5	0.73784	4
s6	0.78735	1

absolutely weak), and the corresponding scale values are (0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, 0.1). Second, the judgment matrix given by the experts in this paper is transformed into the corresponding fuzzy complementary judgment matrix. Finally, using fuzzy complementary judgment matrix sorting algorithm to find the weight formula:$w_i = \frac{\sum_{j=1}^{n} a_{ij} + \frac{n}{2} - 1}{n(n-1)}, i = 1, 2, \cdots, n$. The first-level index, the second-level index, and the final weight are calculated, respectively. The comparison between the first-level index and the final weight is shown in Table 12.

It can be seen from Table 12 that the weight obtained by the fuzzy complementary judgment matrix sorting algorithm has small discrimination and poor comparability. The final weight calculated by the fuzzy complementary judgment matrix sorting algorithm (the weight of the first-level index is multiplied by the weight of the second-level index), the maximum is 0.133, the minimum is 0.077, the range between indicators is 0.057, the interquartile range is 0.023, and the standard deviation is 0.0158, indicating that the degree of dispersion between the weight values is small, and the relative importance of the indicators cannot be better reflected. The weight between indicators lack comparability, and it is difficult to find the true indicators that determine the quality of big data. However, the weight value calculated by CFEM method in this paper, the maximum is 0.350, the minimum is 0.034, the range is 0.316, the interquartile range is 0.118, and the standard deviation is 0.0944, which indicates that there is a large degree of dispersion between weight values and strong comparability between indicators. It reflects that the weight value of the indicators with great impact on the quality of big data is obviously larger, while the index weight with small impact on the quality of big data is obviously smaller. There are obvious differences between weights, which can better reflect the relative importance of indicators and facilitate the objective evaluation of data quality.

The degree of data missing, old and new, accuracy, and reliability in the big data environment directly affect the quality of data sources and the value of data. Therefore, content quality is an important factor that affects the quality of big data. Compared with use quality and result quality, it belongs to a strongly important level, and the weight of content quality should be significantly greater than other factors. By using the fuzzy complementary judgment matrix sorting algorithm, the weight of content quality is 0.383, the weight of use quality is 0.333, and the weight of result quality is 0.283, which does not reflect the strong importance of content quality. However, using the method in this paper, the weight of content quality is 0.633, the weight of use quality is 0.261, and the weight of result quality is 0.106. The weight of content quality is much higher than other factors, which shows that content quality is obviously more important than use quality and result quality, This phenomenon is very consistent with the reality. This also shows that the method proposed

Table 10 | Ranking of alternatives with different methods.

Data No.	CFEM (Proposed)	CoCoSo [43]	C-TOPSIS [44]	A-TOPSIS [45]	EDAS [46]	SECA [47]
s1	2	2	2	2	1	2
s2	3	3	4	4	4	3
s3	6	6	6	6	5	6
s4	5	4	5	5	6	5
s5	4	5	3	3	3	4
s6	1	1	1	1	2	1

Table 11 | Spearman's correlation of different methods.

Methods	CoCoSo [43]	C-TOPSIS [44]	A-TOPSIS [45]	EDAS [46]	SECA [47]
Spearman's correlation	0.942857	0.942857	0.942857	0.828571	1.0000

Table 12 | The table of weight comparison.

Goal	First Indicators	Relative Weights (CFEM)	Relative Weights (FCJM)	Secondary Indicators	Final Weights (CFEM)	Final Weights (FCJM)
Evaluation system of big data quality (U)	Content quality (A1)	0.633	0.383	Completeness (B1)	0.350	0.115
				Timeliness (B2)	0.038	0.077
				Accuracy (B3)	0.152	0.102
				Reliability (B4)	0.078	0.089
	Use quality (A2)	0.261	0.333	Availability (C1)	0.156	0.133
				Compatibility (C2)	0.047	0.103
				Operability (C3)	0.032	0.097
	Result quality (A3)	0.106	0.283	Applicability (D1)	0.077	0.109
				Consistency (D2)	0.034	0.080
				Effectiveness (D3)	0.036	0.094

in this paper is suitable for evaluating the quality of data sources in a big data environment

4.3. Total Discussing

(i) The method of this paper uses the AHP to determine the weight of each impact indicator, decomposes multi-level and complex goals into several levels of multiple indicators, and quantifies the importance of qualitative indicators with fuzzy concepts, reducing the influence of subjective factors, solves the problem that the evaluation index is difficult to accurately define, and ensures the objectivity of the evaluation results. At the same time, the fuzzy mathematics theory is used to comprehensively and quantitatively evaluate the vague and uncertain information, reducing the subjective arbitrariness of decision-makers, effectively improve the reliability and accuracy of judgment and evaluation.

(ii) When evaluating the quality of data sources in a big data environment, single indicator cannot meet the evaluation requirements. Using the CFEM method in this paper, the quality of data sources is regarded as a whole and according to the thinking mode of hierarchical decomposition, comparative judgment and system synthesis, fuzzy evaluation objects are processed by precise digital means. Not only can the evaluation object be evaluated and sorted according to the comprehensive score, but also the grade of the object can be evaluated in the light of the maximum membership principle according to the value on the fuzzy evaluation set. It overcomes the defect of the single result of traditional mathematical methods and provides a certain basis for objectively evaluating the quality of big data.

(iii) Subjective weights and objective weights have their own advantages and disadvantages. The subjective weight reflects the preferences and requirements of the evaluator, but it is easily affected by the subjective factors of the evaluator, and subjective one sidedness cannot be avoided; the objective weight reflects the logical relationship and objective laws of the data itself, but it largely depends on the quality of the sample, and does not reflect the user's point of view. Therefore, this paper uses the distance function method to organically combine the two weights for comprehensive weight determination, which overcomes the shortcomings of a single method for weight determination, and makes the result of the weighting as close as possible to the actual result.

5. CONCLUSION

Big data quality is the prerequisite for big data research and the foundation for all data analysis, mining, and decision support. Evaluating the quality of big data is helpful to understand the essential characteristics of the data and to avoid decision errors caused by data errors to the greatest extent. Aiming at the actual problems of data quality evaluation in the big data environment, this paper proposes a CFEM for big data quality. Through in-depth analysis of the original attributes, usage attributes, and result attributes of big data quality, a big data quality evaluation system based on content quality, usage quality, and result quality is constructed. Calculating the subjective weight and objective weight of each indicator through the AHP and entropy method, and using the distance function method to organically integrate the weights obtained by the two methods to determine the comprehensive weight of each indicator, then through the determined indicator weight, calculating with the degree of membership to get the final quantitative evaluation result. Collecting data from six major Chinese e-commerce platforms through web crawlers and using this method to evaluate these data. The evaluation results show that this method can well adapt to the big data environment, and the weight of quality evaluation of the obtained big data meets the requirements of quality evaluation. Finally, the results are compared with other well-known algorithms. The performance of CFEM method related to those methods indicated acceptable results. However, the data sources in the big data environment are extensive and complex, so how to evaluate the data quality more efficiently and accurately in the complex application environment will be the focus of the next research.

CONFLICT OF INTERESTS

The authors declare that they have no conflict of interest.

AUTHORS' CONTRIBUTIONS

methodology, Wenquan Li; software: Wenquan Li; validation, Wenquan Li; formal analysis, Wenquan Li; investigation, Suping Xu; resources, Suping Xu; data curation, Suping Xu; original draft preparation, Wenquan Li; review and editing, Xindong Peng; visualization, Suping Xu; supervision, Xindong Peng; project administration, Wenquan Li; funding acquisition, Xindong Peng.

Funding Statement

This research was funded by the National Natural Science Foundation of China, grant number 62006155 and research project of the Philosophy and Social Science Planning Project of Guangdong Province grant numbers GD18CFX06 and GD18XFX11.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (no. 62006155) and the Philosophy and Social Science Planning Project of Guangdong Province (nos. GD18CFX06 and GD18XFX11).

REFERENCES

[1] T.Z. Emara, J.Z. Huang, Distributed data strategies to support large-scale data analysis across geo-distributed data centers, IEEE Access. 8 (2020), 26–38.

[2] R. Expósito, R. Galego-Torreiro, J. González-Domínguez, SeQual: big data tool to perform quality control and data preprocessing of large NGS datasets, IEEE Access. 8 (2020), 146075–146084.

[3] M. Li, Z. Liu, X. Shi, ATCS: auto-tuning configurations of big data frameworks based on generative adversarial nets, IEEE Access. 8 (2020), 50485–50496.

[4] I.E. Alaoui, Y. Gahi, Network security strategies in big data context, Procedia Comput. 175 (2020), 730–736.

[5] M. Dai, L. Liu, Risk assessment of agricultural supermarket supply chain in big data environment, Sustain. Comput. Inform. Syst. 28 (2020), 1–9.

[6] X. Kong, L. Chao, S. Yu, Power supply reliability evaluation based on big data analysis for distribution networks considering uncertain factors, Sustain. Cities Soc. 63 (2020), 1–15.

[7] K. Kaur, S. Garg, G. Kaddoum, A big data-enabled consolidated framework for energy efficient software defined data centers in IoT setups, IEEE Trans. Ind. Inform. 16 (2020), 2687–2697.

[8] Y. Zhang, X. Mou, D.M. Chandler, Learning no-reference quality assessment of multiply and singly distorted images with big data, IEEE Trans. Image Process. 29 (2020), 2676–2691.

[9] Q. Liu, G. Feng, X. Zhao, Minimizing the data quality problem of information systems: a process-based method, Decis. Support Syst. 137 (2020), 1–12.

[10] Y. Liu, Y. Wang, K. Zhou, *et al.*, Semantic-aware data quality assessment for image big data, Future Gener. Comp. Syst. 102 (2020), 53–65.

[11] Y. Wang, W. Huang, Z. Zhu, Some thoughts on big data industry and management issues, Sci. Tech. Dev. 1 (2014), 15–194.

[12] K. Desai, Big Data Quality Modeling and Validation, Ph. D. dissertation, San José State University, San Jose, CA, USA, 2018.

[13] D. Ardagna, C. Cappiello, W. Samá, *et al.*, Contextaware data quality assessment for big data, Future Gener. Comp. Syst. 89 (2018), 548–562.

[14] C. Batini, A. Rula, M. Scannapieco, *et al.*, From data quality to big data quality, J. Database Manage. 26 (2015), 60–82.

[15] Z. Mo, Construction of big data quality measurement model, Inform. Stud. Theory Appl. 41 (2018), 11–15.

[16] L. Cai, Y. Liang, Y. Zu, *et al.*, Historical evolution and development trend of data quality, Comp. Sci. 45 (2018), 1–16.

[17] A. Aggarwal, Data quality evaluation framework to assess the dimensions of 3VS of big data, Int. J. Emerg. Tech. Adv. Eng. 7 (2017), 503–506.

[18] A. Kulkarni, A study on metadata management and quality evaluation in big data management, Eng. Tech. Appl. Sci. Res. 4 (2016), 455–459.

[19] X. Zhao, S. Li, W. Yu, *et al.*, Research on web data source quality assessment method in big data, Comp. Eng. 43 (2017), 48–56.

[20] T. Nagle, T. Redman, D. Sammon, Assessing data quality: a managerial call to action, Bus. Hor. 63 (2020), 325–337.

[21] B. Liu, L. Pang, Review of domestic and international research on big data quality, J. China Soc. Sci. Tech. Inform. 38 (2019), 217–226.

[22] S. Salloum, J. Huang, Y. He, Random sample partition: a distributed data model for big data analysis, IEEE Trans. Ind. Inform. 15 (2019), 5846–5854.

[23] M. Mahmud, J. Huang, S. Salloum, *et al.*, A survey of data partitioning and sampling methods to support big data analysis, Big Data Mining Anal. 3 (2020), 85–101.

[24] Y. Hu, S. li, W. Yu, *et al.*, Recognizing the same commodity entities in big data, J. Comp. Res. Dev. 52 (2015), 1794–1805.

[25] L. Gan, Y. Liu, S. Yang, *et al.*, Web temporal inconsistency modeling based on web time axis, J. Comput. Inform. Syst. 8 (2012), 1063–1070.

[26] A. Rezaei, S. Tahsili, Urban vulnerability assessment using AHP, Adv. Civ. Eng. 2018 (2018), 1–20.

[27] X. Zhang, S. Huang, S. Yang, *et al.*, Safety assessment in road construction work system based on group AHP-PCA, Math. Probl. Eng. 4 (2020), 1–12.

[28] H. Zabihi, M. Alizadeh, I.D. Wolf, *et al.*, A GIS-based fuzzy-analytic hierarchy process (F-AHP) for ecotourism suitability decision making: a case study of Babol in Iran, Tour. Manage. Pers. 36 (2020), 1–17.

[29] G. Büyüközkan, C.A. Havle, O. Feyzioğlu, A new digital service quality model and its strategic analysis in aviation industry using interval-valued intuitionistic fuzzy AHP, J. Air Trans. Manage. 86 (2020), 1–16.

[30] T.S. Abdelhamid, J.G. Everett, Identifying root causes of construction accidents, J. Constr. Eng. Manage. 126 (2000), 52–60.

[31] R. Jena, B. Pradhan, Integrated ANN-cross-validation and AHP-TOPSIS model to improve earthquake risk assessment, Int. J. Disaster Risk Red. 50 (2020), 1–35.

[32] S. Wold, K. Esbensen, P. Geladi, Principal component analysis, Chemometr. Intell. Lab. Syst. 2 (1987), 37–52.

[33] L. Chen, S. Gao, B. Liu, *et al.*, FEW-NNN: a fuzzy entropy weighted natural nearest neighbor method for flow-based network traffic attack detection, China Commun. 17 (2020), 151–167.

[34] D. Zhao, C. Li, Q. Wang, *et al.*, Comprehensive evaluation of national electric power development based on cloud model and entropy method and TOPSIS: a case study in 11 countries, J. Clean. Prod. 277 (2020), 1–14.

[35] W. Gong, N. Wang, N. Zhang, *et al.*, Water resistance and a comprehensive evaluation model of magnesium oxychloride cement concrete based on Taguchi and entropy weight method, Const. Build. Mater. 260 (2020), 1–8.

[36] J.B. Liu, M.A. Malik, N. Ayub, *et al.*, Distance measures for multiple-attributes decision-making based on connection numbers of set pair analysis with dual hesitant fuzzy sets, IEEE Access. 8 (2020), 9172–9184.

[37] N. Lam, G. Lu, L. Zhang, Geometric Hardy's inequalities with general distance functions, J. Func. Anal. 279 (2020), 1–35.

[38] A. Rodríguez-Hoyos, J. Estrada-Jiménez, D. Rebollo-Monedero, *et al.*, The fast maximum distance to average vector (F-MDAV): an algorithm for k-anonymous microaggregation in big data, Eng. Appli. Artif. Intell. 90 (2020), 1–12.

[39] L. Kong, B. Ma, Evaluation of environmental impact of construction waste disposal based on fuzzy set analysis, Environ. Tech. Innova. 19 (2020), 1–13.

[40] G.J. Meyer, T. Lorz, R. Wehner, *et al.*, Hybrid fuzzy evaluation algorithm for power system protection security assessment, Elec. Power Syst. Res. 189 (2020), 1–7.

[41] I.M. Jiskani, Q. Cai, W. Zhou, *et al.*, Assessment of risks impeding sustainable mining in Pakistan using fuzzy synthetic evaluation, Res. Policy. 69 (2020), 1–13.

[42] D. Bell, M. Lycett, A. Marshan, *et al.*, Exploring future challenges for big data in the humanitarian domain, J. Bus. Res. 131 (2021), 453–468.

[43] M. Yazdani, P. Zarate, E.K. Zavadskas, *et al.*, A Combined Compromise Solution (CoCoSo) method for multi-criteria decision-making problems, Manage. Decis. 57 (2018), 2501–2519.

[44] C.L. Hwang, K. Yoon, Multiple Attribute Decision Making: Methods and Applications: a State-of-the-Art Survey, Springer Verlag, New York, NY, USA, 1981.

[45] H. Deng, C.H. Yeh, R.J. Willis, Inter-company comparison using modified TOPSIS with objective weights, Comput. Oper. Res. 27 (2000), 963–973.

[46] M.K. Ghorabaee, E.K. Zavadskas, L. Olfat, *et al.*, Multi-criteria inventory classification using a new method of Evaluation Based on Distance from Average Solution (EDAS), Informatica. 26 (2015), 435–451.

[47] K.G. Mehdi, A. Maghsoud, K.Z. Edmundas, *et al.*, Simultaneous Evaluation of Criteria and Alternatives (SECA) for multi-criteria decision-making, Informatica. 29 (2018), 265–280.

[48] H. Li, J. Deng, J. Fang, Housing quality defects influence factors analysis based on fuzzy analytic hierarchy process, Jiangxi Sci. 34 (2016), 85–90.