Research Article

Software Quality Evaluation Model Based on Weighted Mutation Rate Correction Incompletion G1 Combination Weights

Chuanyang Ruan and Jianhui Yang

School of Business Administration, South China University of Technology, Guangzhou 510640, China

Correspondence should be addressed to Chuanyang Ruan; ruancyang@163.com

Aiming at the common problems of quality evaluation method, this paper first establishes a fuzzy software quality evaluation model according to the relationship of software quality subcharacteristics and indicators; furthermore, considering the uncertainty and individual deviations of expert judgment results, this paper corrects and tests the consistency of the incomplete information sorting given by the experts and obtains an integration sorting of gathering different expert opinions through the idea of circling modification; at last, this paper proposes the weighted mutation rate which is used to measure the development balance degree and determines weights of evaluation indicators via weighted mutation rate correction incompletion G1 method, which avoids the problem of integration of subjective and objective weights.

1. Introduction

With the increasing popularity of software products, the market requires high software product quality continuously. How to evaluate the quality of software has become the most concerned problem to users and managers of software organizations, because the evaluation results of software quality can not only guide the users to purchase and use the software, but also guide software developers to develop high-quality software products [1–3].

Previous studies have shown that, due to the characteristics of the software itself and the limitation of public cognitive level, software quality evaluations are always vague and uncertain. Just because of the fuzzy characteristics of software quality, domestic and overseas scholars began to use fuzzy comprehensive evaluation method to solve some core issues of software quality evaluation [4–8]. In the fuzzy comprehensive evaluation method, membership function construction is a quite difficult and complex process, and the construct of the weight is not reasonable. In the construction of membership function, in order to meet the continuous and gradient characteristics of quality, semitrapezoid distribution and trapezoidal distribution were currently used [6–8]. As for the weighting methods, current studies focused on the following aspects.

(a) Subjective weighting method: the analytic hierarchy process (AHP) and the order relation analysis method (G1) [9] are the representative of the weighting method, the characteristic of which is the index weight information which came from the subjective experience judgment of experts.

(b) Objective weighting method: it refers to the weighting method taking entropy method [10], establishing mathematical programming models [11–13] as the representative methods, which are characterized by the index weight information which came from data information of Index, but are affected by subjective factors. It is also the inadequacies of objective weighting method.

(c) Combination of objective and subjective weighting methods: the main representatives are addition synthesis [9] and multiplication synthesis [11–13], but we do not know that whether addition and multiplication methods are reasonable. Literature [14] proposed the combination weights based on standard deviation correction G1 combination weight model, which guaranteed that the combination weights could reflect the expert advices and data information, and only required each expert to give the importance

sorting of all indicators without considering the possible incomplete information phenomenon of expert sorting.

Most of the above combination weighting methods require evaluation experts to consider the same set of evaluation index and give all the evaluation information to each expert. However, in some realistic evaluation of software quality, due to the influence of the knowledge and the experience of each expert, the understanding and cognition of evaluation criteria, psychological scale, and environmental factors, there exist uncertainty, individual judgment bias, and incomplete information in the evaluation result of each expert. In order to eliminate the deviation of individual evaluation results and uncertain information and reduce the impact of randomness and contingency of individual evaluation results, group evaluation should be used. After that, correct uncertain information to increase the effective number of expert evaluation results and conduct the consistency testing [15] or the accuracy analysis [16] of all experts evaluation results, discard the expert evaluation results that do not pass the inspection, finally integrate into the ideal evaluation information. According to the perspective of statistical significance, the more experts, the less dispersion degree of individual evaluation results, the closer group evaluation summary results to the true value of the evaluation object. However, the number of experts is often limited by some factors like evaluation period and the costs in the evaluation practices; sometimes even peer experts themselves in the strict sense are always limited, which is difficult to satisfy the need of expert number. Practice shows that when the number of experts is close to 15, the influence of further increasing number of experts to evaluation accuracy has already turned out to be tiny [16, 17]. Therefore, the number of experts in the evaluation is usually 7 to 15.

According to ISO/IEC 9126 standard, this paper confirms the evaluation index used to measure software quality and then establish a complete set of index systems used as the foundation of software quality evaluation. Design an effective software quality evaluation method, conduct fuzzy processing of index evaluation criterion by the method of constructing membership function in fuzzy mathematics, then obtain, and quantify the fuzzy quality of indicators. For fully considering the different preferences of different experts, this paper corrects the incomplete information meeting certain conditions given by some experts and finally obtains results that passed the consistency test through integrating with the order relation of complete evaluation given by experts, and the integrated order relation had the highest level of consistency. Then propose a weighted mutation rate formula measuring the degree of development difference, by which correct incomplete information G1, and obtain combination weights based on weighted mutation rate correction incompletion G1. The key to this method is how to construct the membership function of software index, how to deal with different preferences of different experts, and how to integrate weights information with subjective and objective factors properly.

2. Software Evaluation Index System

2.1. Software Quality Index System Decomposition. ISO/IEC 9126 standard [18] is a current popular international standard on software metrics and software quality evaluation. ISO/IEC 9126 indicates software quality with six quality characteristics, which are the functionality, reliability, usability, maintainability, efficiency, and portability. Each quality characteristic is indicated with quality subcharacteristic, for example, subcharacteristics of functionality are suitability, accuracy, interoperability, security, and compliance. Each subcharacteristic also contains some relevant measure indexes. Characteristics, subcharacteristics and measurable indexes constitute the three-stage index model of software quality.

The definition, method, and standard of measure in ISO/IEC 9126 standard are general concepts. In order to make the evaluation more visualized and operable, decompose measure in the standard into indicators and data items, formed evaluation index system of software quality, as shown in Figure 1. Each index contains n data items, the value of which is collected from the process software testing, substituting the value of data items into index formula and getting the indicator value. For example, suitability of subcharacteristic contains four indicators: functional attainment ratio, functional specification change ratio, precise input-output definition ratio, and project documentation ratio. Table 1 is the actual example of functional attainment ratio index [7, 19].

2.2. Measurement and Quantification Based on the Index System. One of the key issues of software quality evaluation is to determine the quality evaluation indicators and the common method is as follows: firstly, determine the evaluation standard of indicators and divide the quality of indicators into four grades (Poor, Average, Good, and Excellent); secondly, determine the index threshold through expertise method ($v1$, $v2$, $v3$, $v4$, and $v5$); that is to say, define the value range of indicator on each level as ($v1$, $v2$, $v3$, $v4$, and $v5$); Finally, obtain the quality of index through comparing the value with threshold of indicators [7]. In the above index quality also exist the following shortcomings in the evaluation: (a) disagreement with the continuous and gradient characteristics of quality; (b) losing the evaluation information; (c) poor quantification degree.

The above problems can be solved by fuzzing the evaluation standards of index through constructing the membership function of index quality on all grades. The paper uses semitrapezoid distribution and trapezoidal distribution [20, 21] to construct the membership function of quality grade. Set the domain of discourse V = {Poor, Average, Good, Excellent}, and A is a fuzzy subset of indicator on the domain of discourse. Constructed membership functions, respectively, are

$$A_1(v) = \begin{cases} 1, & v_1 \leq v < v_2, \\ \dfrac{v - c_2}{v_2 - c_2}, & v_2 < v < c_2, \\ 0, & \text{else,} \end{cases}$$
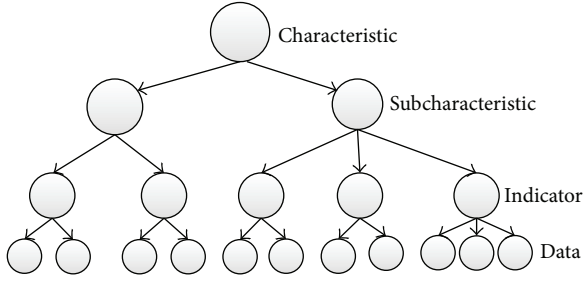
Figure 1: Evaluation index system of software quality.

Table 1: Calculation method of functional attainment ratio index.

Indicator	functional attainment ratio, FA
Application of indicator	The ratio of effective function completion and described function in requirements
Index calculation method	FA = Fe/Fs
Data item	Fe: number of effective function completion Fs: number of described function in requirements

$$
A_2(v) = \begin{cases} 1, & v_2 \le v < v_3, \\ \dfrac{v - c_1}{v_2 - c_1}, & c_1 < v < v_2, \\ \dfrac{v - c_3}{v_3 - c_3}, & v_3 < v < c_3, \\ 0, & \text{else,} \end{cases}
$$

$$
A_3(v) = \begin{cases} 1, & v_3 \le v < v_4, \\ \dfrac{v - c_2}{v_3 - c_2}, & c_2 < v < v_3, \\ \dfrac{v - c_4}{v_4 - c_4}, & v_4 < v < c_4, \\ 0, & \text{else,} \end{cases}
$$

$$
A_4(v) = \begin{cases} 1, & v_4 \le v \le v_5, \\ \dfrac{v - c_3}{v_4 - c_3}, & c_3 < v < v_4, \\ 0, & \text{else,} \end{cases}
$$

(1)

of which c_1, c_2, c_3, and c_4 are central values of intervals (v_1, v_2), (v_2, v_3), (v_3, v_4), and (v_4, v_5).

For example, with regard to the functional attainment ratio in Table 1, assume its threshold value obtained by expertise method [8] as $(0, 0.8, 0.90, 0.95, 1)$, then $v_1 = 0$, $v_2 = 0.8$, $v_3 = 0.9$, $v_4 = 0.95$, and $v_5 = 1$, which is substituted into (1) and obtain the membership function of indicator on each quality grade. If we substitute the measured indicator value $v = 0.92$ into the membership degree equation, then obtain that the membership degrees of the indicator are $A_1(v) = 0$,

$A_2(v) = 0.2$, $A_3(v) = 1$, and $A_4(v) = 0$. So the fuzzy quality of the indicator is

$$
A(v) = \left\{ \frac{0}{\text{Poor}}, \frac{0.2}{\text{Average}}, \frac{1}{\text{Good}}, \frac{0}{\text{Excellent}} \right\}.
$$

(2)

The normalized quality $a(v) = \{0, 0.167, 0.833, 0\}$, calculating the membership degree on different quality grades is a good solution of unclear boundary among the quality grades.

In order to facilitate the comparison of similar software products, the degree of excellent, good, average, and poor can be further quantized. Set $(4, 3, 2, 1)^T$ is the corresponding score set of the remark set {excellent, good, average, poor}, that is, the score of excellent, good, average, and poor are 4, 3, 2, and 1, respectively. Set Z_{ij} to represent the quantized quality of the jth indicator of ith evaluated object and D_{ij} to represent the fuzzy quality of the jth indicator of ith evaluated object; then the quantitative score [8] of this index of specific is

$$
Z_{ij} = D_{ij} \times (4, 3, 2, 1)^T.
$$

(3)

3. Correction Incompletion Group G1 Combination Weights Model

3.1. Difficulty and Solution

(1) Problems of Expert Group Decision Making.

First Problem. As the evaluation experts of different fields often come from different organizations and sectors, and each expert has different knowledge and experience, disagreement always emerges when experts compare the degree of importance of two indicators. For one expert, it is easier to determine which one is more important from two important indicators. But there are often conflict when different experts compare the same two indicators. In a comprehensive evaluation with a number of experts and multiple indicators, pair-wise comparison of different indicators from different experts is involved, which is apparently more trouble and more inconsistent.

Second Problem. It is more prone to have different opinions when multiple experts directly determine the value ratio of how many times one indicator is more important than the other indicator. Obviously, in a comprehensive evaluation with numerous experts and multiple indicators, it is harder to directly determine the value ratio of how many times one indicator is more important than the other indicator, rather than to determine which indicator is important.

(2) Solutions. First Thought. Each expert would be required to sort the importance of indicator. Considering the indicator omission selected by experts according to the actual situation, we integrate expert's incomplete sorting which meets the integration condition and construct a complete information indicator set. Then we unify the sorting of importance degree of two indicators in different expert indicator sets by consistency testing and cycling modification, obtaining a unified sorting that gathers from different experts' opinions.

Second Thought. On the basis of first thought, we could determine the ratio of importance degree of two indicators through comparing their weighted mutation rate (see Section 3.2.3). It was easier to sort two indicators, while it is more difficult to determine the value of importance degree, which tends to cause conflicts, so that we choose to avoid the latter to avoid the contradiction.

3.2. Establishment of G1 Combination Weights Model of Weighted Mutation Rate Correction Incompletion

3.2.1. Introduction of Traditional Order Relation Analysis Method (G1). The order relation analysis method (G1) is a typical subjective weighting method. In this method, the weight information of index is all from the subjective experience of expert and each expert is asked to evaluate all indicators, without considering the expert's preference or reflecting the data information of indicators. Specific steps are as follows.

(1) Weight of Index Layer to Criterion Layer

(1) Experts determine the order relation of indicators.

(2) Experts give rational assignment of importance degree ratio r_k of adjacent indicators x_{k-1} and x_k. Rational assignment is shown in Table 2.

(3) According to r_k value given by experts, the G1 weight of mth indicator on criterion layer is as follows:

$$t_m = \left(1 + \sum_{k=2}^{m} \prod_{i=k}^{m} r_i \right)^{-1}. \tag{4}$$

(4) Weights of the $m-1, m-2, \ldots, 2$th indicators can be obtained from weight t_m:

$$t_{m-1} = r_m t_m, \tag{5}$$

of which t_m represents the weight of mth indicator to the layer on the criterion layer.

(2) Weight of Index Layer to General Objective Layer. Set β_k as the weight of kth indicator to general objective layer on jth criterion layer; x_k is the weight of kth indicator to jth criterion layer on jth criterion layer; $x^{(j)}$ is the weight of jth criterion layer to general objective layer. Then the weight β_k of the indicator to general objective layer is

$$\beta_k = \chi_k \times \chi^{(j)}. \tag{6}$$

3.2.2. Weighted Mutation Rate Correction Incompletion G1 Combination Weights. In the current study, we generally use linear weighting as the integration method of subjective and objective weights. Based on weighted mutation rate, this paper proposed a nonlinear weighting method. We construct the correction incompletion G1 combination weights model based on weighted mutation rate to determine the combination weight from a new perspective. We not only

TABLE 2: Assignment reference table.

r_k	Explain
1.0	indicators x_{k-1} and x_k is equally important
1.2	indicators x_{k-1} and x_k is somewhat important
1.4	indicators x_{k-1} and x_k is obviously important
1.6	indicators x_{k-1} and x_k is highly important
1.8	indicators x_{k-1} and x_k is extremely important

consider experts' preferences and reflect experts' opinions, but also include objective data information. Detailed steps are as follows.

(1) Each expert evaluated the order of importance of indicators in evaluation set and then corrected the incomplete information order relation from some experts, obtaining a corrected complete sorting. Then we implemented consistency testing and information integration with the order relation given by other experts, in order to get an ideal sorting, which can take experts' preferences into account and reflect experts' opinions; see Sections 3.2.4 and 3.2.5.

(2) Determine the importance degree ratio r_k of adjacent indicators x_{k-1} and x_k by calculating the weighted index mutation rate v_k (see Section 3.2.3), which can reflect the data information of indicators and illustrate the objectivity of this method. Consider

$$r_k = \begin{cases} \dfrac{v_{k-1}}{v_k}, & v_{k-1} \geq v_k, \\ 1, & v_{k-1} < v_k. \end{cases} \tag{7}$$

Other steps were the same as traditional G1 method; see Section 3.2.1.

3.2.3. Definition of Weighted Mutation Rate and Its Metric Property. To measure the degree of difference among a set of data, this paper proposed a new measurement method, weighted mutation rate. Weighted mutation rate is a method where we obtain data information variation based on objective data. The degree of data difference can be directly judged from the weighted mutation rate.

Define the weighted mutation rate v_j as

$$v_j = \frac{s_j}{(m-1)\,\overline{y}_j} \times 100\%, \tag{8}$$

of which, y_{ij} was the jth indicator value of ith evaluated objected, $\overline{y}_j = (1/m) \sum_{i=1}^{m} y_{ij}$ was the jth mean of indicator, $w_{ij} = y_{ij} / \sum_{i=1}^{m} y_{ij}$ was the percentage of indicator data, and $s_j = \sqrt{\sum_{i=1}^{m} (y_{ij} - \overline{y_j})^2 \times w_{ij}}$ indicated the weighted standard error of jth indicator.

It is easy to prove that weighted mutation percentage rate v_j has good metric quality. Simplify formula (8):

$$v_j = \frac{1}{m-1} \sqrt{\frac{\sum_{i=1}^{m} m^2 \times y_{ij} \times \left(\left(y_{ij} \right)^2 - 2 y_{ij} \times \overline{y} + \overline{y}^2 \right)}{\left(\sum_{i=1}^{m} y_{ij} \right)^3}}. \tag{9}$$

Then simplify formula (9):

$$v_j = \frac{1}{m-1} \sqrt{\frac{m^2 \sum_{i=1}^{m} y_{ij}^3}{\left(\sum_{i=1}^{m} y_{ij}\right)^3} - \frac{2m \sum_{i=1}^{m} y_{ij}^2}{\left(\sum_{i=1}^{m} y_{ij}\right)^2} + 1}. \qquad (10)$$

As $0 < \sum_{i=1}^{m} y_{ij}^3 / (\sum_{i=1}^{m} y_{ij})^3 \leq \sum_{i=1}^{m} y_{ij}^2 / (\sum_{i=1}^{m} y_{ij})^2 \leq 1$, if $b = \sum_{i=1}^{m} y_{ij}^2 / (\sum_{i=1}^{m} y_{ij})^2$, so formula (10) could be simplified as

$$v_j \leq \frac{1}{m-1} \sqrt{b(m^2 - 2m) + 1} \leq 1. \qquad (11)$$

From formula (11), we knew that $v_j \leq 1$. And formula (11) was greater than or equal to 0. That means that $0 \leq v_j \leq 1$. Therefore, the greater weighted mutation rate v_j is, the higher the degree of development difference among the various indicators is. It is easy to prove that $v_j = 0$ is the necessary and sufficient conditions for absolutely equal development among various indicators. That is $y_{1j} = y_{2j} = \cdots = y_{nj} = \overline{y}$. When $y_{1j} = y_{2j} = \cdots = y_{(n-1)j} = 0$, $y_{nj} \neq 0$, $\overline{y}_{ij} = y_{nj}/n$, $v_j = 1$; Otherwise, if $v_j = 1$, it's easy to prove that $y_{1j} = y_{2j} = \cdots = y_{(n-1)j} = 0$, $y_{nj} \neq 0$, $\overline{y}_{ij} = y_{nj}/n$. That means that $v_j = 1$ was the necessary and sufficient conditions that had absolute differences among the development of various indicators. That is, most indicator items had no progress; only one indicator developed. It was clear that weighted mutation rate had the similar value and meaning as Gini coefficient [22, 23].

3.2.4. Correction Method of Expert Incomplete Information Order Relation. In order to reduce the degree of dispersion of individual evaluation and improve the objectivity of the assessment, selecting experts who possess intelligence and virtue was the key to perfect peer review. Moreover, in order to fully and objectively assess the value of the object, the structural composition of the experts in group evaluation should be representative. That means that the experts should come from different regions and different work units and graduate from different schools, and sometimes they even need to be representative in the industry, subject area, academic schools, and other aspects, so as to offset the individual deviation from each other while calculating the summarized results in group evaluation.

Choose m experts ($7 \leq m \leq 15$) to rank n indicators, assuming t experts give the complete sorting, while there were varying degrees of incomplete indicators in the other sorting of $(m - t)$ experts. Specific correction steps were as follows.

Step 1. Expert classification and modification condition to classify the experts giving the identical index number as one category. If every indicator in this expert sorting was chosen by different expert, that is, each indicator in the category was chosen by at least one expert, then we could conduct the information modification and consistency testing on the incomplete information of expert. Otherwise we would reject the category of expert sorting.

Step 2. Employ the sorting and scoring method [15] to convert the incomplete information given by each expert into scores:

$$R_{ij} = n - r_{ij} + 1, \qquad (12)$$

r_{ij}, $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, m$ was the location of ith indicator in jth sorting of expert. If there is parallel sorting, there was a need to normalize the results of expert, conducting "skip" processing to the following evaluated objects of the parallel sorting. Then we adjusted the same number to the mean of corresponding sequence number. For example, the sorting result of an expert was $1, 2, 2, 3, 4, 5, \ldots$, then the normalized sequence number would be $1, 2.5, 2.5, 4, 5, 6, \ldots$.

Step 3. Calculate the mean value of indicators in different methods:

$$\overline{R}_i = \frac{1}{m} \sum_{j=1}^{m} R_{ij}. \qquad (13)$$

Use the mean of indicators as the score of missing indicators. Then insert them into the sorting of experts according to the numerical size. We would obtain a complete index order relation of each expert. If there was the same mean of two indicators, we should calculate the weighted mutation rate of the indicator score under different expert order relations, in which the small one was superior.

Step 4. For corrected complete indicator order relation of different experts, the results may vary, but the difference should not be too large for the same indicators. Thence, we need to conduct the consistency testing based on Spearman rank correlation coefficient on sorting, which would be discarded if it failed the consistency testing. Set the sorting from jth method as $A_j = (a_{1j}, a_{2j}, \ldots, a_{nj})$; then the formula of Spearman rank correlation coefficient [15] of jth and kth sorting was

$$V_{jk} = 1 - \frac{6 \sum_{i=1}^{n} \left(a_{ij} - a_{ik}\right)^2}{n(n^2 - 1)}. \qquad (14)$$

Discard criteria are as follows. Calculate the Spearman rank correlation coefficient of one expert sorting and other expert sorting, when the mean value was greater than or equal to 0.7, it met the reliability condition and passed consistency testing. Namely, the expert sorting converted to effective sorting, which was equivalent to a complete index order relation. Otherwise, discard the expert sorting.

3.2.5. Establishment of Ideal Sorting. Conduct the consistency testing by formula (14) on the above expert correction sorting that passed the consistency testing and t complete sortings of expert order relation. And then score the expert sorting that passed the consistency testing by formula (12), if β_{ij} was the score of the ith indicator of the jth expert order relation, $\overline{a}_i$ was the synthesis mean score of the ith indicator; then

$$\overline{a}_i = \frac{\sum_{j=1}^{n} \beta_{ij}}{n}. \qquad (15)$$

We would get the integration order relation by reordering the order relation by size. From Theorem 1, integration order relation not only satisfied the condition of Spearman consistency testing, but also achieved the optimal degree of consistency. Therefore, integration order relation was the ideal sorting.

Theorem 1. *If different order relation* $A_1, A_2, \ldots, A_n$ *aiming at the same indicators satisfied the consistency condition of Spearman rank correlation coefficient, the integrated order relation A from the above also satisfied the consistency condition of Spearman rank correlation coefficient and even had the optimal degree of consistency.*

The following was the proof process of Theorem 1. Start with reviewing a lemma.

Lemma 2. *If* $x_i > 0$, $i = 1, 2, \ldots, n$, $\overline{x_i^2}, \overline{x}_i^2$ *were the mean of square and the square of mean respectively, then*

$$\overline{x_i^2} \geq \overline{x}_i^2. \tag{16}$$

If and only if $x_1 = x_2 = \cdots = x_n$, *both were equal.*

Proof. a_{ij} represented the location of the ith indicator by the jth expert, $\overline{V}$ was the mean of integration order relation and the Spearman rank correlation coefficient ranked by experts, $\overline{V}_j$ represented the mean of the jth expert order relation and Spearman rank correlation coefficient ranked by other experts.

By (14), the mean of Spearman rank correlation coefficient can be deduced:

$$\overline{V}_j = 1 - \frac{6 \sum_{i=1}^{n} \sum_{k=1, k \neq j}^{m} \left(a_{ij} - a_{ik}\right)^2}{(m-1) n \left(n^2 - 1\right)}. \tag{17}$$

Then the mean of the integration order relation as follows:

$$\overline{V} = 1 - \frac{6 \sum_{i=1}^{n} \sum_{k=1}^{m} \left(\overline{a}_i - a_{ik}\right)^2}{mn \left(n^2 - 1\right)}. \tag{18}$$

Contrasted the consistency degree of the integration order relation and expert sorting:

$$\overline{V} - \overline{V}_j$$

$$= \frac{6 \sum_{i=1}^{n} \sum_{k=1, k \neq j}^{m} \left(a_{ij} - a_{ik}\right)^2}{(m-1) n \left(n^2 - 1\right)} - \frac{6 \sum_{i=1}^{n} \sum_{k=1}^{m} \left(\overline{a}_i - a_{ik}\right)^2}{mn \left(n^2 - 1\right)}$$

$$= \frac{6 \sum_{i=1}^{n} \left[m^2 \left(a_{ij} - \overline{a}_i\right)^2 + \left(a_{i1}^2 + a_{i2}^2 + \cdots + a_{im}^2\right) - m\overline{a}_i^2\right]}{mn (m-1) \left(n^2 - 1\right)}. \tag{19}$$

Based on Lemma 2, $\left(a_{i1}^2 + a_{i2}^2 + \cdots + a_{im}^2\right) - m\overline{a}_i^2 = m\overline{a_i^2} - m\overline{a}_i^2 \geq 0$, and other items were greater than or equal to 0; then $\overline{V} \geq \overline{V}_j$; namely, the mean of Spearman rank correlation coefficient of the integration order relation is optimal. □

Some scholars [24–26] got the comprehensive sorting by weighting the expert consistency degree, where the greater the weight is, the greater the degree of consistency is. The integration sorting obtained by this weighted method was certain to satisfy the consistency condition, which could be proved by the above method, but not necessarily the best sorting, and the consistency degree of the method proposed by this paper was higher than the weighted method.

3.3. Advantage of Weighted Mutation Rate Correction Incompletion G1 Method. In conclusion, the weighted mutation rate correction incompletion G1 combination weights have the following advantages comparing with other weighting methods.

(1) In the evaluation process, we can fully consider the preferences of the experts and modify the possible incomplete information of experts. Then we construct a correction method based on incomplete information. Finally, we integrate into the ideal sorting with complete information sorting of other experts. This method makes full use of the knowledge and experience of experts, reflecting uncertainty and incompletion of information in the evaluation process.

(2) By comparing the weighted mutation rate of indicators, we are able to determine the ratio of importance degree of adjacent indicators. With this, we combine the experience of expert and data information of indicators reasonably, which not only makes the combination weights reflect the expert information (indicator sorting) and the data information (determination of the ratio of indicator importance degree by the weighted mutation rate), but also avoids the optional chose of determining the important ratio of indicators and the thorny issue of how to distribute the objective and subjective weights.

(3) Wide applicability of weighting method is another advantage. The combination weighting method has extensive applicability. The only requirement is that there must be quantified indicators data. Calculating the weighted mutation rate directly during the calculation, the combination weights can be obtained by integrating ideal sorting according to the importance sorting given by evaluator.

4. Evaluation Equation

The comprehensive evaluation score could be obtained by weighted summation of quantified scores of indicators in Section 2.2 and combination weights in Section 3.2.2. Set d_i: as the evaluated score of the ith evaluated object, t_{ik} as the incompletion G1 method combination weights of the kth indicator of the ith evaluated object, and x_{ik} as the quantified score of the kth indicator of the ith evaluated object; then the evaluation equation was as shown in

$$d_i = \sum_{k=1}^{n} x_{ik} t_{ik}. \tag{20}$$

5. Application of Comprehensive Evaluation of Software Quality

5.1. Principle of Evaluation of Software Quality. As shown in Figure 1, each subcharacteristic of software quality includes a number of quality indicators, namely, the quality of subcharacteristic is decided by the quality of indicators it contains. So the quality of subcharacteristic can be obtained from the quality of comprehensive indicators. Therefore, based on the quality of indicators and the corresponding weights, the quality of subcharacteristic and characteristic could be evaluated through fuzzy comprehensive evaluation method.

On the basis of constructing the index system, this paper researched fuzzy comprehensive evaluation method based on weighted mutation rate correction incompletion G1 combination weights. First, the value of data items in the index system was from the actual software development and testing process, which was the basis for quantitative evaluation. After that, we substituted the data item into the formula of membership degrees and calculated fuzzy quality and quantified score of indicators. The index quality was not only drawn on expert experience, but also based on the actual measurement data which was more credible. Secondly, as to possible different situations of experts and obtained objective quantitative data, this paper put forward a weighting method integrating subjective and objective evaluation based on weighted mutation rate correction incompletion G1 combination weights. Finally, according to the relationship between the subcharacteristic of software quality and indicators, we employed fuzzy comprehensive evaluation method to evaluate the quality of subcharacteristic, in which way we could obtain the characteristic quality and overall quality of software. Its comprehensive evaluation schematic diagram was shown in Figure 2.

5.2. Process of Evaluation of Software Quality. The following took suitability, the subcharacteristic of functionality as the example to introduce how to evaluate the subcharacteristic by fuzzy comprehensive evaluation method. The subcharacteristic suitability contains four indicators [8]: functional attainment ratio, functional specification change ratio, precise input-output definition ratio, and project documentation ratio, signified with U_1, U_2, U_3, U_4 respectively. Set the factor set $U = \{U_1, U_2, U_3, U_4\}$ and judgment set $V = \{$Poor, Average, Good, Excellent$\}$. Assume that Z_{ij} represented the quantified quality of the jth indicator of the ith evaluated object and D_{ij} represented the fuzzy of the jth indicator of the ith evaluated object; there are three evaluated software, Z_1, Z_2, Z_3.

5.2.1. Fuzzy Quantified Score of Software Quality. For each indicator of the factors set U, evaluated by the method in Section 2.2, tested and substituted into the membership degree formula (1), we obtained the fuzzy quality of each indicator of the factor set U separately, $D_{11} = \{0, 0.167, 0.833, 0\}$, $D_{12} = \{0, 0.375, 0.625, 0\}$, $D_{13} = \{0.286, 0.714, 0, 0\}$, $D_{14} = \{0, 0, 0.333, 0.667\}$, $D_{21} = \{0.7, 0, 0, 0.3\}$, $D_{22} = \{0.15, 0.25, 0.6, 0\}$, $D_{23} = \{0, 0.55, 0, 0.45\}$, $D_{24} = \{0.525, 0, 0.475, 0\}$, $D_{31} = \{0.15, 0.85, 0, 0\}$, $D_{32} = \{0, 0, 0.8, 0.2\}$, $D_{33} = \{0.3, 0.15, 0.2, 0.35\}$, $D_{34} = \{0.2, 0, 0.5, 0.3\}$, then substituted into formula (3), obtained the matrix of quantized score:

$$Z = \begin{bmatrix} 2.167 & 2.375 & 3.286 & 1.333 \\ 3.100 & 2.550 & 2.100 & 3.050 \\ 3.150 & 1.800 & 2.400 & 2.100 \end{bmatrix}. \tag{21}$$

5.2.2. Solution of Combination Weights

Step 1 (determination of ideal sorting of indicator importance degree). If we invited 9 experts, of which three are complete index order relations of expert, three are order relations of expert missing one item, another three are order relations of expert missing two items, signified with A, B, and C, respectively, expert sorting of A category was $A_1 = \{U_3, U_4, U_1, U_2\}$, $A_2 = \{U_1, U_3, U_4, U_2\}$, $A_3 = \{U_1, U_3, U_2, U_4\}$; expert sorting of B category was $B_1 = \{U_3, U_1, U_2\}$, $B_2 = \{U_1, U_3, U_4\}$, $B_3 = \{U_3, U_4, U_2\}$; and expert sorting of C category was $C_1 = \{U_3, U_1\}$, $C_2 = \{U_1, U_4\}$, $C_3 = \{U_3, U_4\}$ from Section 3.2.4, we knew A category was regarded as valid as complete index sorting; B category existed indicator omission but all indicators satisfied the correction condition, so retained; C category did not satisfied the correction condition, should be discarded.

Substituted B category into formulas (12) and (13) and obtain the mean value of each indicator:

$$\overline{R}_1 = 2.5, \qquad \overline{R}_2 = 1, \qquad \overline{R}_3 \approx 2.67, \qquad \overline{R}_4 = 1.5. \tag{22}$$

Contrast the mean value with the incomplete information of expert sorting score of B category, respectively, insert the deletion indicator into the expert incomplete sorting by size, and contrast the weighted mutation rate if the same size, and all index order relations of each expert were as follows:

$$X_1 = \{U_3, U_1, U_4, U_2\}, \qquad X_2 = \{U_1, U_3, U_2, U_4\},$$
$$X_3 = \{U_3, U_1, U_4, U_2\}. \tag{23}$$

Calculate the Spearman rank correlation coefficient of one expert with the other two experts and obtain $\overline{V}_1 = 0.8$, $\overline{V}_2 = 0.6$, and $\overline{V}_3 = 0.8$, since 0.8 is greater than the critical value 0.7, the correction sorting of expert B_1 and expert B_3 passed the consistency testing and the correction sorting of expert B_2 was discarded. Conduct Spearman rank correlation coefficient test on these two correction sortings that passed the consistency testing and three complete expert sortings and obtain the mean value

$$\overline{V}_{B_1} = 0.8, \qquad \overline{V}_{B_3} = 0.8, \qquad \overline{V}_{A_1} = 0.5,$$
$$\overline{V}_{A_2} = 0.7, \qquad \overline{V}_{A_3} = 0.5. \tag{24}$$

Using the formulas (12) and (13) to calculate the mean value of expert B_1 and expert B_3 passed the consistency testing and expert A_2: $\overline{R}_1 = 10/3$, $\overline{R}_2 = 1$, $\overline{R}_3 = 11/3$, $\overline{R}_4 = 2$, sorted according to the size, then the overall sorting result was $X = \{U_3, U_1, U_4, U_2\}$, from Theorem 1, we knew it was the ideal sorting.
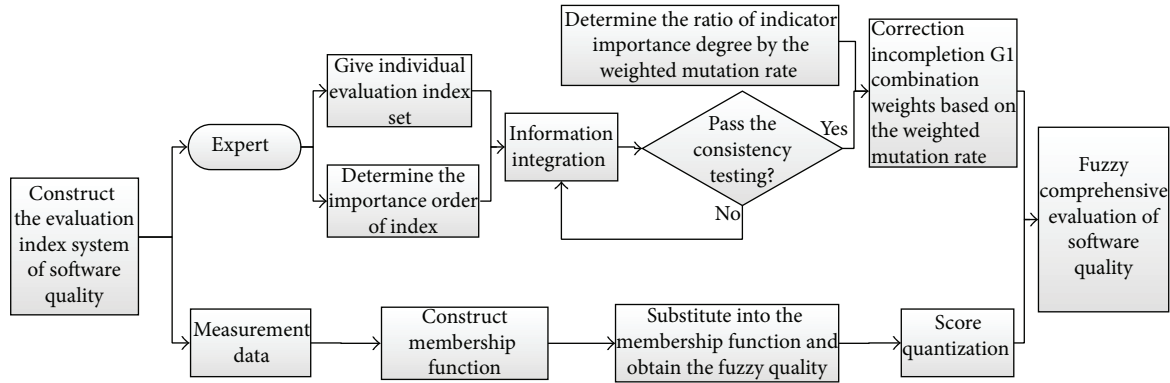
FIGURE 2: Fuzzy software evaluation schematic diagram based on weighted mutation rate correction incompletion G1 combination weights.

Step 2 (calculation of weighted mutation rate of index). We used formula (8) to calculate the weighted mutation rate of quantitative score matrix of indicators in Section 5.2.1:

$$v(U_1) = 0.0759, \qquad v(U_2) = 0.0686,$$
$$v(U_3) = 0.1028, \qquad v(U_4) = 0.1659. \tag{25}$$

Step 3 (calculation of importance degree ratio of indicators). We used formula (7) to calculate the value of weighted mutation rate of adjacent indicators:

$$r_2 = \frac{v(U_3)}{v(U_1)} = 1.35, \qquad r_3 = \frac{v(U_1)}{v(U_4)} = 1,$$
$$r_4 = \frac{v(U_4)}{v(U_2)} = 2.42. \tag{26}$$

Step 4 (determination of index weights). We substituted into formulas (4) and (5) and obtained the index weight vector $\beta = (0.266, 0.110, 0.359, 0.266)^T$.

5.2.3. Comprehensive Evaluation Result. Substitute the score matrix and the weight vector into the evaluation formula (20), obtain overall evaluation result of each software $Z_1 = 2.372$, $Z_2 = 2.670$, and $Z_3 = 2.456$, and from $\text{Score}(Z_2) > \text{Score}(Z_3) > \text{Score}(Z_1)$ we knew that the subcharacteristic suitability of functionality of software Z_2 was optimal, followed by Z_3, Z_1.

For the quality evaluation of each software characteristic, we used the same quality evaluation method as subcharacteristic. First evaluated all the subcharacteristics and then used the fuzzy comprehensive evaluation method based on weighted mutation rate correction incompletion G1 combination weights to evaluate its characteristic.

6. Conclusions

This paper considered the situation that each expert may have different individual evaluation set, and expert evaluation was expressed with uncertain information, by which expert could express individual subjective judgment flexibly, integrated the evaluation results of experts and obtained the ideal sorting, then determined the ratio of indicators by weighted mutation rate, making the index weights include subjective opinions and objective test data properly. After that, we calculated the fuzzy quality and quantified score of indicators by the construction of membership degree in fuzzy mathematics and then obtained comprehensive evaluation result by weighted summation with combination weights. This method is a combined evaluation method including quantitative and qualitative method, which could take full advantage of expert knowledge and experience and objective data obtained by testing, conforms to the characteristic of poor visibility and difficulty in measure; the evaluation process is also in compliance with human thinking judgment process and shows good flexibility, effectiveness, and rationality.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

References

[1] B. W. Boehm, J. R. Brown, and M. Lipow, "Quantitative evaluation of software quality," in *Proceedings of the of the 2nd International Conference on Software Engineering*, pp. 592–605, 1976.

[2] S. Sarkar, G. M. Rama, and A. C. Kak, "API-based and information-theoretic metrics for measuring the quality of software modularization," *IEEE Transactions on Software Engineering*, vol. 33, no. 1, pp. 14–32, 2007.

[3] F. Yue, Z. Su, Y. Lu, and G. Zhang, "Comprehensive evaluation of software quality based on fuzzy soft sets," *Systems Engineering and Electronics*, vol. 35, no. 7, pp. 1460–1466, 2013.

[4] N. J. Pizzi, "Software quality prediction using fuzzy integration: a case study," *Soft Computing*, vol. 12, no. 1, pp. 67–76, 2008.

[5] V. Y. Khramov and P. N. Besedin, "Use of fuzzy situations in assessment of software quality," *Telecommunications and Radio Engineering*, vol. 64, no. 6, pp. 455–464, 2005.

[6] Z. Shi and X. He, "Fuzzy software quality synthesis evaluation," *Systems Engineering and Electronics*, vol. 24, no. 12, pp. 121–122, 2002.

[7] J. H. Zhou, Z. Wang, Z. K. Yang et al., "Research on software quality evaluation based on fuzzy method," *Systems Engineering and Electronics*, vol. 26, no. 7, pp. 988–991, 2004.

[8] Y. A. Yang, "synthetic evaluation method for software quality," *Mini-Micro System*, vol. 21, no. 3, pp. 313–315, 2000.

[9] Y. J. Guo, *Comprehensive Evaluation Theory, Methods and Extensions*, Science Press, Beijing, China, 2012.

[10] Y. M. Zhang, G. T. Chi, and L. A. Xu, "Comprehensive evaluation of human all-round development based on entropy method: model and empirical study," *Chinese Journal of Management*, vol. 6, no. 5, pp. 1047–1055, 2009.

[11] G. T. Chi, G. Li, and Y. Q. Cheng, "The human all-round development evaluation model based on AHP and standard deviation and empirical study," *Chinese Journal of Management*, vol. 7, no. 2, pp. 301–310, 2010.

[12] Y. Wang, Y. Jiao, and H. Li, "An evolutionary algorithm for solving nonlinear bilevel programming based on a new constraint-handling scheme," *IEEE Transactions on Systems, Man and Cybernetics C*, vol. 35, no. 2, pp. 221–232, 2005.

[13] Y. Wang and C. Dang, "An evolutionary algorithm for global optimization based on level-set evolution and latin squares," *IEEE Transactions on Evolutionary Computation*, vol. 11, no. 5, pp. 579–595, 2007.

[14] G. Li, B. Wang, L. Zhou, M. Zhang, and K. Chen, "The human all-round development evaluation model based on revised G1 by standard deviation and empirical research," *System Engineering Theory and Practice*, vol. 32, no. 11, pp. 2473–2485, 2012.

[15] G. T. Chi, F. Qi, and G. Li, "The evaluation model of scientific development concept for Chinese provinces based on combination weighting of improved Group-G1 and its application," *System Engineering Theory & Practice*, vol. 33, no. 6, pp. 1448–1457, 2013.

[16] H. Hong and Q. Wang, "Analysis of the accuracy of expert judgment and method to identify the star-level of expert," *Fudan Education Forum*, vol. 10, no. 6, pp. 59–63, 2012.

[17] M. X. Sun, *Prediction and Evaluation*, Zhejiang Education Publishing House, Hangzhou, China, 1986.

[18] ISO/IEC JTC1/SC7/WG6, ISO/IEC 9126-1: Information Technology Software Quality Characteristics and Metrics—Part 1: Quality Model.

[19] ISO/IEC JTC1/SC7/WG6, ISO/IEC 14598 Part 1-Part 6: Information Technology Evaluation of Software Product.

[20] L. B. Yang and Y. Y. Gao, *Principle and Application of Fuzzy Mathematics*, South China University of Technology Press, Guangzhou, China, 2002.

[21] Z. Shi and X. G. He, "The fuzziness of software quality and its representation," *Computer Engineering and Design*, vol. 22, no. 4, pp. 1–4, 2001.

[22] Z. J. Wang and R. D. Hu, "Amended weighed coefficient of variation: the useful index for measuring the degree of income distribution equality," *The Journal of Quantitative & Technical Economics*, no. 6, pp. 134–137, 2006.

[23] X. J. Hong and J. C. Li, "A further study about some problems of the Gini coefficient," *The Journal of Quantitative & Technical Economics*, vol. 2, pp. 86–96, 2006.

[24] Z. S. Xu, "On consistency improving method in analytic Hierarchy process," *European Journal of Operational Research*, vol. 126, pp. 683–687, 2000.

[25] F. Liu, W. G. Zhang, and Z. X. Wang, "A goal programming model for incomplete interval multiplicative preference relations and its application in group decision-making," *European Journal of Operational Research*, vol. 218, no. 3, pp. 747–754, 2012.

[26] P. Groselj and L. Z. Stirn, "Acceptable consistency of aggregated comparison matrices in analytic hierarchy process," *European Journal of Operational Research*, vol. 223, no. 2, pp. 417–420, 2012.