An Applied View to Determine the Weights of Experts' Scores Based on an Evidential Reasoning Approach Under Two-Dimensional Frameworks

Xiaoqing Huang, Peng Gui, Jingui Yao, Wenxing Zhu, Chufan Zhou, Xin Li and Shaorong Li

***Abstract*—At present, many decision-making fields require collecting and organizing the opinions and evaluation values of experts. However, they evaluation values given by experts used in the decision-making should be further analysed from a scientific angle. In this paper the evidence reasoning approach under two-dimensional frameworks (ERTDF) is used to mine the relative validity of the expert scoring value. This approach takes the self-evaluated familiarity with the decision-making objects to be evaluated and the authority of the position and title as the main characteristics of the expert evaluation body. The unique characteristic information of the above experts is transformed into evidence to assist in correcting the original score results and to express their respective reliability distribution. This method also could support some intelligent decisions under the evaluation background of experts in various fields. The application results show that this method could better consider the uncertainty of the evaluation results caused by the various characteristics of the individual experts, improve the effect of experts' scores, and make the comprehensive results of multiple experts' evaluation values more reasonable and accurate.**

***Index Terms*—experts' characteristic, evidential reasoning approach (ER), two-dimensional frameworks (ERTDF), weights of experts' scores**

Manuscript submitted December 05, 2021; revised July 31, 2023.

This work was supported in part by the Science and Technology Project of Jiangxi Provincial Department of Education (GJJ220084); National Social Science Foundation of China (20XGL016) and (22XJY040); Jiangxi University of Science and Technology(22ZKZB-03) and (XZG-15-01-05) ; Jiangxi Education Reform Project (jxjg-20-85-1) and (jxjg-21-85-1); Science and Technology Project of Jiangxi Provincial Department of Education (GJJ207103); Jiangxi Humanities and Social Science Research Project (YS20107), (JJ19112) and (JJ20226).

X. Huang is a lecturer at Jiangxi University of Science and Technology Nanchang, 330000, China (e-mail: huangxiaoqing2111 @163. com).

P. Gui is a lecturer at Gongqing Institute of Science and Technology, Jiujiang, 332020, China (e-mail: lxy@ email.ncu. edu.cn).

J. Yao is a professor at the Nanchang JiaoTong Institute, Nanchang, 330000, China (corresponding author phone: +86 18179149149; fax: 0791-82129652; e-mail: justin149jq@163.com).

W. Zhu is a professor at Jiangxi University of Science and Technology Nanchang, 330000, China (e-mail: Wx_zhu52000@126.com)

C. Zhou is an undergraduate student at Jiangxi University of Science and Technology, Nanchang, 330000, China (e-mail: zhouchufan720@163. com).

X. Li is a lecturer at Jiangxi University of Science and Technology, Nanchang, 330000, China (e-mail: ry007@163.com)

S. Li is a lecturer at Jiangxi University of Science and Technology, Nanchang, 330000, China (e-mail: lishaorong2016 @163.com)

I. Introduction

EVALUATIVE and selective work is an important decision-making activity that occurs frequently, especially related to opinion summaries. As a manager of evaluation or selective work, it is necessary to consider a more effective and scientific way to deal with the summary results of multiple experts' opinions. Sometimes the simplest method may be the arithmetic average method. However, the characteristics, knowledge structure, understanding predilections, and evaluation scales of distinct experts are different, so they should not be simply summarized under the same standard [1]–[4].

Therefore, for similar decision-making problems, especially group decision-making problems, a large number of scholars, experts, leaders and other strata are dissimilar, and it is necessary to scientifically analyze the summary of experts' evaluation results, to address the verification from different angles and to improve the accuracy of the evaluation conclusions of the research object. For example, Weiye Jiang proposed adjusting the experts' scores according to the differences of experts with respect to the sensitivity of various indicators, this approach has a certain effect on boosting the different scores consistent with the objective facts [5].

II. Literature Review

In the application of expert scoring, experts with particular skills evaluate decision-making objects in various fields [1]–[22]. This scoring is widely used in many research projects with various methods [1]–[17].

Owing to incomplete or lacking information, experts cannot provide precise evaluation opinions under certain situations. To address these situations, some scholars have represented evaluation opinions by using fuzzy concepts [9]. Some authors have used fuzzy mathematics methods to deal with experts' score values in some uncertain situations, such as bipolar intuitionistic fuzzy soft expert sets and their application [10], examining the fuzzy attributes alongside gray characteristics of expert assessment [11], applying the fuzzy analytical hierarchy process (FAHP) approach [12], rough set theory [13], the intuitionistic fuzzy technique [16], and the interval evidential reasoning approach that is used to deal with fuzziness and ignorance [17].

The evaluation opinions of multiple experts should be aggregated to reduce the effect of experts' shortcomings at different levels. Research shows that the aggregated opinion is consistently superior to the opinions of individual experts, although it is collected by merely using a simple calculation of the expert scoring method [10]. For example, Liu and Qiu [6] inversely determined the evaluation level of each expert by comparing the relative deviation of the point set corresponding to the interval number of the experts' judgment matrix and the experts' comprehensive judgment matrix. For the fuzzy multiattribute group decision-making problem with completely unknown expert evaluations and attribute weights, Nan *et al.* [7] constructed a multiattribute group decision-making method based on an experts' trust network given incomplete information. Zhu *et al.* [8] discussed methods to determine the evaluation experts' weights in four cases. These methods revised the differences in experts' evaluations and evaluation results, but they are mainly considered from the external observable determination elements of experts. Shanthi *et al.* [10] found the difference in the sum of the grades for agree and disagree bipolar intuitionistic fuzzy soft expert sets with and without possibility values. Liu [14] viewed the efficiency values of each DMU in a cross-efficiency matrix as efficiency scores determined by different experts. They considered the distinction of different experts with respect to education backgrounds, work experiences and other aspects, and their efficiency scores to DMUs should be treated differently and allocated variant weights in the final overall assessment. Anderer *et al.* [15] directly scored their sleep scoring system independently by 2 experts and by a consensus scorer. Independent multiple human expert scoring was required to specify the equivocal epochs problems and to examine possible solutions.

Considering these differences, in this paper the expert evidence reasoning method based on two-dimensional frameworks (ERTDF) is adopted, and self-judgment is taken as the main subject. Among them, the I-dimensional recognition framework describes the experts' evaluation opinions for the decision-making object or the object to be selected. The Ⅱ-dimensional recognition framework reflects the difference characteristic information after the experts' self-judgment, which is used to modify and to supplement the I-dimensional evidence recognition framework and to increase the information content of the evidence recognition framework. Then, the second-dimension recognition framework helps to obtain the evidence correction factor and to modify the evaluation information of experts' scores in the I-dimension framework. Finally, the evidence reasoning method aggregates the evaluation information of multiple experts to realize the ranking and optimization of decision objects or objects to be selected. Through the example analysis, it is verified that the method is suitable and effective for decision-making projects or selection problems.

III. Evidential Reasoning Approach under Two-Dimensional Frameworks and an Influence Mechanism

A. Proposed ERTDF

The main research content of this paper is obtaining the utility value of experts' evaluation information based on an evidential reasoning approach under II-dimensional frameworks. When considering how to make decisions or choices, in most cases, multiple experts evaluate the relevant results. The results often depend on the opinions of multiple peer experts, and the scores given by multiple experts may be inconsistent. To effectively integrate the experts' scores, in this paper an evidential reasoning approach under two-dimensional frameworks is selected. The traditional framework constructed by the traditional evidential reasoning method reflects only the decision scores of experts and cannot reflect the quality of experts' decisions. Therefore, it is necessary to add one-dimensional information to reflect the multiple characteristics of experts and to modify the original decision scores. This extension is more accurate and effective use of experts' decision scores. The specific calculation steps of the method proposed in this section are as follows: ① On the basis of the original I-dimensional framework of evidential reasoning, we add the II-dimensional framework of evidential reasoning to reflect the characteristics of experts and form a II-dimensional framework; ② The II-dimension of the recognition framework generates evidence correction factors to correct the reliability distribution of experts' decision scores; ③ The ER method is used to fuse the evaluation of multiple experts; ④ The comprehensive evaluation results facilitate the comparison and ranking of multiple alternative or decision objects [9].

B. Construction of two-Dimensional Frameworks

Definition 1. Let θ_i be a possible result of a decision-making problem and $\Theta=\{\theta_1,\theta_2,\cdots,\theta_n\}$ be the set of all possible results. The elements in the set are mutually exclusive and exhaustive, that is, $\theta_i \cap \theta_j = \Phi$ for any $i, j \in \{1,2,\cdots n\}$ and $i \neq j$, where Φ is an empty set. Θ is called a framework of discernment. The power set of Θ is denoted as $P(\Theta)$ or 2^Θ, which can also be represented as follows:

$$P(\Theta) = 2^\Theta = \{\Phi,\{\theta_1\},\cdots,\{\theta_n\},\{\theta_1,\theta_2\},\cdots,\{\theta_1,\theta_n\},\cdots,\Theta\}$$

Definition 2. For a decision-making problem, on the basis of constructing the traditional framework of discernment Θ, the one-dimensional framework of discernment Ψ is added to reflect the evidence source features or the evidence acquisition process. Consequently, II-dimensional frameworks of discernment are formed, expressed as follows:

$$\Upsilon : \{\Psi_1,\Psi_2,\cdots,\Psi_n\} \rightarrow \Theta$$

Among them, Ψ:$\{\Psi_1,\Psi_2,\cdots,\Psi_n\}$ is the II-dimensional framework that is used to represent n features of the evidence source. Ψ_i represents feature i of the evidence source. Similar to Θ, the elements in Ψ_i are mutually exclusive and exhaustive, that is, $\Psi_i = \{\varphi_{i,1},\varphi_{i,2},\ldots\varphi_{i,m}\}$. For simplicity, the II-dimensional frameworks can be abbreviated as $\Upsilon : (\Psi \rightarrow \Theta)$. The II-dimensional framework is used to modify and supplement the I-dimensional framework. II-dimensional frameworks contain additional information and can assist decision-makers in improving the accuracy of their decisions.

Definition 3. If a function $m:2^{\Theta}\rightarrow[0,1]$ satisfies the following conditions:

$$\begin{cases} m(\Phi)=0 \\ \sum_{\theta\subseteq\Theta} m(\theta)=1 \\ 0\le m(\theta)\le 1 \end{cases} \tag{1}$$

then m is called a basic probability assignment (BPA) function or mass function. $\forall\theta\subset\Theta$, $m(\theta)$ called the basic probability number, and it can be interpreted as the degree of belief that the evidence supports θ. The basic probability number assigned to Θ represents the degree of global ignorance, denoted by $m(\Theta)$.

During the research project selection problem, the I-dimensional framework describes the evaluation opinions of experts, that is, the evaluation grades for the 'comprehensive evaluation' and the 'decision opinion' sets. The set of 'comprehensive evaluation' evaluation grades is expressed as:

$$\Theta_1=\{\text{excellent,good,average,poor}\}=\{\theta_{1,1},\theta_{1,2},\theta_{1,3},\theta_{1,4}\}$$

The set of 'decision opinion' evaluation grades is expressed as follows:

$$\Theta_2=\{\text{priority considered, considered,non- considered}\}=\{\theta_{2,1},\theta_{2,2},\theta_{2,3}\}$$

We assume that L experts E_i $(i=1,2,\ldots,L)$ are present to evaluate the object e_j $(j=1,2,\ldots,J)$. If e_j is evaluated to a grade $\theta_{t,n}$ on the t evaluation criterion with a belief degree $m_{t,n}^{i,j}$ by E_i, then the assessment of decision object e_j can be described by a belief distribution as follows:

$$S_1^i(e_j)=\left\{\left(\theta_{t,n},m_{t,n}^{i,j}\right),t=1,2;n=1,\ldots,N_t;(\Theta_t,m_{t,\Theta}^{i,j})\right\} \tag{2}$$

Here, $0\le m_{t,n}^{i,j}\le 1$, $\sum_{n=1}^{N_t} m_{t,n}^{i,j}\le 1$ and $m_{t,\Theta}^{i,j}=1-\sum_{n=1}^{N_t} m_{t,n}^{i,j}$, where $m_{t,\Theta}^{i,j}$ is the degree of global ignorance [9].

C. Impact Mechanism

Because the assessed value of experts contains complex information and influencing factors, which also have person-specific characteristics, the evaluation result of each expert has certain influencing factors, such as self-cognition, knowledge structure, understanding predilections, and evaluation scales. To better identify or separate the data information value with a certain deviation from the real value, in this paper the influence angle that is easier to obtain or determine as the main influence mechanism source of the experts' evaluation is selected.

Specifically, the subject of the II-dimension recognition framework represents the experts' characteristic information, which is used to reflect the quality of experts' evaluation information. Experts' characteristics are described from two perspectives in the paper.

① Experts' familiarity with the decision object. Most of the different decision objects belong to a special field, and many decision objects also involve the experts' knowledge structure background. Even peer experts may not be able to fully understand the specific situation for all decision objects. The 'familiarity' of experts in the peer evaluation opinion table reflects the experts' understanding of the relevant fields of the decision objects. Generally, the higher the familiarity of experts with the decision objects, the higher the reliability of experts' evaluation.

② The influence of experts' title and position authority. The professional titles of evaluation experts often have higher discursive power in the management of decision-making or the selection of objects. This comes from the authoritative influence brought by professional titles. The existing professional titles of experts can reflect the influence of the evaluation preference of the decision-making or selection objects to a certain extent. Generally, the higher the influence of the experts' professional titles, the higher the preference reliability of experts' evaluation of response samples.

The II-dimensional framework is constructed as follows:

If expert E_i is evaluated to a grade $\varphi_{t,m}$ on characteristic c_t with a belief degree $\beta_{t,m}^i$, then the assessment of expert E_i can be described by a belief distribution as follows:

$$S_2^i(c_t)=\left\{\left(\varphi_{t,m},\beta_{t,m}^i\right),t=1,2;m=1,2,3;\left(\Psi_t,\beta_{t,\Psi}^i\right)\right\} \tag{3}$$

Here, $0\le\beta_{t,m}^i\le 1$, $\sum_{m=1}^{3}\beta_{t,m}^i\le 1$ and $\sum_{m=1}^{3}\beta_{t,m}^i\le 1$, where $\beta_{t,\Psi}^i$ is the degree of global ignorance.

In the II-dimension recognition framework, 'familiarity' is a qualitative index, and experts directly give evaluation information according to their understanding of the subject field. 'Influence of professional title and position authority' is a quantitative index, which is obtained according to the experts' professional title and position authority statistics, and the basic reliability distribution of professional title and position is inequitable at different levels.

If the authority influence rate of decision object h_m is equivalent to a grade $\varphi_{2,m}$ of the experts' title and position authority, then the authority influence rate of decision object h_i can be transformed to the degree of belief as follows:

$$\beta_{2,m}^i=\frac{h_{m-1}-h_i}{h_{m-1}-h_m},\ \beta_{2,m-1}^i=1-\beta_{2,m}^i \text{ and } h_{m-1}\ge h_i\ge h_m \tag{4}$$

In conclusion, the II-dimensional frameworks that are constructed in this paper can be expressed as follows:

$$\Upsilon:(\Psi\rightarrow\Theta)=\left(\{\Psi_1,\Psi_2\}\rightarrow\{\Theta_1,\Theta_2\}\right)$$

After the II-dimensional frameworks are constructed, all the evidence should be aggregated. Compared with the traditional one-dimensional framework, II-dimensional frameworks can describe the features of the evidence sources by using the II-dimensional framework. First, we can fit the II-dimensional evidence information into the I-dimensional evidence information. That is, the evidence correction factor $a=\left(a_1,a_2,\ldots,a_i,\ldots,a_L\right)$, where $0\le a_i\le 1$, can be generated based on the II-dimensional evidence information, and the evidence correction factor a is used to discount the I-dimensional evidence information. Subsequently, we aggregate the discounted I-dimensional evidence information. Fig. 1 illustrates the aggregation logic.

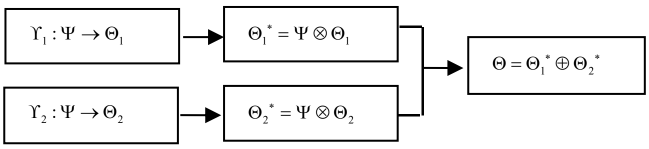

Fig. 1. Logic chart for information aggregation under II-dimensional frameworks.

The evidence correction factor $a=\left(a_1,a_2,\ldots,a_i,\ldots,a_L\right)$ is calculated by the evidential reasoning approach.

The fusion of multiple expert evaluation information is completed based on the decision matrix and the evidence correction factor. Then, the generated evidence correction factor can be used as the next step for the decision-making management party to process the data scored by each expert. Multiplication and addition can be directly considered to reduce the proportion of data with little significance. The results of comprehensive processing of expert evaluation data are more reasonable and effective, which is also a more appropriate processing method and worthy of further promotion and adoption. However, due to the variety of actual problem objects, the indicators or parameters that may be used to modify the evaluation value need to be further adjusted or replaced to better match the actual background of the specific research objects and to obtain more valuable evaluation results flexibly and effectively [9]–[11].

IV. Illustrative Example

In this paper, a set of expert information is simulated, and an example is solved and verified by using an evidential reasoning approach under II-dimensional frameworks.

First, we assume the basic setup of a group of six peer experts with a typical combination of characteristics. We design them to self-score and to evaluate the familiarity of the decision objects with the influence of professional title and position authority. We input their initial self-evaluation data into the model, and we set the consideration weight of familiarity and professional title and position authority at the same time. The initial weight of this paper was set to 0.5.

Second, according to our preference, six relatively reliable experts in the industry in the same or similar research fields for the decision objects of this paper participate in the evaluation of qualitative indicators. The experts choose the 'familiarity', 'professional title and position authority', 'comprehensive evaluation', and 'consideration' of the decision objects according to the relevant contents set by us. 'Familiarity' refers to the self-assessment of experts' familiarity with the field of the decision objects, which is divided into three levels: 'familiar', 'partly familiar' and 'less familiar', and 'familiar' is mainly used as a reference index. 'Professional title and position authority' refers to the authority influence rate of decision objects, and it is divided into five levels: 'high', 'less high', 'medium', 'less low' and 'low'. 'Comprehensive evaluation' and 'consideration opinions' are the comprehensive evaluation indices for each decision object. 'Comprehensive evaluation' can be divided into four grades (or other grades) according to the specific object. Experts evaluate the decision objects according to the evaluation criteria and select the corresponding evaluation level.

Then, we assume the relevant initial values before performing ER analysis on the experts' reliability formed after the self-assessment of these experts.

Taking an item in the data set as an example to illustrate the implementation process of this method in detail, the setting of experts' title and position authority of the initial value required for the decision-making or selection problem is shown in Table I. The characteristic information of experts' self-assessment familiarity of the initial value required for the decision-making or selection problem is shown in Table II.

TABLE I
Experts' Title and Position Authority Influence Value Setting

	Title and Position Authority	Title and Position Authority Influence(Initial Value)
Expert E_1	Less high	90
Expert E_2	Less low	80
Expert E_3	Less high	90
Expert E_4	Medium	85
Expert E_5	High	95
Expert E_6	Low	75

TABLE II
Experts' Self-assessment familiarity Influence Value Setting

	Self-assessment Familiarity	Self-assessment Familiarity Influence(Initial Value)
Expert E_1	Less familiar	0.8
Expert E_2	Partly familiar	0.6
Expert E_3	Familiar	0.95
Expert E_4	Partly familiar	0.6
Expert E_5	Less familiar	0.8
Expert E_6	Familiar	0.95

The steps of experts' evaluation information utility value estimation are as follows:

Step 1: The experts' original evaluation information and characteristic information are transformed into belief distributions by using (2) to (4). The reliability distribution of the six experts in the II-dimensional recognition framework and experts' scores for one object to be evaluated in the I-dimensional recognition framework are shown in Table III.

TABLE III
Belief Distributions of the Six Experts for the II-Dimensional Frameworks of One Object

	The I-Dimensional framework		The II-Dimensional framework	
	Comprehensive Evaluation	decision Opinion	Self-assessment Familiarity	Title and Position Authority
Expert E_1	$(\theta_{1,1}, 1)$	$(\theta_{2,2}, 1)$	$(\phi_{1,2}, 1)$	$(\phi_{2,2}, 1)$
Expert E_2	$(\theta_{1,3}, 1)$	$(\theta_{2,2}, 1)$	$(\phi_{1,3}, 1)$	$(\phi_{2,4}, 1)$
Expert E_3	$(\theta_{1,1}, 1)$	$(\theta_{2,1}, 1)$	$(\phi_{1,1}, 1)$	$(\phi_{2,2}, 1)$
Expert E_4	$(\theta_{1,2}, 1)$	$(\theta_{2,2}, 1)$	$(\phi_{1,3}, 1)$	$(\phi_{2,3}, 1)$
Expert E_5	$(\theta_{1,2}, 1)$	$(\theta_{2,1}, 1)$	$(\phi_{1,2}, 1)$	$(\phi_{2,1}, 1)$
Expert E_6	$(\theta_{1,4}, 1)$	$(\theta_{2,3}, 1)$	$(\phi_{1,1}, 1)$	$(\phi_{2,5}, 1)$

Step 2: The evidence correction factors are generated by using the characteristic information of the experts in the II-dimensional framework. The relative weights of the experts' characteristics are determined by using the direct assignment method. The two features of the experts in the II-dimensional framework are assumed to be equally important, that is, $\omega_1 = \omega_2 = 0.5$.

Then, the overall reliability distribution of each expert in the II-dimension recognition framework is obtained by the basic formula of the evidential reasoning approach. Assuming that the utility values of each level of the II-dimension recognition framework are 0.95, 0.75, and 0.65, then the evidence reliability correction factor can be calculated. We use the model calculation results directly. The operation results and the reliability data are shown in Table IV. Then, we can use a as weights for experts' scores. The calculation results generated by the data after the solution of each expert's belief distribution are also shown in Table IV. According to this method, we can further calculate the weights of multiple experts' scores under different actual combinations with two different features of experts' weights in Table V.

Step 3: According to the data generated by the evidential reasoning approach in the two-dimensional framework (ERTDF), we obtain a as the correction coefficient of the scoring results made by each expert. According to the correction coefficient, we can multiply a by the scoring value of each expert for the decision-making or selection problem to generate the final experts' effect scoring value in the I-dimensional framework. This utility value considers the reliability of the experts' self-assessment. It also comprehensively considers the scoring results of each expert on the decision-making or selection objects to ensure the effectiveness of the results given by the experts over all [14]. The belief distributions of experts in the I-dimensional framework can be modified by using the evidence correction factor and then aggregated by ER models. Table VI reflects the final results.

Step 4: Similar to the existing method of other evaluation of decision objects, the utilities of the four grades in the 'Comprehensive evaluation' are assigned scores of 9, 8, 7 and 6; the utilities of the three grades in the 'consideration' are assigned scores of 9, 7 and 5. Subsequently, the combined belief distribution of the decision object can be quantified as 7.8702 by using the ER model in the I-dimensional framework.

Step 5: Using the same model method, we calculate the original scores of the other five objects to be selected. Table VII shows the original scores of the other five objects given by the same six experts. To enhance the effectiveness of expert evaluation and verify the feasibility of the ERTDF, the evaluation results of the above six objects to be selected are sorted by using the ERTDF, ER [9] and improved TOPSIS [11] simultaneously. All of the results are shown in Tables VIII-X.

In summary, we find that the ranking results given by the ERTDF and ER are roughly consistent with the calculation results of the method proposed in this paper, especially with the method of the improved TOPSIS. It can be verified that the method used in this paper is feasible and effective for ranking evaluation. Among them, for object e_5, which has different sorting results calculated by ER, we can consider the facts that expert E_2 and expert E_6 have a lower level on self-assessment familiarity or title and position authority, their ranking results are significantly different from those of other experts', so ranking result of object e_5 perform different from other methods. The traditional ER does not reduce their scoring impact. As this illustrative example is a typical UMADM problem, the ERTDF is an effective and reasonable method to solve the UMADM problem. By using the ERTDF algorithm to synthesize the evidence of the evaluation, data loss can be avoided for uncertain information, and we can improve the effectiveness of the evaluation results. Compared with that of ER algorithms, the evaluation result of the ERTDF method used in this paper is not a simple value but an integration of evaluation grade distributions and related uncertainties. Through the verified example of various model algorithms, this paper demonstrates that this method is more reasonable and effective than the traditional ER algorithm.

In Fig. 2, to better compare the differences in the real scoring effects of experts, we compared the actual expert scores with the ERTDF evaluation results. The scores of various experts are different from the final comprehensive evaluation values. The disparity of different experts was not the same, as they had various radar displays. For example, experts E2 (α=0.9311) and E5 (α=0.8762) have better fitting, and expert E6 (α=0.7464) has the most deviation. The expert's weight of the ERTDF algorithm modified by expert E6 (α=0.7464) is also the largest, so it can be roughly and preliminarily concluded that the ERTDF algorithm is relatively effective in correcting the actual scores of different experts.

In addition, we further distinguished the differences between the scores of the experts and the final results of the ERTDF. We calculated the average absolute deviation (AAD) of each expert's score without correcting by ERTDF, as shown in Fig. 3. Among them, the AADs of expert E2 and expert E5 are relatively small, and the AAD of expert E6 is relatively large. In ERTDF, we previously assigned lower correction weights to experts with larger AAD to reduce their relative impact on the final results, such as the weights of expert E2, whose AAD was the lowest, set to 0.9311. The weight of expert E6 who had the largest AAD, set to 0.7464. Then, the purpose of improving the effectiveness of overall scores could be realized.

To facilitate comparison, we sorted the results calculated by the improved TOPSIS method and reset the scores. The rule was that the first ranking score equals 9, the second equals 8.5, the third equals 8, and so on. Then, we reset the scores of 6 objects as follows: 8, 9, 7.5, 6.5, 7, 8.5 [19], [20].On this basis, to better verify the quality of experts' scoring, we multiplied the experts' scoring value by the weight obtained in the II-dimension under the ERTDF and compared it with the three evaluation methods. According to Figs. 4, we found that the results of the ERTDF and the improved TOPSIS algorithms are more consistent and closer to the trend of different experts' scores. Comparing the ERTDF and the improved TOPSIS with the ER algorithm, there is still a small amount of discrepancy, so we could select the ERTDF algorithm as the better algorithm. Among them, due to the different calculation methods, the data calculated by the improved TOPSIS method are quite different from the previous two methods.

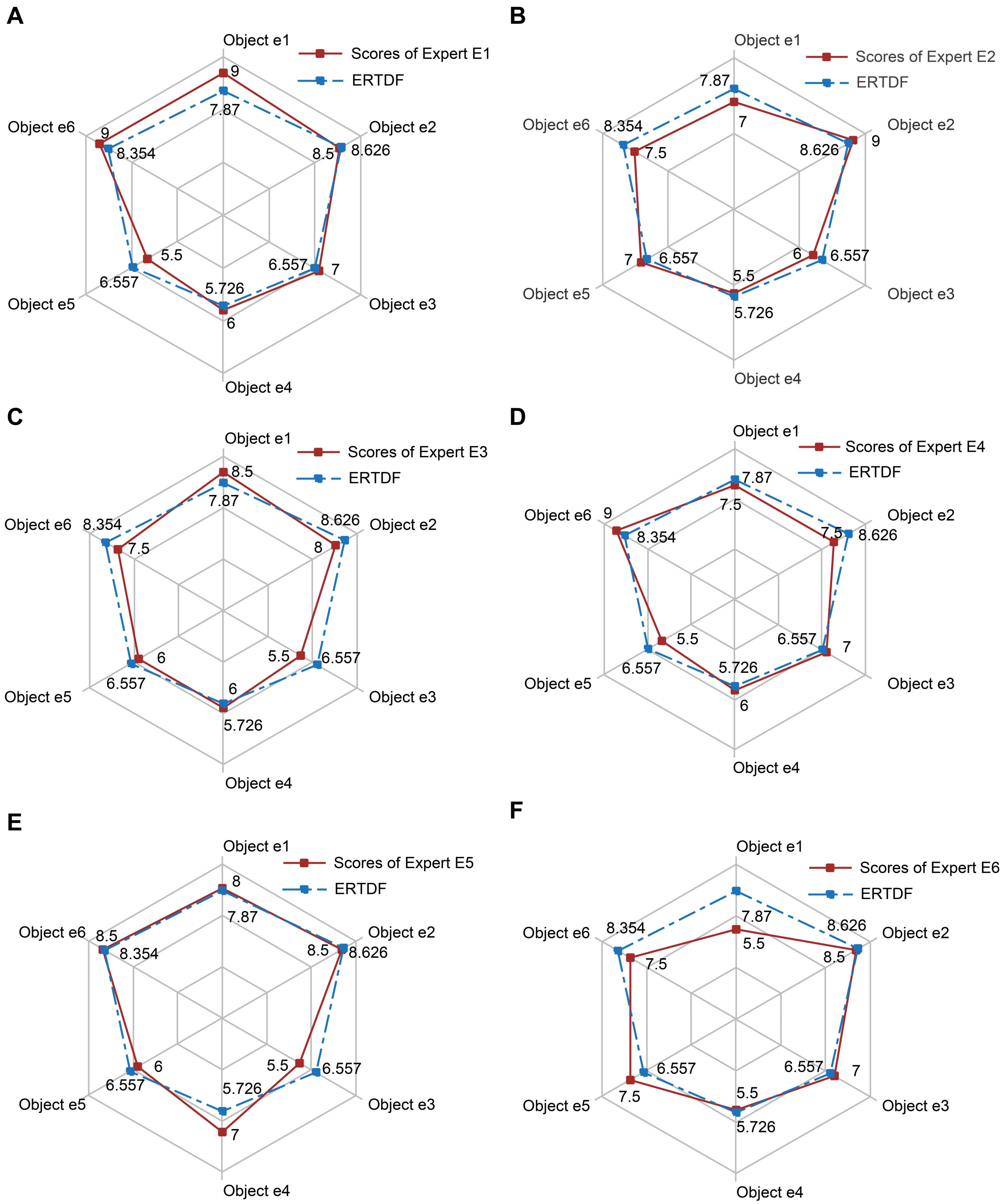

Fig. 2. Comparison of the experts' scores with the evaluation results of the ERTDF.

TABLE IV
COMBINED BELIEF DISTRIBUTION AND THE EVIDENCE CORRECTION FACTOR OF EACH EXPERT IN THE II-DIMENSIONAL FRAMEWORK

	The Belief Distribution in the Ⅱ-Dimensional Framework	Weights of Expert's Score(a)
Expert E_1	(Φ1,0.6822), (Φ2,0.1395), (Φ3,0.0233)	0.8762
Expert E_2	(Φ1,0.1196), (Φ2,0.5792), (Φ3,0.2802)	0.7464
Expert E_3	(Φ1,0.9286), (Φ2,0.0714), (Φ3,0.0000)	0.9311
Expert E_4	(Φ1,0.3543), (Φ2,0.3776), (Φ3,0.2861)	0.8141
Expert E_5	(Φ1,0.3529), (Φ2,0.4047), (Φ3,0.2424)	0.8762
Expert E_6	(Φ1,0.3529), (Φ2,0.4047), (Φ3,0.2424)	0.7941

TABLE V
WEIGHTS FOR DIFFERENT TYPES OF EXPERTS' SCORES WITH VARIOUS WEIGHTS WITH TWO FEATURES FOR THE EXPERTS

	Self-assessment Familiarity	Title and Position Authority	Weights of Different Types of Expert's Score								
	ω_1	ω_2=1-ω_1	ω_1=0.1	ω_1=0.2	ω_1=0.3	ω_1=0.4	ω_1=0.5	ω_1=0.6	ω_1=0.7	ω_1=0.8	ω_1=0.9
Type 1	Familiar	High	0.9494	0.9474	0.9436	0.9381	0.9311	0.9233	0.9154	0.9081	0.9019
Type 2	Familiar	Less High	0.9494	0.9474	0.9436	0.9381	0.9311	0.9233	0.9154	0.9081	0.9019
Type 3	Familiar	Medium	0.8534	0.8583	0.8648	0.8722	0.8799	0.8868	0.8920	0.8953	0.8969
Type 4	Familiar	Less Low	0.7517	0.7580	0.7710	0.7907	0.8156	0.8416	0.8643	0.8810	0.8915
Type 5	Familiar	Low	0.7028	0.712	0.7307	0.7589	0.7941	0.8299	0.8598	0.8804	0.8921
Type 6	Less Familiar	High	0.9480	0.9406	0.9259	0.904	0.8762	0.8464	0.8188	0.7965	0.7804
Type 7	Less Familiar	Less High	0.9480	0.9406	0.9259	0.904	0.8762	0.8464	0.8188	0.7965	0.7804
Type 8	Less Familiar	Medium	0.8479	0.8434	0.8361	0.826	0.8141	0.8016	0.7902	0.7810	0.7743
Type 9	Less Familiar	Less Low	0.7502	0.7510	0.7525	0.7547	0.7575	0.7605	0.7636	0.7662	0.7684
Type 10	Less Familiar	Low	0.7018	0.7057	0.7123	0.7214	0.7323	0.7435	0.7535	0.7613	0.7667
Type 11	Partly Familiar	High	0.9477	0.9393	0.9231	0.8990	0.8686	0.8353	0.8034	0.7761	0.755
Type 12	Partly Familiar	Less High	0.9477	0.9393	0.9231	0.8990	0.8686	0.8353	0.8034	0.7761	0.755
Type 13	Partly Familiar	Medium	0.8479	0.8434	0.8361	0.8260	0.8141	0.8016	0.7902	0.7810	0.7743
Type 14	Partly Familiar	Less Low	0.7499	0.7495	0.7488	0.7477	0.7464	0.7449	0.7434	0.7420	0.7409
Type 15	Partly Familiar	Low	0.7012	0.7035	0.7073	0.7123	0.7182	0.7243	0.7299	0.7345	0.7378

TABLE VI
RESULTS GENERATED BY AGGREGATING THE EVALUATION INFORMATION OF SIX EXPERTS

Comprehensive evaluation	Excellent	Good	Average	Poor
	0.4793	0.3034	0.2171	0.3375
Consideration	Priority considered	Considered	Non-considered	
	0.3418	0.4887	0.1695	

TABLE VII
ORIGINAL EVALUATION INFORMATION OF THE OTHER FIVE OBJECTS WITH DIFFERENT CATEGORIZATION RESULTS

Object	Expert E_1	Expert E_2	Expert E_3	Expert E_4	Expert E_5	Expert E_6
Object e_2	$\theta_{1,2}\theta_{2,1}$	$\theta_{1,1}\theta_{2,1}$	$\theta_{1,1}\theta_{2,2}$	$\theta_{1,2}\theta_{2,2}$	$\theta_{1,2}\theta_{2,1}$	$\theta_{1,2}\theta_{2,1}$
Object e_3	$\theta_{1,3}\theta_{2,2}$	$\theta_{1,3}\theta_{2,3}$	$\theta_{1,4}\theta_{2,3}$	$\theta_{1,3}\theta_{2,2}$	$\theta_{1,4}\theta_{2,3}$	$\theta_{1,3}\theta_{2,2}$
Object e_4	$\theta_{1,3}\theta_{2,3}$	$\theta_{1,4}\theta_{2,3}$	$\theta_{1,3}\theta_{2,3}$	$\theta_{1,3}\theta_{2,3}$	$\theta_{1,3}\theta_{2,2}$	$\theta_{1,4}\theta_{2,3}$
Object e_5	$\theta_{1,4}\theta_{2,3}$	$\theta_{1,3}\theta_{2,2}$	$\theta_{1,3}\theta_{2,3}$	$\theta_{1,4}\theta_{2,3}$	$\theta_{1,3}\theta_{2,3}$	$\theta_{1,2}\theta_{2,2}$
Object e_6	$\theta_{1,1}\theta_{2,1}$	$\theta_{1,2}\theta_{2,2}$	$\theta_{1,2}\theta_{2,2}$	$\theta_{1,1}\theta_{2,1}$	$\theta_{1,2}\theta_{2,1}$	$\theta_{1,2}\theta_{2,2}$

TABLE VIII
THE BELIEF DISTRIBUTION OF DIFFERENT OBJECTS BASED ON ERTDF

The Belief Distribution Based on ERTDF	Object e_1	Object e_2	Object e_3	Object e_4	Object e_5	Object e_6
Worst	18.62%	0.00%	55.07%	58.34%	65.02%	0.00%
Poor	19.04%	12.47%	22.47%	9.98%	16.41%	21.55%
Average	19.04%	12.47%	22.47%	9.98%	16.41%	21.55%
Good	43.30%	75.06%	0.00%	21.70%	2.16%	56.91%

TABLE IX
THE BELIEF DISTRIBUTION OF DIFFERENT OBJECTS BASED ON ER

The Belief Distribution Based on ER	Object e_1	Object e_2	Object e_3	Object e_4	Object e_5	Object e_6
Worst	16.75%	0.00%	56.68%	57.52%	65.20%	0.00%
Poor	20.99%	12.32%	21.66%	10.68%	16.46%	21.66%
Average	20.99%	12.32%	21.66%	10.68%	16.46%	21.66%
Good	41.26%	75.36%	0.00%	21.12%	1.88%	56.68%

TABLE X
RANKING RESULTS OF EACH OBJECT TO BE DETERMINED BY DIFFERENT ALGORITHMS

Algorithms	Comprehensive evaluation value of each object						Ranking results of each object
	e_1	e_2	e_3	e_4	e_5	e_6	
ERTDF	7.8702	8.6259	6.6747	5.7260	6.5571	8.3544	$e_2>e_6>e_1>e_3>e_5>e_4$
ER	7.8669	8.6304	6.6498	6.9540	6.5502	8.3502	$e_2>e_6>e_1>e_4>e_3>e_5$
Improved TOPSIS	0.2005	0.2651	0.1050	0.0769	0.1007	0.2519	$e_2>e_6>e_1>e_3>e_5>e_4$

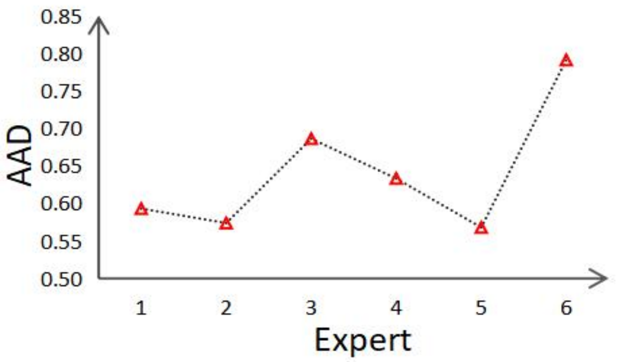

Fig. 3. The average absolute deviation (AAD) between the scores of the experts without correcting and the final results of the ERTDF.

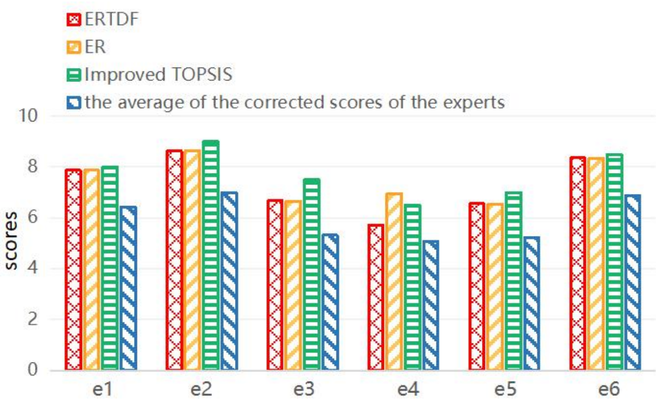

Fig. 4. Comparison of the average of the corrected scores of the experts by ERTDF with the evaluation results of three different models.

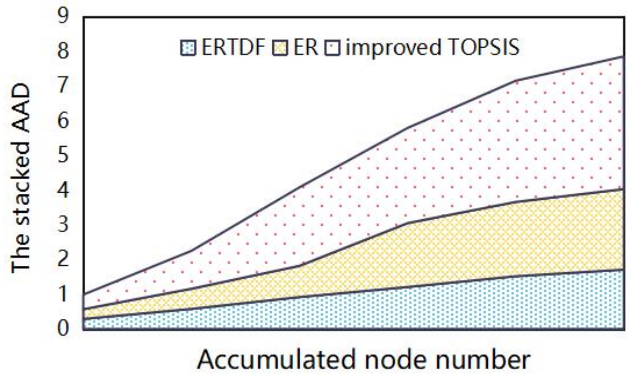

Fig. 5. The stacked AAD without weight correction given by ERTDF

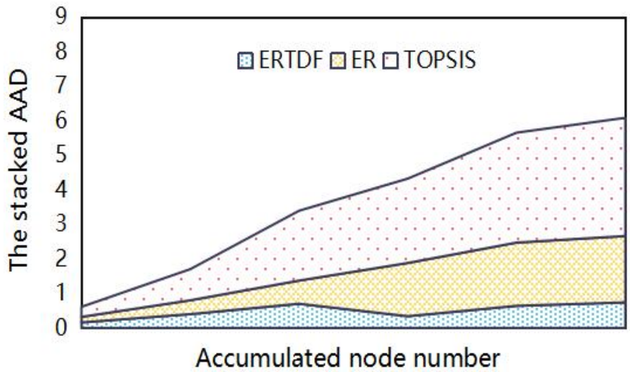

Fig. 6. The stacked AAD with weight correction given by ERTDF

To better compare the effects of using ERTDF, we accumulated the AAD without weight correction given by ERTDF and the AAD with weight correction given by ERTDF in the same model, resulting in a stacked graph of two AAD accumulations, as shown in Figures 5 and 6.

The display of the AAD accumulation stacking diagram of the below two figures, compared to the final value calculated by the models of ERTDF, ER, and improved TOPSIS, shows that the total AAD of the experts' scoring results with weight correction given by ERTDF is relatively smaller, indicating that the application of the model has a certain significance to reduce differences and errors.

Showing as the insufficient differences in the graphical representations of Figures 5 and 6, this paper took "the average of the corrected scores of the experts" as the dependent variable, and used the Improved TOPSIS, ER, and ERTDF as independent variables for OLS regression analysis. Robust standard error regression method was used for this study. From the TABLE XI below, it can be seen that the

R-squared value of OLS regression is 0.998, which means Improved TOPSIS, ER, ERTDF can explain the 99.84% change in the average of the corrected scores of the experts. When conducting F-test on the OLS, it was found that it passed the F-test (F=640.978, p=0.002<0.05), which means that at least one method of Improved TOPSIS, ER, or ERTDF will have an impact on the average of the corrected scores of the experts. The specific impact relationship is shown in the following figure.

TABLE XI
OLS REGRESSION ANALYSIS RESULTS

	Regression coefficient
Constant	-0.268 (-0.687)
Improved TOPSIS	-0.145 (-1.110)
ER	0.463** (9.149)
ERTDF	0.535** (5.036)
R^2	0.998
Adjusted R^2	0.996
F	F (3,2)=640.978,p=0.002

Dependent variable：the average of the corrected scores of the experts

D-W：2.541

* p<0.05 ** p<0.01

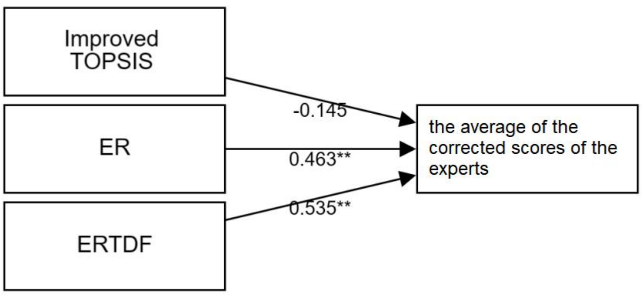

Fig.7. The Impact Relationship Diagram

Compared with ER and Improved TOPSIS, ERTDF can have the best relationship with the the average of the corrected scores of the experts, which means that the ERTDF performs the best fit for such way to make decision-making.

By extension, due to the limited length of this paper and the limited number of experts and application objects involved, it is not easy to demonstrate that ERTDF is equally effective in other or more environmental contexts. Therefore, one of the author in our team of this paper had published a SCI paper in JCR II, who using ERTDF as an algorithm model, which can better demonstrate the effectiveness of ERTDF [9]. That paper studied on the evaluation process of 19 projects by multiple experts, according to the characteristics of the experts themselves and the accuracy of previous 19 projects, it was found that ERTDF can describe the knowledge background and historical evaluation performance of experts and sequentially determine the correlation of experts, and the aggregated results are in line with the actual situation of expert evaluations, as shown in the following Figure 8.

Then, the TABLE XI shows that the 29 projects can further be screened in the panel evaluation by ERTDF algorithm, and the final project funding rate is 72%. The results obtained by using the ERTDF algorithm in the TABLE XI have a high fit with the actual results, indicating that ERTDF can effectively combine the quantitative transformation results of expert unique information and achieve the evaluation with smaller errors.

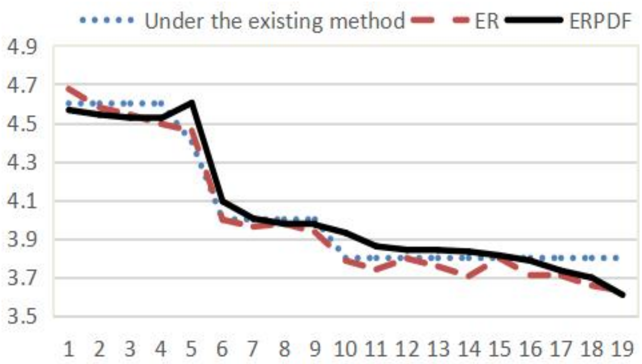

Fig.8. The Impact Relationship Diagram

TABLE XI
THE CATEGORIZING RESULTS OF 100 PROJECTS

Grade	The categorizing results by ERTDF		
	Number of projects	Number of actual funded projects	Number of actual non-funded projects
A(⩾4.8)	1	1	0
A-(⩾4.6)	1	1	0
B(⩾4)	18	14	4
E(⩾3.8)	9	5	4
C(<3.8)	71	0	71
Total	100	21	79

Therefore, through the above comparison from different perspectives, we see that the ERTDF algorithm can be used to scientifically adjust the experts' weights in advance so that the experts' scores can be better used more practically. Moreover, the error or uncertainty caused by the experts' subject familiarity and professional title or authority can be abated or reduced, ultimately improving the scoring quality.

V. CONCLUSION

In this paper, we start from the utility of experts' internal evaluation of their own scoring value, assuming that experts' reliability is directly related to the degree of self-assessment familiarity and the influence of professional title and position authority. The selection of experts' characteristic attributes required by various decision objects may be different. The selection can be redesigned during specific implementation. If the evaluation of some decision objects does not consider the characteristic response of experts' titles and positions, they can be replaced and adjusted as needed [21].

To some extent, this method could mine some information that we cannot directly obtain from intuitive judgment, and the information that has a direct impact on the evaluation results is hard to score. However, the experts themselves can be variant, so it has a high degree of dependence on experts' own self-cognition. During specific implementation, we can also improve the credibility of experts' self-cognition by providing corresponding reference standards for self-evaluation and by improving the effectiveness of the evaluation results. The important advantage of this method is that expands the traditional evidence reasoning approach to II-dimensional framework modeling; that is, on the basis of constructing the traditional I-dimensional recognition framework to describe the experts' evaluation opinions, we construct the II-dimensional framework to describe the influence value of the self-assessment familiarity and the professional title and position authority under the experts' self-awareness. This approach reflects the reliability of the evaluation results given by the experts. The quality of the I-dimensional evidence information is improved by the II-dimensional evidence information to modify the I-dimensional evidence information. This modification means that the synthesized results can better meet the actual situation of the final evaluation object and can realize more effective integration of multiple expert opinions. In addition, this new method makes comprehensive use of experts' characteristic information and experts' decision information, improves the quality of experts' decision information, and can be applied to general experts' decision-making problems [22].

In brief, by the method proposed in this paper, the weights of experts' feature attributes and the utility value of the feature level are set by a direct assignment method, which has a certain subjectivity. In follow-up research, we can also propose a more appropriate method to scientifically set the framework for experts' attribute weight features and feature - level utility indexes according to specific problems to optimize the model. Thus, the model could integrate multiple experts' opinions more reasonably and fairly and complete the scientific evaluation of decision-making or selection objects more reasonably[23].

REFERENCES

[1] F. Xiang, "Expert evaluation of core reviewers of scientific and technological journals," *News Res. Guide,* vol. 11, no. 22, pp. 213–214, 2020.

[2] M. L. Yue, F. S. Li, H. B. Tang, X. H. Lu, and T. C. Ma, "Consistency eva- luation method of periodical reviewers and its effectiveness verification," *Data Anal. Knowl. Discovery,* no. 10, 2020.

[3] S. N. Zhan, *Comprehensive Evaluation of Blast Furnace Condition Based on Expert Knowledge and Data*. China: Inner Mongolia University of Science and Technology, 2020.

[4] L. Y. Zhang and X. R. Chen, "Research on citation behavior of experts as evaluators," *Foreign Lang. Foreign Lang. Teaching,* no. 6, pp. 121–130, 151, 2020.

[5] W. Y. Jiang, *Research on Heterogeneity of Expert Evaluation*. China: Yunnan University of Finance and Economics, 2020.

[6] W. Liu and J. D. Qiu, "Research on optimal aggregation of interval number judgment matrix and evaluation of expert evaluation level of group decision making," *J. Chifeng Univ. (Natural Sci. Ed.),* vol. 36, no. 9, pp. 1–6, 2020.

[7] N. Nan, Y. J. Dou, J. Jiang, K. W. Yang, and Y. J. Tan, "Incomplete infor - mation weapon selection decision based on expert trust network," *Syst. Eng. Theory Pract.,* no. 9, 2020.

[8] G. C. Zhu, D. P. Zhao, and S. Y. Zhang, "Fuzzy multi-attribute group decision making method based on determining evaluation expert weight scheme," *J. Guangdong Normal Univ. Technol.,* vol. 41, no. 3, pp. 56–66, 2020.

[9] W. Zhu, S. Li, Q. Ku, and C. Zhang, "Evaluation information fusion of scientific research project based on evidential reasoning approach under two-dimensional frames of discernment," *IEEE Access,* vol. 8, pp. 8087–8100, 2020.

[10] S. Shanthi, N. Nishadevi, M. Saranya, and S. Sampathu, "Bipolar intuitionistic fuzzy soft expert score functions," *Mater. Today Proc.,* vol. 49, pp. 1880–1884, 2021.

[11] S. S. Shang, W. F. Lyv, and L. J. Luo, "Improved grey FMEA evaluation with interval uncertain linguistic variables and TOPSIS," *Eng. Lett.,* vol. 29, no. 2, pp. 516–525, 2021.

[12] T. K. L. Nguyen, Q. M. Le, and T. H. D. Vu, "Fuzzy analytical hierarchy process approach of attracting investment capital into industrial parks in Hai Duong province, Vietnam," *Eng. Lett.,* vol. 29, no. 3, pp. 1083–1088, 2021.

[13] T. Kashima, S. Matsumoto, and H. Ishii, "A well-balanced menu planning with fuzzy weight," *Eng. Lett.,* vol. 16, no. 3, pp. 419–424, 2008.

[14] P. Liu, "Determination of the weights for the ultimate cross efficiency using expert scoring method," *Eng. Lett.,* vol. 29, no. 3, pp. 1035–1043, 2021.

[15] P. Anderer, A. Moreau, M. Ross, A. Cerny, and S. Thusoo, "0306 scoring of sleep according to AASM: Artificial intelligence versus multiple human expert scorings," *Sleep,* vol. 41, no. suppl_1, pp. A118, 2018.

[16] İ. Otay and M. Jaller, "Multi-expert disaster risk management & response capabilities assessment using interval-valued intuitionistic fuzzy sets," *J. Intell. Fuzzy Syst.,* vol. 38, pp. 835–852, 2020.

[17] W. Liang and Y. M. Wang, "A probabilistic interval-valued hesitant fuzzy gained and lost dominance score method based on regret theory," *Comput. Ind. Eng.,* vol. 159, p. 107532, 2021.

[18] G. Tabor, K. Nankervis, J. Fernandes, and J. Williams, "Generation of domains for the equine musculoskeletal rehabilitation outcome score: Development by expert consensus," *Animals (Basel),* vol. 10, no. 2, p. 203, 2020.

[19] J. Yang, "Research on the forecasting of enterprise credit scoring based on SVR model," *Academic J. Eng. Technol. Sci.,* vol. 3, no. 1, pp. 55–64, 2020.

[20] T. Harasawa, K. Kawai-Kowase, J. Tamura, and M. Nakamura, "Accurate and quick predictor of necrotizing soft tissue infection: Usefulness of the LRINEC score and NSTI assessment score," *J. Infection Chemotherapy,* vol. 26, no. 4, pp. 331–334, 2020.

[21] J. Sullivan, E. Bak, M. J. Sullivan, and P. K. Gurnani, "Predictive value of scoring tools in determining heparin-induced thrombocytopenia in patients on extracorporeal membrane oxygenation," *Perfusion,* vol. 35, no. 5, pp. 378–383, 2020.

[22] X. Huang, X. Lu, Y. Sun, J. Yao, and W. Zhu, "A comprehensive performance evaluation of Chinese energy supply chain under "double-carbon" goals based on AHP and three-stage DEA," *Sustainability,* vol. 14, no. 16, p. 10149, 2022.

[23] X. Huang," Research on Intelligent Decision Optimization Based on Multiple Model Construction and Solution - Taking Sinotrans Logistics as an Example ," *Beijing Publishing House*, ISBN: 978-7 -200-17460-1, 2022.