Micro-Extended Belief Rule-Based System with Activation Factor and Parameter Optimization for Industrial Cost Prediction

Suhui Wang [a], Fei-Fei Ye [b, c], Long-Hao Yang [a, c,*], Jun Liu [c], Hui Wang [c], and Luis Martínez [c, d]

[a]Decision Sciences Institute, Fuzhou University, Fuzhou, PR China

[b]School of Public Management, Fujian Normal University, Fuzhou, PR China

[c]School of Computing, Ulster University, Northern Ireland, UK

[d]Department of Computer Science, University of Jaén, Jaén, Spain

*The corresponding author: more026@hotmail.com

Abstract: Industrial cost is a comprehensive indicator to reflect industrial behaviors, whereas industrial data with annotation are rare because the annotation process is very expensive. Data increment transformation is a feasible solution to enrich annotated industrial data, but it bring a new challenge in system modeling because the size of transformational data is the quadratical relationship with that of collected data, and even turn into big data problem. Hence, a novel rule-based system proposed for handling big data problems, called micro-extended belief rule-based system (Micro-EBRBS), is introduced for industrial cost prediction. Firstly, the Micro-EBRBS is improved by 1) the use of activation factor to revise the calculation of individual matching degrees; 2) the use of parameter optimization to determine the optimal value of basic parameters. Afterwards, on the basis of data increment transformation, a novel industrial cost prediction model, called data increment-based Micro-EBRBS (DIME) model, is developed to accurately predict industrial costs. In case study, 13 state-own holding industries with historical data from 1999 to 2019 in China are used to illustrate the effectiveness of the DIME model. Comparative results show that the DIME model is more accurate than some existing models in industrial cost prediction.

Keywords: Extended belief rule base; Data increment;Parameter optimization; Activation factor; Industrial cost prediction

Notations

U_i	The ith (i=1,…, T) antecedent attribute
$A_{i,j}$	The jth (j=1,…, J_i) referential value of the ith antecedent attribute
D_n	The nth (n=1,…, N) consequent of consequent attribute
$\alpha_{i,j}^{j_1\cdots j_M}$	The belief degree of referential value $A_{i,j}$ for extended belief rule $R_{j_1\cdots j_M}$
$\beta_n^{j_1\cdots j_M}$	The belief degree of consequent D_n for extended belief rule $R_{j_1\cdots j_M}$
β_n	The integrated belief degree of consequent D_n.
$\theta_{j_1\cdots j_M}$	The weight of extended belief rule $R_{j_1\cdots j_M}$
δ_i	The weight of antecedent attribute U_i
$u(A_{i,j})$	The utility value of referential value $A_{i,j}$ in the ith antecedent attribute U_i;
$(\boldsymbol{x}_t, y_t)$	The tth collected input-output data pair.
R_t	The tth extended belief rule generated from the tth collected input-output data pair
$\alpha_{i,j}^t$	The belief degree of reference value $A_{i,j}$ generated from input data $x_{t,i}$
$D(A_{i,j_i}; i=1,...,M)$	The division domain related to M referential value A_{i,j_i}
$R_{j_1\cdots j_M}$	The extended belief rule related to $D(A_{i,j_i}; i=1,...,M)$
$S^{j_1\cdots j_M}(x_i, U_i)$	The individual matching degree of U_i of $R_{j_1\cdots j_M}$
$S_\lambda^{j_1\cdots j_M}(x_i, U_i)$	The individual matching degree of U_i of $R_{j_1\cdots j_M}$ based on activation factor λ
$w_{j_1\cdots j_M}$	The activation weight of extended belief rule $R_{j_1\cdots j_M}$

$AR(\boldsymbol{x})$	A set of activated rules to reply the given input data $\boldsymbol{x}$.
$CS(\boldsymbol{x})$	A candidate set of historical input-output industrial data pairs to reply the given input data $\boldsymbol{x}$
$(\Delta\boldsymbol{x}_{t,s}, \Delta y_{t,s})$	The data increment of the tth and the sth collected input-output data pairs
$f(\Delta\boldsymbol{x}_s)$	The inference output to reply data increment $\Delta\boldsymbol{x}_s$
$f(\boldsymbol{x})$	The inference output to reply the given input data $\boldsymbol{x}$

1. Introduction

Industrial cost is a comprehensive indicator to reflect industrial behaviors, *e.g.*, the quality of business management and the level of labor consumption, an effective cost planning is therefore conducive to improve production efficiencies and promote the sustainable development of industries. An effective prediction model can avoid industries' production risks to a certain extent and make reasonable allocations of input resources according to the target production output, as well as finally promote the allocations of industry resources and enhance the competitiveness of industrial production. According to China Statistical Yearbook, the industrial costs increased by 6.7% in the past 16 years, and the ratio of sales cost to sales revenue reached 81.5% from the latest data [1]. High production and sales costs bring serious challenges to the long-term development of industries. As a result, how to carry out a reasonable cost planning based on the target output and improve the cost control level of the industry is becoming an urgent problem to be solved.

Industrial costs-related studies have attracted the attention of many scholars in recent decade and the research topics mainly focused on the cost-benefit analysis in industrial production [2], cost efficiency of different industries [3], carbon dioxide abatement costs of industries [4] and the low cost of industry wastewater treatment [5]. The scholars also studied on power prediction [6] and energy consumption [7] in industrial production, *i.e.*, Ma *et al*. indicated that predicting a price range is practical and desirable [8]. For the construction of cost prediction model, Chakraborty *et al*. developed a new construction cost prediction model using hybrid natural and light gradient boosting [9]; Jiang *et al*. proposed the cost prediction model for products remanufacturing judgment based on backward propagation artificial neural network [10].

From the previous studies on proposing industrial cost prediction models, it can be found that a limited number of industrial data is a serious but common problem, which would result in the over-fitting of a cost prediction model because of the lack of training data. Moreover, as the annotation process can be very expensive, it is difficult to collect industrial data. For example, in the case of predicting total assets of state-own holding industries, the available data can be collected from Chinese industrial Yearbooks and its scale only includes 1999 - 2019. This poses a special challenge in data analysis and addressing such small data challenges needs special methodologies in what can be called data increment transformation, *e.g.*, the original data x_1, x_2, and x_3 can be transformed into $\Delta x_{1,2}=x_1-x_2$, $\Delta x_{2,1}=x_2-x_1$, $\Delta x_{1,3}=x_1-x_3$, $\Delta x_{3,1}=x_3-x_1$, $\Delta x_{2,3}=x_2-x_3$, and $\Delta x_{3,2}=x_3-x_2$, resulting in the fact that the size of transformational data is the quadratical relationship with that of collected data and it bring a new challenge in system modeling because the size of new data sometime will be large scale.

In this context, the motivations of this study include: 1) the existing studies on industrial cost prediction rarely carried out the cost prediction researches according to the target production output of industries, so it is worthy of considering the production output of industries to propose a novel cost prediction model; 2) the data increment transformation enlarges the data scale of industrial cost prediction, and even turn into big data problem, so it is necessary to construct the cost prediction model based on big data technique.

Extended belief rule-based system (EBRBS) was proposed by Liu *et al*. [11] and it has been successfully applied to various kinds of prediction problems, *e.g.*, bridge risk assessment [12], and consumer references prediction [13]. Recently, the EBRBS showed its excellent capacity in cost prediction modeling, Wang *et al*. proposed a joining learning method for the cost prediction of environmental management [14]. The results revealed that the EBRBS can be applied in cost prediction and has higher accuracy than other models. However, the construction of an EBRBS is required to transform one collected data into one extended belief rule, leading to the fact that the computing efficiency of EBRBS has to be weakened in the situation of large numbers of data. As an extension of EBRBS to big data, Micro-EBRBS [15] was recently developed to handle big data problems. For this purpose, the Micro-EBRBS is introduced in this study to achieve industrial cost prediction with overcoming the following challenges:

(1) For the construction scheme of Micro-EBRBS, there are many basic parameters needed to be determined. Previous studies were mainly based on expert knowledge to assign the value of basic parameters, which has certain subjectivity and sometime is impossible for experts because of the lack of necessary data and information. Therefore, how to determine the optimal value of basic parameters for Micro-EBRBS is one of the challenges to be solved.

(2) For the inference scheme of Micro-EBRBS, the calculation of individual matching degrees is easily affected by subjective factors and previous studies utilized a boundary value 1 to calculate individual matching degrees, leading to the inconsistency and incompleteness of Micro-EBRBS when the Euclidean distances for all rules are greater than 1 or the Euclidean distances are all smaller than 0. Hence, it is necessary to find a solution to determine the boundary value.

In order to overcome the above-mentioned challenges, parameter optimization is introduced to optimize the values of basic parameters, instead of the use of expert knowledge, so that Micro-EBRBS is constructed based on the optimal value of basic parameters. Moreover, activation factor is used to revise the calculation of individual matching degrees, so that Micro-EBRBS not only is able to avoid inconsistency and incompleteness issues, but also can activate consistent rules to produce inference output for each input data. Furthermore, by utilizing data increment transformation to enrich training data, a new industrial cost prediction model, called data increment-based Micro-EBRBS (DIME) model, is proposed in the present work. The contributions of the present work can be therefore summarized as follows:

(1) Parameter optimization is introduced to improve the construction scheme of Micro-EBRBS, so the improved Micro-EBRBS has the optimal value of basic parameters to increase the accuracy of industrial cost prediction.

(2) Activation factor is defined to improve the inference scheme of Micro-EBRBS, so the improved Micro-EBRBS has a dynamic way to activate consistent rules and accurately predict industrial costs for given input data.

(3) The improved Micro-EBRBS with data increment transformation is used to propose a DIME model for overcoming the challenge that large scale of transformational data are available for industrial cost prediction modeling.

To verify the effectiveness and accuracy of the proposed DIME model, 13 state-owned holding industries in China and the corresponding input-output data pairs from 1999 to 2019 from Chinese industrial Yearbooks are collected to provide the illustration of constructing the DIME model based on the Micro-EBRBS and data increment. Comparative studies are also carried out to show excellent accuracy of the DIME model better than some existing industrial cost prediction models.

The remainder of the paper is organized as follows: Section 2 introduces the basic of Micro-EBRBS. Section 3 provides the improvements of Micro-EBRBS based on parameter optimization and activation factor. Section 4 proposes the DIME model based on the Micro-EBRBS and data increment, and Section 5 provides the case study of comparative analysis for the DIME model. Finally, Section 6 concludes this work.

2. Basics of Micro-EBRBS

In this section, the basics of Micro-EBRBS and its construction scheme and inference scheme are introduced firstly, followed by the introduction of potential challenges for Micro-EBRBS application.

2.1. Micro-EBRBS and its construction scheme

Micro-EBRBS is an advanced rule based-system and consists of a series of extended belief rules in the unit of division domain [15]. Suppose that there are M antecedent attributes U_i (i=1,…, M) with J_i referential values $A_{i,j}$ (j= 1,…, J_i) and one consequent attribute D with N consequents D_n (n=1,…, N). Therefore, the extended belief rule related to division domain $D(A_{i,j_i}; i=1,...,M)$ is written as follows:

$$\begin{aligned} R_{j_1\cdots j_M}: & \ IF\ U_1\ is\ \{(A_{1,j}, \alpha_{1,j}^{j_1\cdots j_M}); j=1,...,J_1\} \wedge \cdots \wedge U_M\ is\ \{(A_{M,j}, \alpha_{M,j}^{j_1\cdots j_M}); j=1,...,J_M\}, \\ & THEN\ D\ is\ \{(D_n, \beta_n^{j_1\cdots j_M}); n=1,...,N\},\ with\ \theta_{j_1\cdots j_M}\ and\ \{\delta_i; i=1,...,M\} \end{aligned} \tag{1}$$

where $\alpha_{i,j}^{j_1\cdots j_M}$ and $\beta_n^{j_1\cdots j_M}$ denote the belief degrees of referential value $A_{i,j}$ and consequent D_n in rule $R_{j_1\cdots j_M}$. Moreover, the belief degrees in antecedent attribute satisfy $\alpha_{i,j_i}^{j_1\cdots j_M} > \alpha_{i,j}^{j_1\cdots j_M}$ (j=1,…, J_i; $j \neq j_i$; i=1,…, M); $\theta_{j_1\cdots j_M}$ denotes the weight of rule $R_{j_1\cdots j_M}$; δ_i denotes the weight of attribute U_i.

In order to generate the extended belief rule shown in Eq. (1), Micro-EBRBS construction scheme should be performed to generate belief distributions, reduce extended belief rules, and calculate rule weights. Fig. 1 shows the basic framework of Micro-EBRBS construction scheme.

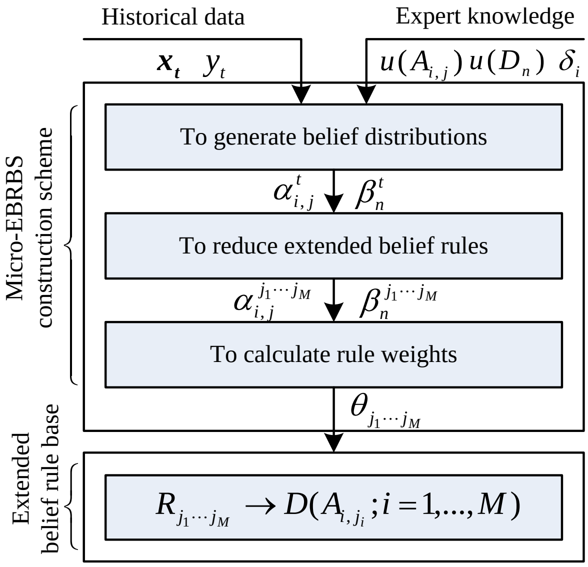

Fig. 1. Framework of Micro-EBRBS construction scheme

According to Fig. 1, the Micro-EBRBS construction scheme includes the following steps:

Step 1: To generate belief distributions. Suppose $x_{t,i}$ is the tth (t=1,…, T) input data of attribute U_i. A belief distribution $S(x_{t,i}) = \{(A_{i,j}, \alpha_{i,j}^t); j=1,...,J_i\}$ can be generated as follows:

$$\alpha_{i,j}^t = \frac{u(A_{i,j+1}) - x_{t,i}}{u(A_{i,j+1}) - u(A_{i,j})} \text{ and } \alpha_{i,j+1}^t = 1 - \alpha_{i,j}^t, \text{ if } u(A_{i,j}) \leq x_{t,i} \leq u(A_{i,j+1}) \tag{2}$$

$$\alpha_{i,k}^t = 0,\ for\ k=1,...,J_i\ and\ k \neq j, j+1 \tag{3}$$

where $u(A_{i,j})$ denotes the utility value of reference value $A_{i,j}$ in the ith attribute U_i; $\alpha_{i,j}^t$ denotes the belief degree of reference value $A_{i,j}$ generated from data $x_{t,i}$.

Next, when y_t is assumed to be the tth output data of attribute D and the utility values are $\{u(D_n); n=1,…, N\}$, a belief distribution $S(y_t) = \{(D_n, \beta_n^t); n = 1,..., N\}$ can be also generated using Eqs. (2) to (3).

Step 2: To reduce extended belief rules. All belief distributions generated from the *t*th input-output data pair < $x_{t,i}$, y_t, i=1,…, M> is regarded as an initial extended belief rule R_t (t=1,…, T). All these rules should be mapped into a division domain according to the following map function:

$$R_t \to D(A_{i,j_i}; i = 1,..., M);\ j_i = \arg\max_{j=1,...,J_i} \{\alpha_{i,j}^t\} \tag{4}$$

where the map function means the collection of the rules with the maximum belief degree in the same referential values.

Consequently, for the division domain which has one rule at least, all rules in the same division domain are used to generate a new extended belief rule, in which the belief degrees of new rule $R_{j_1 \cdots j_M}$ are calculated as follows:

$$\alpha_{i,j}^{j_1 \cdots j_M} = \frac{\sum_{t=1}^{T_{j_1 \cdots j_M}} \alpha_{i,j}^t}{T_{j_1 \cdots j_M}}; i = 1,..., M;\ j = 1,..., J_i \tag{5}$$

$$\beta_n^{j_1 \cdots j_M} = \frac{\sum_{t=1}^{T_{j_1 \cdots j_M}} \beta_n^t}{T_{j_1 \cdots j_M}}; n = 1,..., N \tag{6}$$

where $T_{j_1 \cdots j_M}$ is the number of extended belief rules in division domain $D(A_{i,j_i}; i = 1,..., M)$.

Step 3: To calculate rule weights. According to the analysis of rule weight calculation in EBRBS, whose rule weight will approximate to 1 when using a large number of data to generate extended belief rules [15], the rule weight of $R_{j_1 \cdots j_M}$ is calculated by

$$\theta_k = \frac{T_{j_1 \cdots j_M}}{T} \tag{7}$$

Remark 1: Comparing to the Micro-EBRBS proposed in [15] for classification problems, two adjustments are made for better solving regression problems, the first one is Eq. (6) which is based on $T_{j_1 \cdots j_M}$ to normalize belief degrees; the second one is Eq. (7), which considers the number of data in each division domain to calculate weight rules.

2.2. Micro-EBRBS inference scheme

After constructing a Micro-EBRBS, the resulting Micro-EBRBS can be used to reply any given input data according to the Micro-EBRBS inference scheme, whose basic framework is shown in Fig. 2.

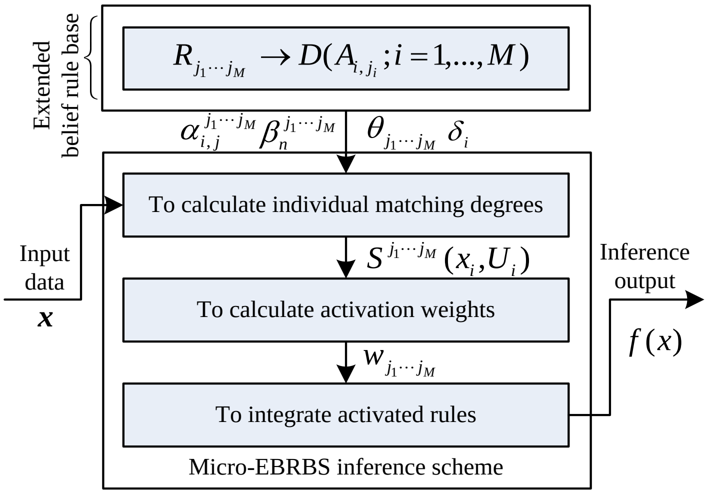

Fig. 2. Framework of Micro-EBRBS inference scheme

According to Fig. 2, the Micro-EBRBS inference scheme includes the following steps:

Step 1: To calculate individual matching degrees. For a given input data $\boldsymbol{x}$=(x_1,…, x_M), each input x_i (i=1,…, M) needs to be transformed into belief distribution $S(x_i)$={($A_{i,j}$, $\alpha_{i,j}$); j=1,…, J_i} via Eqs. (2) to (3). Thereafter, the individual matching degree $S^{j_1\cdots j_M}(x_i,U_i)$ between rule $R_{j_1\cdots j_M}$ and data $\boldsymbol{x}$ for attribute U_i is calculated based on the similarity measure of belief distributions as follows:

$$S^{j_1\cdots j_M}(x_i,U_i)=\begin{cases}1-d_i^{j_1\cdots j_M}, & d_i^{j_1\cdots j_M}\le 1\\ 0, & d_i^{j_1\cdots j_M}>1\end{cases}, d_i^{j_1\cdots j_M}=\sqrt{\sum_{j=1}^{J_i}(\alpha_{i,j}-\alpha_{i,j}^{j_1\cdots j_M})^2} \tag{8}$$

where $\alpha_{i,j}^{j_1\cdots j_M}$ is the belief degree of attribute U_i of rule $R_{j_1\cdots j_M}$; $d_i^{j_1\cdots j_M}$ denotes the Euclidean distance between the belief distributions of rule $R_{j_1\cdots j_M}$ and data $\boldsymbol{x}$ in attribute U_i.

Step 2: To calculate activation weights. Based on the individual matching degrees shown in Eq. (8), the activation weight of rule $R_{j_1\cdots j_M}$, denoted as $w_{j_1\cdots j_M}$, is calculated by

$$w_{j_1\cdots j_M}=\frac{\theta_{j_1\cdots j_M}\prod_{i=1}^{M}(S^{j_1\cdots j_M}(x_i,U_i))^{\bar{\delta}_i}}{\sum_{R_k}^{AR(\boldsymbol{x})}\theta_k\prod_{i=1}^{M}S^k(x_i,U_i)^{\bar{\delta}_i}}, \bar{\delta}_i=\frac{\delta_i}{\max_{t=1,\ldots,M}\{\delta_t\}} \tag{9}$$

where $\theta_{j_1\cdots j_M}$ is the weight of rule $R_{j_1\cdots j_M}$; δ_i is the weight of attribute U_i; $AR(\boldsymbol{x})$ is the rule set to produce an inference output for replying data $\boldsymbol{x}$ and it constrains all rules of Micro-EBRBS.

Step 3: To integrate activated rules. Suppose that all rules in $AR(\boldsymbol{x})$ are activated for replying data $\boldsymbol{x}$ and are further integrated using the analytical ER algorithm [16] as follows:

$$\beta_n=\frac{\prod_{R_{j_1\cdots j_M}}^{AR(\boldsymbol{x})}(w_{j_1\cdots j_M}\beta_n^{j_1\cdots j_M}+1-w_{j_1\cdots j_M})-\prod_{R_{j_1\cdots j_M}}^{AR(\boldsymbol{x})}(1-w_{j_1\cdots j_M})}{\sum_{i=1}^{N}\prod_{R_{j_1\cdots j_M}}^{AR(\boldsymbol{x})}(w_{j_1\cdots j_M}\beta_n^{j_1\cdots j_M}+1-w_{j_1\cdots j_M})-N\prod_{R_{j_1\cdots j_M}}^{AR(\boldsymbol{x})}(1-w_{j_1\cdots j_M})} \tag{10}$$

where β_n is the integrated belief degree of consequent D_n. Hence, when the utility value of consequent D_n is $u(D_n)$, the inference output of Micro-EBRBS is obtained as follows.

$$f(\boldsymbol{x})=\sum_{n=1}^{N}u(D_n)\beta_n+\frac{1-\sum_{n=1}^{N}\beta_n}{2}(u(D_1)+u(D_N)) \tag{11}$$

2.3. Challenges of Micro-EBRBS application

Micro-EBRBS is an extension of EBRBS and has been successfully applied in handling big data problems [15]. Owing to the fact that EBRBS is one of useful decision support systems and has shown its superior performance in the field of oil pipeline leak detection [11] and environmental governance cost prediction [14] , it is believed that Micro-EBRBS has the potential to be an advance decision support system for industrial cost prediction. However, according to Micro-EBRBS construction and inference schemes detailed in Sections 2.1 and 2.2, the following challenges of Micro- EBRBS must be overcome and solved for better applications.

Challenge 1: The influence of basic parameters on the Micro-EBRBS construction scheme

In the Micro-EBRBS construction scheme, the basic parameters, including attribute weights, utility values of antecedent and consequent attributes, are vital to generate an extended belief rule base. However, all these basic parameters are always initialized by domain experts based on personal experiences, since it difficult or even impossible for domain experts to provide right values for those basic parameters in a complex decision-making problem, original Micro-EBRBS usually fails to generate extended belief rule base with sufficiently high performance, Hence, to find a solution to determine the optimal value of basic parameters is necessary for the Micro-EBRBS construction scheme.

Challenge 2: The influence of rule activation on the Micro-EBRBS inference scheme

In the Micro-EBRBS inference scheme, one extended belief rule is activated if and only if its activation weight is greater than 0. Based on the formulas shown in Eqs (8) and (9), it can be found that the boundary value 1 has a direct relationship to rule activation, so it will lead to the incompleteness and inconsistency issues commonly found in data-driven systems, *i.e.*, incompleteness occurs in a Micro-EBRBS when the Euclidean distances shown in Eq. (8) for all rules are greater than 1; Conversely, inconsistency occurs when the Euclidean distances for all rules are smaller than 0. Hence, it is necessary to find a solution to determine the boundary value to calculate individual matching degrees.

The above challenges clearly show the necessary conditions for improving Micro-EBRBS for better applications. Thus, in the present work, two concepts, namely parameter optimization and activation factor, are introduced to improve Micro-EBRBS so that the improved Micro-EBRBS can overcome and address the two challenges mentioned earlier.

3. Improving Micro-EBRBS Using Parameter Optimization and Activation Factor

In this section, parameter optimization and activation factor are applied to improve the Micro-EBRBS construction and inference schemes firstly, followed by the introduction of the framework of improved Micro-EBRBS in Section 3.3.

3.1. Parameter optimization-based Micro-EBRBS construction scheme

In order to overcome the ***Challenge 1*** detailed in Section 2.3, a solution needs to be proposed to determine the optimal value of basic parameters for the Micro-EBRBS construction scheme. Inspired by the previous studies on improving rule-based systems [17][18][19], parameter optimization is introduced to optimize the basic parameters, including attribute weights, utility values of antecedent and consequent attributes, for Micro-EBRBS. Fig. 3 shows the framework of improved Micro- EBRBS construction scheme.

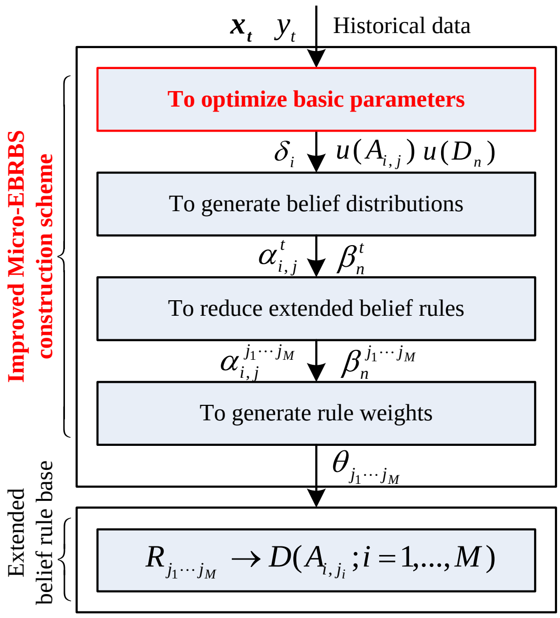

Fig. 3. Framework of improved Micro-EBRBS construction scheme

From Fig. 3, historical data need to be collected to optimally train the basic parameters of Micro-EBRBS. For the implementation of this purpose, a parameter optimization model is proposed to determine the optimal value of basic parameters. Suppose that Micro-EBRBS has M antecedent attribute with M attribute weights, and J_i (i=1,…, M) utility values $u(A_{i,j})$ (j=1,…, J_i)for the ith antecedent attribute and N utility values $u(D_n)$ (n=1,…, N) for one consequent attribute. Hence, the parameter optimization model of Micro-EBRBS can be written as follows:

$$\min MAE(\{\delta_i, u(A_{i,j}), u(D_n)\}) = \sum_{t=1}^{T} |y_t - f(\boldsymbol{x}_t)|$$

$$s.t.\begin{cases} 0 \le \delta_i \le 1; i=1,...,M \\ u(A_{i,j}) \le u(A_{i,j+1}); j=1,..., J_i-1; i=1,...,M \\ u(A_{i,1}) = lb_i; u(A_{i,J_i}) = ub_i; i=1,...,M \\ u(D_n) \le u(D_{n+1}); n=1,..., N-1 \\ u(D_1) = lb; u(D_N) = ub \end{cases} \tag{12}$$

where $MAE(\delta_i, u(A_{i,j}), u(D_n))$ denotes the mean absolute error (MAE) of the Micro-EBRBS composed by basic parameters δ_i, $u(A_{i,j})$, and $u(D_n)$; lb_i and ub_i are the lower and upper bounds of the ith antecedent attribute, respectively; lb and ub are the lower and upper bounds of consequent attribute, respectively; $\boldsymbol{x}_t$ and y_t are the tth (t=1,.., T) input and output; and $f(\boldsymbol{x}_t)$ is the inference output of the Micro-EBRB for replying input data $\boldsymbol{x}_t$.

Remark 5: For the above-mentioned parameter optimization model, it is worth noting that the previous studies on rule-based systems have proposed various kinds of parameter learning algorithms, such as the commonly used optimization toolbox in MATLAB [20], evolutionary algorithms [21][22], and expectation maximization algorithm [23][24]. All these algorithms can be used to solve the parameter optimization model shown in Eq. (12) to obtain the optimal parameter values of Micro-EBRBS.

3.2. Activation factor-based Micro-EBRBS inference scheme

In order to overcome the ***Challenge 2*** shown in Section 2.3, another solution needs to be proposed to activate consistent rules for the Micro-EBRBS inference scheme. Inspired by the previous studies on rule activation for EBRBS [25][26][27], activation factor is defined to revise the calculation of individual matching degrees, so that the Micro-EBRBS can activate consistent rules by avoiding inconsistency and incompleteness issues. Fig. 4 shows the framework of the improved Micro-EBRBS inference scheme.

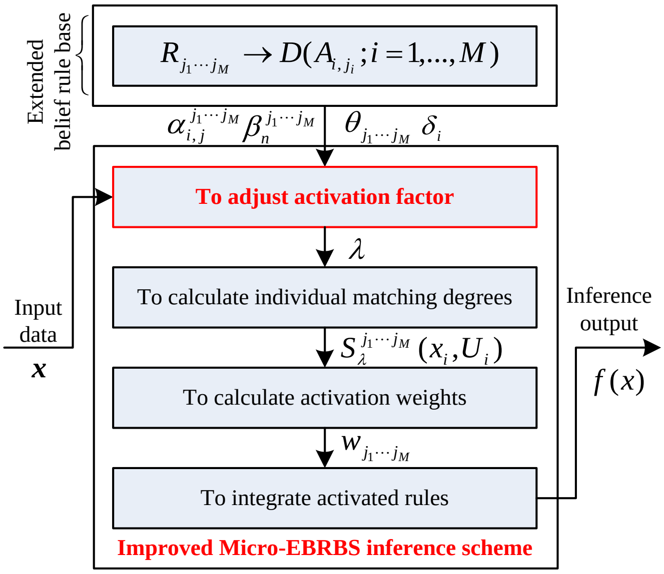

Fig. 4. Framework of improved Micro-EBRBS inference scheme

From Fig. 4, a new definition regarding activation factor-based individual matching degree is provided as follows:

Definition 1 (Activation factor-based individual matching degree): Suppose that the belief distribution of rule $R_{j_1\cdots j_M}$ and data $\boldsymbol{x}$ in attribute U_i is $\{(A_{i,j}, \alpha_{i,j}^{j_1\cdots j_M}); j=1,\dots, J_i\}$ and $\{(A_{i,j}, \alpha_{i,j}); j=1,\dots, J_i\}$, respectively. The new individual matching degree of rule $R_{j_1\cdots j_M}$ and data $\boldsymbol{x}$ for attribute U_i is calculated by

$$S_{\lambda}^{j_1\cdots j_M}(x_i,U_i)=\begin{cases}\lambda-d_i^{j_1\cdots j_M}, & d_i^{j_1\cdots j_M}\le\lambda\\ 0 & ,d_i^{j_1\cdots j_M}>\lambda\end{cases}, d_i^{j_1\cdots j_M}=\sqrt{\sum_{j=1}^{J_i}(\alpha_{i,j}-\alpha_{i,j}^{j_1\cdots j_M})^2} \tag{13}$$

where λ denotes activation factor and it has the following characteristics: 1) all rules will be activated when $\lambda=\sqrt{2}$; 2) none of rules will be activated when $\lambda=0$.

Based on ***Definition 1***, it can be found that the value of activation factor λ is vital to activate consistent rules for Micro-EBRB. Hence, the detailed step of adjusting the value of activation factor is provided as follows:

Step 1: To adjust activation factor. For a given input data $\boldsymbol{x}=(x_1,\ldots, x_M)$, the activation weight of rule $R_{j_1\cdots j_M}$, denoted as $w_{j_1\cdots j_M}$, can be calculated based on ***Step 1*** and ***Step 2*** detailed in Section 2.2, in which the individual matching degree is calculated according to ***Definition 1***. Afterwards, rule $R_{j_1\cdots j_M}$ should be put into rule set $\varDelta_\lambda$, namely $\varDelta_\lambda=\varDelta_\lambda\cup R_{j_1\cdots j_M}$ when $w_{j_1\cdots j_M}>0$. The consistency of $\varDelta_\lambda$ needs to be evaluated for determine the value of activation factor λ, in which the evaluation formula is shown as follows:

$$C(\varDelta_\lambda)=\frac{\max_{n=1,\ldots,N}\{C_n\}}{|\varDelta_\lambda|} \tag{14}$$

where C_n is calculated by

$$C_n=|D_n;n=\arg\max_{i=1,\ldots,N}\{\beta_{i,k}\};R_k\in\varDelta_\lambda| \tag{15}$$

Finally, by adjusting the value of λ to seek the maximum consistency of $\varDelta_\lambda$, and the corresponding rules can be regarded as activated rules.

3.3. Framework of improved Micro-EBRBS

According to the introduction of the original Micro-EBRBS in Section 2 and the improved construction and inference schemes in Section 3.1 and Section 3.2, respectively, an improved Micro-EBRBS is proposed in this section. Fig. 5 gives the framework of improved Micro-EBRBS.

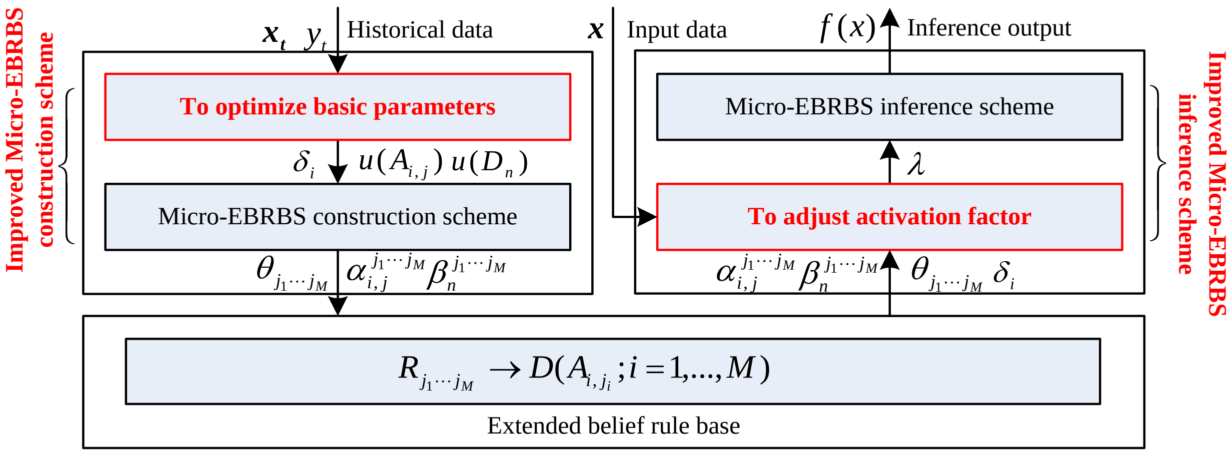

Fig.5. Framework of improved Micro-EBRBS

It is clear from Fig. 5 that an improved Micro-EBRBS consists of an extended belief rule base, an improved Micro-EBRBS construction scheme, which is based on the proposed parameter optimization (Please see Section 3.1 for details) to optimize basic parameters and further utilizes the Micro-EBRBS construction scheme (Please see Section 2.1 for details) to generate the extended belief rule base, and an improved Micro-EBRBS inference scheme, which is based on the proposed activation factor (Please see Section 3.2 for details) to revise the calculation of individual matching degrees and further utilizes the Micro-EBRBS inference scheme (Please see Section 2.2 for details) to produce inference outputs.

4. Industrial Cost Prediction Model based on Data Increment and Improved Micro-EBRBS

In this section, a DIME model is proposed on the basis of data increment transformation and the improved Micro-EBRBS for industrial cost prediction. The framework of the DIME model is introduced in Section 4.1 and its components are introduced in Sections 4.2 and 4.3, respectively.

4.1. Framework of DIME model for industrial cost prediction

According to the introduction of the improved Micro-EBRBS in Section 3, a new prediction model is developed for industrial cost prediction based on data increment transformation and the improved Micro-EBRBS and it is so called DIME model. Fig. 6 gives the framework of the DIME model.

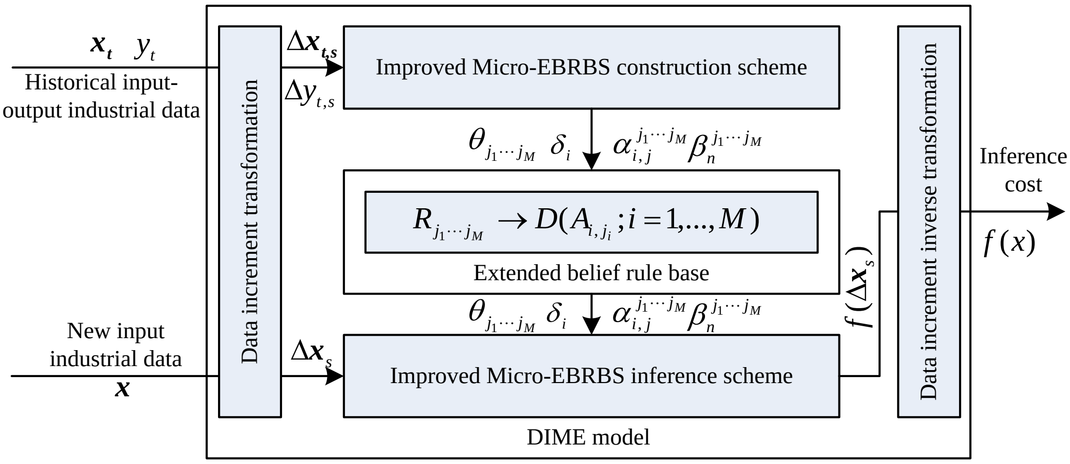

Fig. 6. Framework of DIME model for industrial cost prediction

It is clear from Fig. 6 that the DIME model has two main components, in which the first component is the construction of an extended belief rule base using historical input-output industrial data based on data increment transformation and the improved Micro-ERBBS construction scheme; the second component is the response of new input industrial data using the extended belief rule base based on the improved Micro-EBRBS inference scheme and data increment inverse transformation. The details of these two components are introduced in the following sections.

4.2. Using data increment to construct an improved Micro-EBRBS

For the industrial cost prediction, a limited number of industrial data is a serious but common problem which usually results in the over-fitting problem and the low accuracy of an industrial cost prediction model because of the lack of training data. Moreover, as the annotation process can be very expensive, it is very difficult to collect industrial data for the aim of industrial cost prediction in practical environment. For these reasons, a novel definition of data increment is introduced to enrich the limited industrial data as follows:

Definition 2 (Data increment) [28]. Suppose there has a function $y=f(x)$ with its definition domain $[a, b]$. When there exists an input-output data pair $<x_0, y_0>$ in the function $y=f(x)$, for any $x_1 \in [a, b]$, the data increment regarding the input and output data can be written as $\Delta x=x_1-x_0$ and $\Delta y=f(x_1)-f(x_0)=f(x_0+\Delta x)-f(x_0)$, respectively.

According to ***Definition 2*** and the parameter optimization-based Micro-EBRBS construction scheme shown in Section 4.1, the specific steps of constructing DIME model are as follows:

Step 1: To calculate the data increment of input-output industrial data pairs. Suppose that there are M input indicators and one output indicator with T input-output industrial data pairs $<\boldsymbol{x}_t, y_t>$ ($t=1,\ldots, T$) for industrial cost prediction, in which $\boldsymbol{x}_t=(x_{t,i}; i=1,\ldots, M)$. According to ***Definition 2***, the data increment of any two input-output industrial data pairs, *e.g.*, $<\boldsymbol{x}_t, y_t>$ and $<\boldsymbol{x}_t, y_t>$ ($t, s=1,\ldots, T; t\neq s$), can be calculated as follows:

$$\Delta \boldsymbol{x}_{t,s} = \boldsymbol{x}_t - \boldsymbol{x}_s \tag{16}$$

$$\Delta y_{t,s} = y_t - y_s \tag{17}$$

where the new T×(T-1) input-output data pairs regarding data increment can be obtained by selecting any two from original T input-output industrial data pairs, these T×(T-1) new input and output data shown in Eqs. (16) and (17) are denoted as $<\boldsymbol{x}_k, y_k>$ (k=1,…, L; L=T×(T-1)).

Step 2: To construct an improved Micro-EBRBS. Firstly, L new input and output data should be used to optimize the basic parameters of Micro-EBRBS via Eq. (12) and they can be denoted as δ_i, $u(A_{i,j})$, and $u(D_n)$; Secondly, L sets of belief distributions should be transformed from L new input and output data using Eqs. (2) and (3); Thirdly, based on the location relationship between extended belief rules shown in Eq. (4), extended belief rules can be generated from L sets of belief distributions using Eqs. (5) and (6); Finally, the rule weight of each extended belief rule is calculated based on Eq. (7).

4.3. Using the improved Micro-EBRBS to predict industrial costs

In this section, the improved Micro-EBRBS based on the data increment of industrial input-output data pairs is applied to predict industrial costs. The detailed steps of cost prediction process are provided as follows:

Step 1: To calculate the data increment of new input data. Suppose that a new input data of industrial cost prediction is $\boldsymbol{x}=(x_1,\ldots, x_M)$. A candidate set of historical input-output industrial data pairs $CS(\boldsymbol{x})$ should be selected from historical T data pairs $<\boldsymbol{x}_t, y_t>$ (t=1,…, T) as follows:

$$CS(\boldsymbol{x}) = CS(\boldsymbol{x}) \cup \{(\boldsymbol{x}_s, y_s)\}, if\ |x_i - x_{s,i}| < |x_i - x_{t,i}|; i = 1,\ldots, M; t = 1,\ldots, T \tag{18}$$

where $(\boldsymbol{x}_s, y_s)$ is the data increment of any two input-output industrial data pairs in industrial cost prediction. Next, the data increment of new input data $\boldsymbol{x}$ can be calculated as follows:

$$\Delta \boldsymbol{x}_s = \boldsymbol{x} - \boldsymbol{x}_s; (\boldsymbol{x}_s, y_s) \in CS(\boldsymbol{x}) \tag{19}$$

Step 2: To produce inference outputs for replying input data. For each data increment of new input data $\boldsymbol{x}$ shown in Eq. (19), the activation factor should be calculated firstly based on Eqs. (14) to (15), followed by the calculation of individual matching degrees using ***Definition 1***. Afterwards, the calculation of activation weights and the integration of activated rules are performed to produce an inference output $f(\Delta\boldsymbol{x}_s)$ for replying data increment $\Delta\boldsymbol{x}_s$ according to Eqs. (9) to (11). Finally, the final inference output to reply input data $\boldsymbol{x}$ can be obtained by:

$$f(\boldsymbol{x}) = \frac{\sum_{(x_s, y_s) \in CS(x)} (f(\Delta \boldsymbol{x}_s) + y_s)}{|CS(\boldsymbol{x})|} \tag{20}$$

5. Case Study of Industrial Cost Prediction in China

To verify the effectiveness of the DIME model, the historical data of 13 state-own holding industries are collected from Chinese industrial Yearbooks 1999-2019 to analyze the industrial cost prediction process and compare some existing cost prediction models.

5.1. Data sources and variable definition

Considering the different types of state-own holding industries in China, the input and output indicators of 13 Chinese state-own holding industries are selected as research objects for industrial cost prediction. Table 1 shows the classification of these 13 state-own holding industries. Based on the input and output data collected from Chinese industrial Yearbooks,

Table 2 shows the data statistic analysis of two input indicators and three output indicators from 1999-2019.

Table 1 Classification of state-own holding industries

Categories	Detailed information of 13 industries
Mining industries	1. Nonferrous metal mining and dressing; 2. Mining professional and auxiliary activities;
Manufacturing industries	1. Agricultural and sideline food processing; 2. Textile; 3. Clothing; 4. Leather, fur and feather products; 5. Wood processing and wood products; 6. Furniture manufacturing; 7. Paper products; 8. Printing and recording media reproduction; 9. Culture and education products; 10. Pharmaceutical manufacturing; 11. Electrical machine and equipment manufacturing;

Table 2 Statistic analysis of input and output indicators

	Indicator name	Minimum	Maximum	Average	Standard deviation
Input indicators	Operating income(10^8yuan)	-32.33	3353.20	97.42	337.35
	Total profit(10^8yuan)	1.24	4148.17	574.75	693.91
Output indicators	Total Assets(10^8yuan)	28.39	5891.74	829.64	821.76
	Total liabilities(10^8yuan)	4.84	1761.00	372.43	401.58
	Selling expenses(10^8yuan)	18.18	3662.59	577.70	679.67

From the statistic analysis results in Table 2, the industrial output indicators include total assets, total liabilities, and selling expenses; the input indicators include industrial operating income and total profit, it is obvious that there have large differences in input and output indicators among the 13 state-own holding industries, and the standard deviation of input and output indicators is also large, indicating that the production costs and the performance of different state-own holding industries are significantly different. Meanwhile, the data from 1999 to 2018 of each state-own holding industry in China is used as training data and the remaining data as testing data in constructing the DIME model.

5.2. Development procedure of the DIME model for industrial cost prediction

In this section, the procedures of developing a DIME model are analyzed via the data increment transformation and parameter optimization, the improved Micro-EBRBS construction, and industrial cost prediction in Sections 5.2.1 to 5.2.3.

5.2.1. The 1st part: data increment transformation and parameter optimization

In this section, according to data increment transformation and parameter optimization, the corresponding development and results of the DIME model are provided, which takes the example of clothing industry by using operating income and total profit as two input indicators and using selling expenses as an output indicator.

Firstly, based on the steps in Section 4.2, the data increment of input and output indictors can be transformed and the statistic analysis of these data increments are showed in Table 3. From Table 3, it can be found that 1) the absolute value of minimum value is equal to that of maximum value; 2) the average value is equal to 0 for all indicators. This is because the data increment transformation is based on any two input-output data pairs to generate new data so that there always exists $\Delta \boldsymbol{x}_{t,s}=-\Delta \boldsymbol{x}_{s,t}$. For example, there are two input-output data pairs x_t=10 and x_s=5, the resulting data increments are $\Delta x_{t,s}$=5 and $\Delta x_{s,t}$=-5, respectively. Therefore, the absolute values of minimum and maximum values are equal to each other for all indicators and the average values are all equal to 0.

Table 3 Statistic analysis of data increment of clothing industry

Indicator name	Relationship	Minimum	Maximum	Average	Standard deviation
Operating income(10^8yuan)	Input indicator	-144.59	144.59	0.00	115.99
Total profit(10^8yuan)	Input indicator	-1835.31	1835.31	0.00	142.10
Selling expenses(10^8yuan)	Output indicator	-1513.69	1513.69	0.00	150.63

Secondly, to construct an extended belief rule base, a set of referential values and consequents should be given using expert knowledge and they are assumed as {*Negative High* (*NH*), *Negative Low* (*NL*), *Zero* (*Z*), *Positive Low* (*PL*), *Positive High* (*PH*)}. Based on the minimum and maximum values of each indicator shown in Table 3, Table 4 provides the utility values for each referential value and consequent, which are evenly distributed within the minimum and maximum value of each indicator. For example, five utility values of the selling expenses are {(*NH* -1513.69), (*NL*, -756.85), (*Z*, 0.00), (*PL*, 756.85), (*PH*, 1513.69)}. Additionally, the weight of each input indicator should be also provided by expert knowledge and they are set as 1.000 for operating income and total profit.

Table 4. Initial basic parameter values of selling expenses prediction of clothing industry

Indicator name	Relationship	Weight	*NH*	*NL*	*Z*	*PL*	*PH*
Operating income(10^8yuan)	Input indicator	1.000	-144.59	-72.30	0.00	72.30	144.59
Total profit(10^8yuan)	Input indicator	1.000	-1835.31	-917.65	0.00	917.65	1835.31
Selling expenses(10^8yuan)	Output indicator	--	-1513.69	-756.85	0.00	756.85	1513.69

Thirdly, the parameter optimization model shown in Eq. (12) with the existing differential evolution (DE) algorithm [17] is used to iteratively optimize the basic parameters based on the data increment of operating income, total profit, and selling expenses. When the number of iterations and individuals in the DE algorithm is set as 300 and 60, respectively, the change of the absolute mean error (MAE) obtained from the target function is shown in Fig. 7. It can be found from Fig. 7 that the MAE of selling expenses prediction is significantly decreased after 300 iterations and gradually tends to converge.

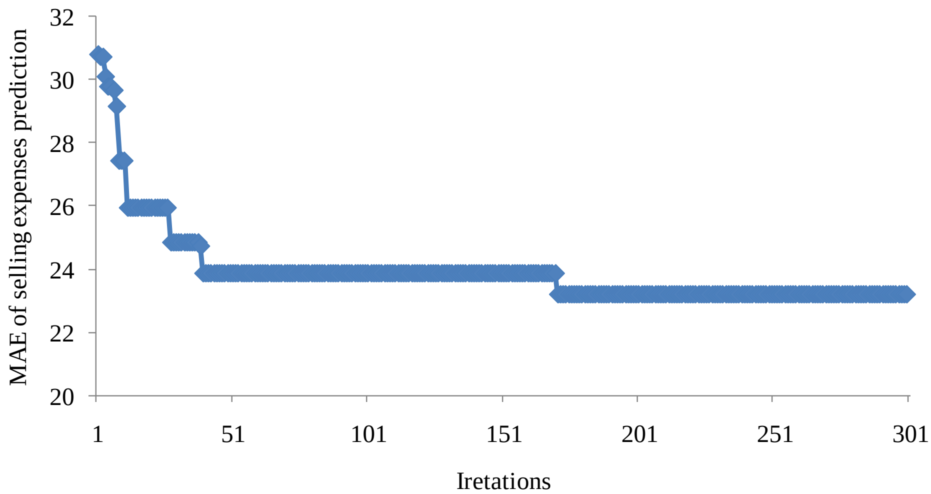

Fig. 7 MAE of Micro-EBRBS in predicting selling expense for clothing industry

Finally, after performing parameter optimization, the optimized basic parameters for Micro-EBRBS in predicting selling expenses for clothing industry are shown in Table 5. It can be found from Table 5 that the weights and utility values of input and output indicators have some changes after iterative parameter optimization. For example, the initial weights of two input indicators are 1, but both of them are 0.2584 and 0.6131 after iterative parameter optimization. Five utility values of selling expenses are {(*NH*, -1513.69), (*NL*, -242.58), (*Z*, 283.18), (*PL*, 966.67), (*PH*, 1513.69)}, which are different from the initial utility values {(*NH* -1513.69), (*NL*, -756.85), (*Z*, 0.00), (*PL*, 756.85), (*PH*, 1513.69)}.

Table 5 Optimal parameter values of selling expenses prediction of clothing industry

Indicator name	Relationship	Weight	*NH*	*NL*	*Z*	*PL*	*PH*
Operating income	Input indicator	0.258	-144.59	-56.75	23.33	73.83	144.59
Total profit	Input indicator	0.613	-1835.31	-745.73	-67.42	188.91	1835.31
Selling expenses	Output indicator	--	-1513.69	-242.58	283.18	966.67	1513.69

5.2.2. The 2nd part: the improved Micro-EBRBS construction

Continuing with the example of the selling expenses prediction for clothing industry, the improved Micro-EBRBS construction scheme is further used to generate extended belief rules and calculate rule weights in this section.

Firstly, according to the utility values of all indicators shown in Table 5, a total of 380 input-output industrial data increment are used to generate belief distributions according to Eqs. (2) and (3). For example, one input-output industrial data increment is $x_{t,1}$=10.77, $x_{t,2}$=65.91, and y_t=25.52, the corresponding belief distributions for each indicators are $S(x_{t,1})$= {(*NL*, 0.157), (*Z*, 0.843)}, $S(x_{t,2})$={(*Z*, 0.52), (*PL*, 0.48)}, and $S(y_t)$= {(*NL*, 0.49), (*Z*, 0.51)}, respectively. Consequently, the division domain of these belief distributions is $D(A_{1,3}, A_{2,3})$ or $D(Z, Z)$.

Table 6 Rule weights and belief distribution of selling expenses prediction of clothing industry

$R_{j_1\cdots j_M}$	$\theta_{j_1\cdots j_M}$	Belief distributions		
		Operating income	Total profit	Selling expenses
$R_{3,3}$	0.442	{(*NL*, 0.322), (*Z*, 0.678)}	{(*NL*, 0.004), (*Z*, 0.763), (*PL*, 0.233)}	{(*NL*, 0.577), (*Z*, 0.423)}
$R_{1,4}$	0.084	{(*NH*, 0.919), (*NL*, 0.081)}	{(*Z*, 0.035), (*PL*, 0.951), (*PH*, 0.014)}	{(*NL*, 0.654), (*Z*, 0.346)}
$R_{3,4}$	0.197	{(*NL*, 0.220), (*Z*, 0.780)}	{(*Z*, 0.307), (*PL*, 0.693)}	{(*NL*, 0.451), (*Z*, 0.549)}
$R_{2,1}$	0.084	{(*NL*, 0.850), (*Z*, 0.141)}	{(*NH*, 0.843), (*NL*, 0.157)}	{(*NH*, 0.943), (*NL*, 0.057)}
$R_{5,3}$	0.084	{(*PL*, 0.101), (*PH*, 0.899)}	{(*NL*, 0.101), (*Z*, 0.899)}	{(*NL*, 0.424), (*Z*, 0.576)}
$R_{4,1}$	0.011	{(*PL*, 0.754), (*PH*, 0.246)}	{(*NH*, 0.967), (*NL*, 0.033)}	{(*NH*, 0.895), (*NL*, 0.105)}
$R_{4,5}$	0.042	{(*Z*, 0.348), (*PL*, 0.652)}	{(*PL*, 0.080), (*PH*, 0.920)}	{(*PL*, 0.081), (*PH*, 0.919)}
$R_{2,5}$	0.008	{(*NH*, 0.350), (*NL*, 0.650)}	{(*PL*, 0.014), (*PH*, 0.986)}	{(*PL*, 0.226), (*PH*, 0.774)}
$R_{3,5}$	0.042	{(*Z*, 0.744), (*PL*, 0.256)}	{(*PL*, 0.120), (*PH*, 0.880)}	{(*PL*, 0.185), (*PH*, 0.815)}
$R_{1,5}$	0.003	{(*NH*, 0.520), (*NL*, 0.480)}	{(*PL*, 0.040), (*PH*, 0.960)}	{(*PL*, 0.297), (*PH*, 0.703)}
$R_{2,3}$	0.003	{(*NL*, 0.541), (*Z*, 0.459)}	{(*NL*, 0.001), (*Z*, 0.999)}	{(*NL*, 0.647), (*Z*, 0.353)}

Secondly, after transforming 380 industrial data increment into belief distributions, the all rules located in the same division domain are used to generate a new extended belief rule according to Eqs. (5) and (6). For example, apart from the rule consisted of {(*NL*, 0.157), (*Z*, 0.843)} for operating income, {(*Z*, 0.52), (*PL*, 0.48)} for total profit, and {(*NL*, 0.49), (*Z*, 0.51)} for selling expenses, there is another rule consisted of {(*Z*, 0.843), (*PL*, 0.157)} for operating income, {(*NL*, 0.48), (*Z*, 0.52)} for total profit, and {(*Z*, 0.51), (*PL*, 0.49)} for selling expenses at the division domain $D(Z, Z)$, the resulting new extended belief rule can be calculated and its belief distributions is {(*NL*, 0.0785), (*Z*, 0.843), (*NL*, 0.0785)} in operating income, {(*NL*, 0.24), (*Z*, 0.52), (*PL*, 0.24)} for total profit, and {(*NL*, 0.245), (*Z*, 0.51), (*PL*, 0.245)} for selling expenses. Additionally, due to the fact that there are only two rules at the division domain $D(Z, Z)$, the weight of the new rule is 2/380=0.053. Table 6 shows the weight and belief distribution of all new extended belief rules.

5.2.3. The 3rd part: cost prediction using the improved Micro-EBRBS and data increment

In this section, according to the inference of the improved Micro-EBRBS with data increment inverse transformation, the corresponding development and results are provided as follows:

Firstly, by taking the example of clothing industry using operating income and total profit as two input indicators and using selling expenses as an output indicator, the new input industrial data is $\boldsymbol{x}$=(5.2, 210.90), the candidate set of historical input-output industrial data pairs can be obtained, namely $CS(\boldsymbol{x})$={<$\boldsymbol{x}_t$=(10.41, 207.64), y_t=149.15>, <$\boldsymbol{x}_s$=(11.02, 187.65), y_s=142.59>}. Hence, the input of the improved Micro-EBRBS are $\Delta\boldsymbol{x}_t$=(-5.21, 3.26) and $\Delta\boldsymbol{x}_s$=(-5.82, 13.25).

Secondly, on the basis of the extended belief rules shown in Table 6, the improved Micro-EBRBS inference scheme is performed to produce the inference output of $\Delta\boldsymbol{x}_t$=(-5.21, 3.26) and $\Delta\boldsymbol{x}_s$=(-5.82, 13.25). Table 7 shows the activation factors, the integrated belief distributions, and the predicted cost increments of the improved Micro-EBRBS.

Table 7 Activation factor and integrated belief distribution of improved Micro-EBRBS

	Activation factor	Integrated belief distribution					Predicted cost increment
		NH	*NL*	*Z*	*PL*	*PH*	
Selling expenses	0.7242	0.0000	0.5650	0.4350	0.0000	0.0000	-13.87
	0.7242	0.0000	0.5440	0.4560	0.0000	0.0000	-2.86

Finally, according to the predicted cost increments in Table 7, the actual costs in $CS(\boldsymbol{x})$, and Eq. (20), the final predicted cost of the data $\boldsymbol{x}$=(5.2, 210.90) can be calculated by $f(\boldsymbol{x})=(y_t+f(\Delta\boldsymbol{x}_t)+y_s+f(\Delta\boldsymbol{x}_s))/2$=(149.15+(-13.87)+142.59+(-2.86))/2= 137.505. Similarly, by using the improved Micro-EBRBS to predict three types of costs in 13 state-own holding industries of China, Figs. 8 to 10 show the fitting degree between predicted costs and actual costs of 13 state-own holding industries.

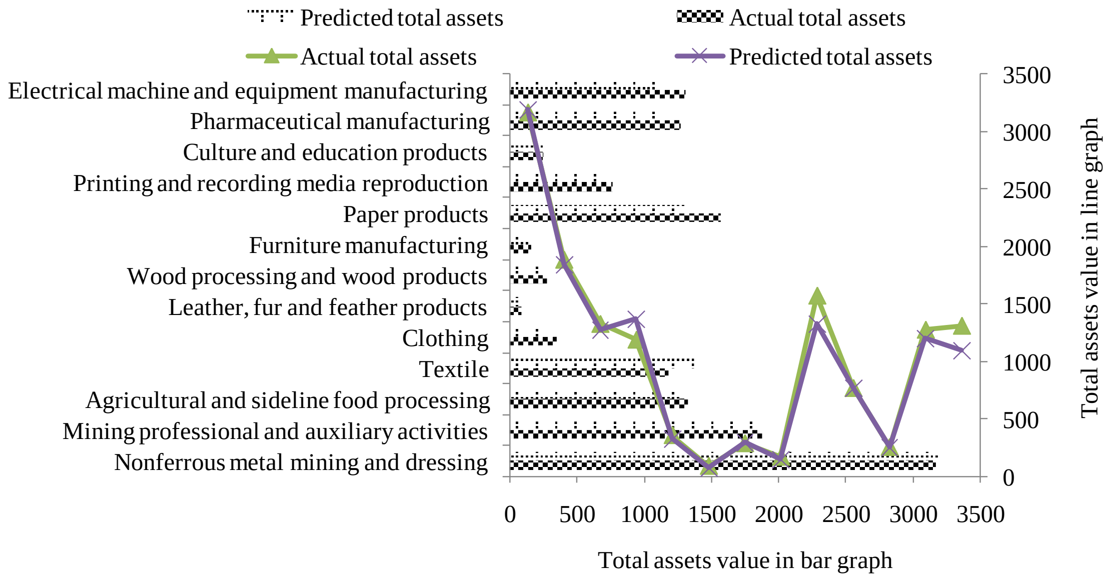

Fig. 8. Fitting degree between actual total assets and predicted total assets

From Fig. 8, it can be seen that the predicted value of total assets basically fits the actual value, and the culture and education products industry and printing and recording media reproduction industry have the highest fitting degree. From the perspective of total assets difference, it indicates that nonferrous metal mining and dressing industry has the highest assets, while leather, fur and feather products industry has the lowest assets. The prediction results between actual total assets and predicted assets show that the improved Micro-EBRB has strong applicability and good prediction performance in different industrial cost prediction. The difference in the total assets of different state-owned holding industries is related to the national industries development policy and the focus of economic development, and it is also determined by the differences in the historical development factors and internal management modes of state-owned industries.

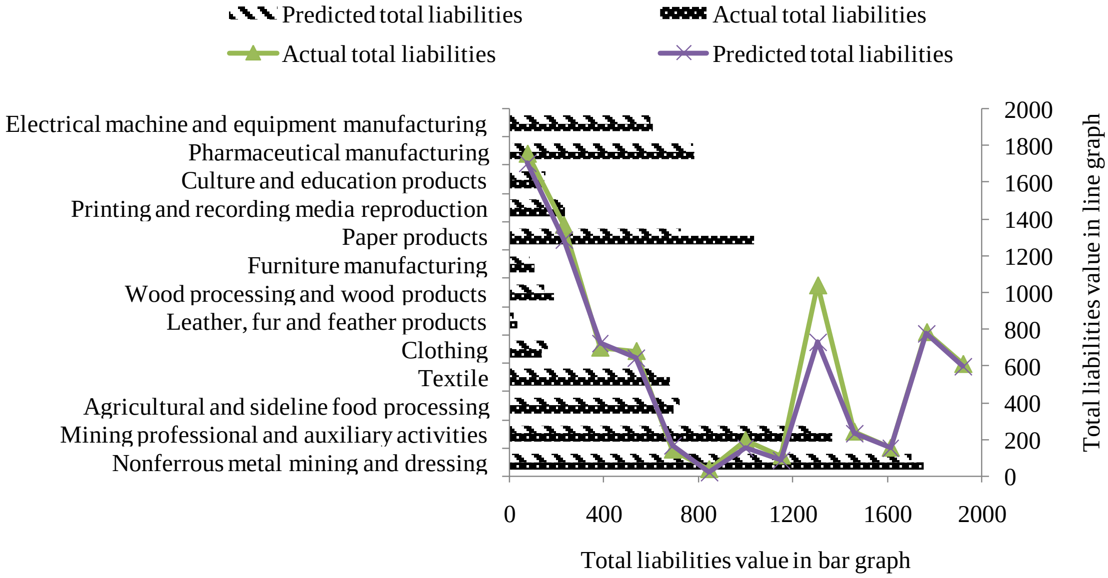

Fig.9. Fitting degree between actual total liabilities and predicted total liabilities

From Fig. 9, it worth noting that the total liabilities in nonferrous metal mining and dressing industry has higher than other industries, and paper products industry has the lowest fitting degree between actual total liabilities and predicted total liabilities, however, the predicted value of the total liabilities of most industries in Fig. 9 is highly consistent with the actual value. The total liabilities of state-owned holding industries are closely related to the internal management mode and market operation status of industries, but it does not mean that the more total liabilities, the worse the operation status of state-own holding industries. This is because the production and operation scale of different industries varies greatly, that is, although the total liabilities of some industries are large, the total assets and selling expenses of these industries are also higher.

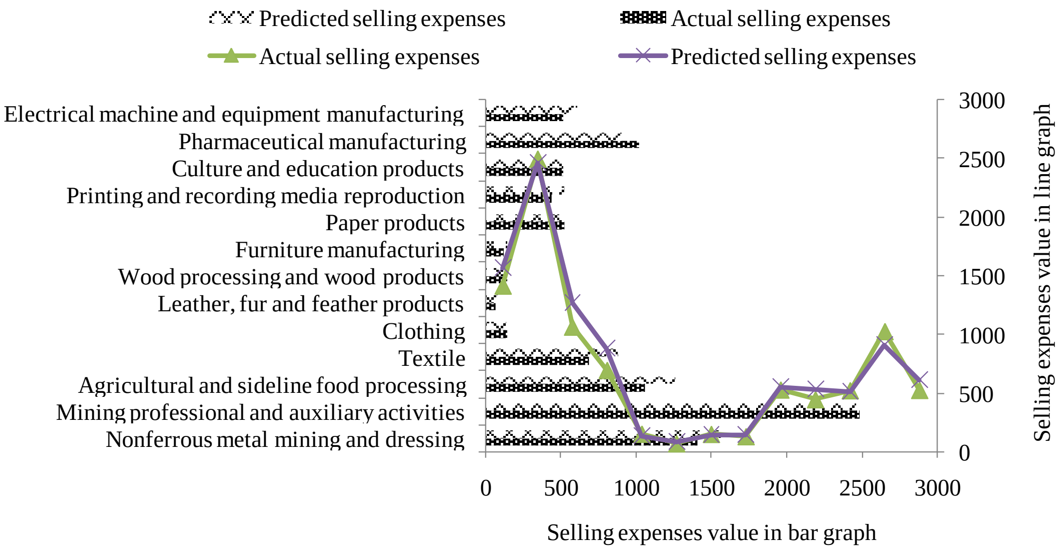

Fig. 10. Fitting degree between actual selling expenses and predicted selling expense

Fig. 10 shows that the predicted value of the selling expenses of 13 state-owned holding industries basically fits the actual value. Among them, the Mining professional and auxiliary activities industry has the highest selling expenses, while the leather, fur and feather products industry and furniture manufacturing industry have significantly lower selling expenses than other industries, this difference is related to the needs of current Chinese economic development and market policies.

In summary, the results within this section demonstrate that three kinds of cost prediction based on the DIME model have a high fitting degree with the actual industry costs, which shows that the effectiveness of using the improved Micro-EBRBS and data increment to propose the DIME model for better industrial cost prediction.

5.3. Comparative analysis of different models for industrial cost prediction

In order to validate the effectiveness of the DIME model, the result of the predicted costs is measured with MAE. Table 8 shows the comparison results of the Micro-EBRBS with or without parameter optimization. As Table 8 illustrates, the parameter optimization can effectively improve the accuracy of Micro-EBRBS, so the MAE between actual cost and predicted cost are 91.06, 57.07 and 90.14 by Micro-EBRBS without parameter optimization, while MAE between actual cost and predicted cost are 70.07, 49.03 and 72.49 by Micro-EBRBS with parameter optimization. Based on the parameter optimization, the prediction error of three industrial costs predicted by Micro-EBRBS is significantly decreased.

Table 8. Comparison of Micro-EBRBS with or without parameter optimization

Predicted costs	MAE	
	Without parameter optimization	With parameter optimization
Total Assets	91.06	70.07
Total liabilities	57.07	49.03
Selling expenses	90.14	72.49

Additionally, the DIME model is further used to compare with some existing prediction models, including adaptive neuro fuzzy inference system (ANFIS), EBRBS, and original Micro-EBRBS. Table 8 shows the MAE of these models for the industrial cost prediction of total assets, total liabilities, and selling expenses. From Table 8, it can be found that the DIME model obtains the 1st best accuracy in the three types of industrial cost prediction, while the original Micro-EBRBS obtains the 2nd best accuracy in total assets and total liabilities prediction, and 3rd best accuracy in selling expenses prediction. Additionally, in the comparison of average ranks, the result is DIME (1) > Micro-EBRBS (2.33) > EBRBS (3.33) > ANFIS (3.66), indicating that the DIME model can produce satisfactory results comparing to other models.

Table 9. Comparison of the DIME model with other models

Predicted costs	ANFIS	EBRBS	Micro-EBRBS	DIME
Total Assets	149.96(3)	178.26(4)	91.06(2)	70.07(1)
Total liabilities	124.95(4)	152.01(4)	57.07(2)	49.03(1)
Selling expenses	173.27(4)	80.79(2)	90.14(3)	72.49(1)

6. Conclusions

In this study, a novel industrial cost prediction model, named DIME model, was developed based on the data increment transformation and the Micro-EBRBS, in which the former one is a data preprocessing technique that used to enrich data for small data analytics; the latter one is an advanced rule-based system for handling big data problems. The case study of 13 state-own holding industries in China demonstrated that the DIME model is an effective and accurate industrial cost prediction model. The conclusions of this study can be further summarized as follows:

(1) By considering the limited data available for industrial cost prediction, the data increment transformation was used as the first stage of constructing industrial cost prediction model. Due to the large scale of the training data obtained from the data increment transformation, the Micro-EBRBS was used as the second stage of constructing industrial cost prediction model. As a result, the DIME model was proposed for the first time to predict industrial costs.

(2) Due to the artificial subjectivity in parameters determination, the parameter optimization was introduced as the first improvement to enhance the prediction accuracy of Micro-EBRBS. Moreover, the activation factor was used as the second improvement to revise the calculation of individual matching degrees, so Micro-EBRBS can avoid the incompleteness and inconsistency issues in the process of producing an inferential output for the given input data.

(3) The comparative studies demonstrated that the DIME model could effectively improve the accuracy of Micro-EBRBS by using data increment transformation, parameter optimization, and activation factor, the predicted costs of 13 state-own holding industries have high fitting degree with actual industrial costs, and the prediction errors obtained by the DIME model are much lower than the other cost prediction models.

The present study has several limitations. The first limitation is related to the number of state-owned holding industries and input indicators where the 13 industries and the 2 input indicators were selected in the case study. More industries and input indicators should be used in the modeling process. The second limitation is related to indicator screening which was carried out according to literature review and expert experience. The objective feature selection method should be used to select the representative indicators in the modeling process.

In the future researches, more types of industries can be applied to industrial cost prediction with the consideration of more indicators and data, as well as some famous features selection methods. Owing to different characteristics of industrial management, the influences of technological innovation and regional policies on industrial cost prediction can be analyzed in future studies. Furthermore, an offline method can be studies to determine the value of activation factor for the Micro-EBRBS, which would promote the application of the Micro-EBRBS for various complex prediction problems.

Acknowledgements

This research was supported by the National Natural Science Foundation of China (Nos. 72001043 and 61773123), the National Science Foundation of Fujian Province, China (No. 2020J05122), the Humanities and Social Science Foundation of the Ministry of Education of China (Nos. 20YJC630188).

Data Availability Statement

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Reference:

[1] China National Bureau of Statistics, http://www.stats.gov.cn/tjsj/ndsj/.

[2] InHaeng, N.J., Amit S., Anna, S. M. Exploring employees' perceptions of costs and benefits of unionization in the hospitality industry. International Journal of Hospitality Management, 2020,87,102465.

[3] Joanna, P. Cost efficiency and endogenous regulatory choices: evidence from the transport industry in France. Journal of Regulatory Economics, 2021, 59, 25-46.

[4] Lin, B. Q., Chen, Y. The rapid development of the photovoltaic industry in China and related carbon dioxide abatement costs. Regional Environmental Change, 2020, 20, 49-59.

[5] Lyu, Y. Z., Ye, H.Y., Zhao, Z. N., Tian, J. P., Chen, L. J. Exploring the cost of wastewater treatment in a chemical industrial Park: Model development and application. Resources, Conservation & Recycling, 2020, 155, 104663.

[6] Tohry, A., Chelgani C. S., Matin S. S., Noormohammadi, M. Power-draw prediction by random forest based on operating parameters for an industrial ball mill. Advanced Powder Technology, 2020, 31, 967-972.

[7] Guo, H., Deng, S. X., Yang J. B., Liu, J. W., Nie, C. D. Analysis and prediction of industrial energy conservation in underdeveloped regions of China using a data pre-processing grey model. Energy policy, 2020,139,111244.

[8] Ma, C., Liu Z. B., Cao, Z.G., Song, W., Zhang, J., Zeng, W. L. Cost-sensitive deep forest for price prediction. Pattern Recognition, 2020,

107,107499.

[9] Chakraborty, D., Elhegazy, H., Elzarka,H., Gutierrez,L. A novel construction cost prediction model using hybrid natural and light gradient boosting. Advanced Engineering Informatics,2020,46,101201.

[10] Jiang, Z. G., Ding, Z. Y., Liu, Y., Wang, Y., Hu, X. L., Yang, Y. H. A data-driven based decomposition–integration method for remanufacturing cost prediction of end-of-life products. Robotics and Computer Integrated Manufacturing, 2020, 61, 101838.

[11] Liu J., Martínez L., Calzada A., Wang H., A novel belief rule base representation, generation and its inference methodology, Knowledge-Based Systems, 2013, 53: 129-141.

[12] Yang L. H., Wang Y. M., Chang L. L., Fu Y. G. A disjunctive belief rule-based expert system for bridge risk assessment with dynamic parameter optimization model. *Computers & Industrial Engineering*, 2017, 113, 459-474.

[13] Yang, Y., Fu, C., Chen, Y. W., Xu, D. L., Yang, S. L. A belief rule based expert system for predicting consumer preference in new product development.Knowledge-Based Systems,2016, 94,105-113.

[14] Wang Y. M., Ye F. F., Yang L. H. Extended belief rule based system with joint learning for environmental governance cost prediction. Ecological Indicators, 2020, 111, 1-14.

[15] Yang L.H., Liu J., Wang Y.-M., Martinez L., A Micro-Extended Belief Rule-Based System for Big Data Multiclass Classification Problems, IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2021, 51(1): 420-440.

[16] Wang, Y. M., Yang, J. B., Xu, D. L. Environmental impact assessment using the evidential reasoning approach. European Journal of Operation Research, 2006,174(3):1885-1913.

[17] Yang L. H., Liu J., Wang Y. M., Martínez L., New activation weight calculation and parameter optimization for extended belief rule-based system based on sensitivity analysis, Knowledge and Information Systems, 2019, 60(2): 837-878.

[18] Chang L. L., Zhang L. M., Xu X. J., Correlation-oriented Complex System Structural Risk Assessment using Copula and Belief Rule Base, Information Sciences, 2021, In Press.

[19] Chang L. L., Fu C., Zhu W., Liu W. Y., Belief rule mining using the evidential reasoning rule for medical diagnosis. International Journal of Approximate Reasoning, 2021, 130: 273-291.

[20] Sachan S., Yang J. B., Xu D. L., Benavides D. E., Li Y., An explainable AI decision-support-system to automate loan underwriting, Expert Systems with Applications, 2020, 144: 1-24.

[21] Xu X. J., Yan X. P., Sheng C. X., Yuan C. Q., Xu D. L., Yang J. B., A Belief Rule-Based Expert System for Fault Diagnosis of Marine Diesel Engines, IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2020, 50(2): 656-672.

[22] Chang L. L., Zhou Z. J., Liao H. C., Chen Y. W., Tan X., Herrera F., Generic Disjunctive Belief Rule Base Modeling, Inferencing, and Optimization, IEEE Transactions On Fuzzy Systems, 2019, 27(9): 1866-1880.

[23] Zhang B. C., Zhang Y., Hu G. Y., Zhou Z. J., Wu L. H., Lv S. Y., A method of automatically generating initial parameters for large-scale belief rule base, Knowledge-Based Systems, 2020, 199: 1-12.

[24] Zhou Z. J., Feng Z. C., Hu C. H., Hu G. Y., He W., Han X.-X., Aeronautical relay health state assessment model based on belief rule base with attribute reliability, Knowledge-Based Systems, 2020, 197: 1-13.

[25] Calzada A., Liu J., Wang H., Kashyap A., A New Dynamic Rule Activation Method for Extended Belief Rule-Based Systems, IEEE Transactions On Knowledge and Data Engineering, 2015, 27(4): 880-894.

[26] Zhu H. Z., Xiao M. Q., Yang L. H., Tang X. L., Liang Y. J., Li J. F., A minimum centre distance rule activation method for extended belief rule-based classification systems, Applied Soft Computing, 2020, 91: 1-14.

[27] Chang L. L., Zhou Z. J., You Y., Yang L.-H., Zhou Z.-G., Belief rule based expert system for classification problems with new rule activation and weight calculation procedures. Information Sciences, 2016, 336: 75-91.

[28] Yang L. H., Wang S., Ye F. F., Liu J., Wang Y. M., Hu H., Environmental investment prediction using extended belief rule-based system and evidential reasoning rule, Journal of Cleaner Production, 2021, 289: 1-19.