New Activation Weight Calculation and Parameter Optimization for Extended Belief Rule-Based System Based on Sensitivity Analysis

Long-Hao Yang[a, c], Jun Liu[d], Ying-Ming Wang[a, b, *], Luis Martínez[c]

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

[b]Key Laboratory of Spatial Data Mining & Information Sharing of Ministry of Education, Fuzhou University, Fuzhou, PR China

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

[d]School of Computing and Mathematics, Ulster University, Northern Ireland, UK

*Corresponding author. E-mail: msymwang@hotmail.com; Tel: +86-0591-22866677

Abstract: An extended belief rule-based (EBRB) system is superior to existing rule-based systems in managing several types of uncertain information and modeling complex issues effectively and efficiently. However, the accuracy and interpretability of the EBRB system still need to be enhanced by addressing the following shortcomings: the interpretability of the intermediate variables in the EBRB system should be definite and the system parameters must be effectively determined. Therefore, we distinguish discrete and continuous data types to perform sensitivity analysis twice: first, on the rule inference scheme to study the interpretability of individual matching degrees and activation weights; and second, on the rule generation scheme to examine the effect of utility values and attribute weights on the accuracy of the EBRB system. Based on the analyses, we propose a novel activation weight calculation method and parameter optimization method to enhance the interpretability and accuracy of the EBRB system, respectively. We then present three case studies to elucidate the effectiveness of the proposed methods. The results indicate that the enhanced EBRB system prevents counterintuitive and insensitive situations and obtains better accuracies than some studies.

Keywords: Extended belief rule-based system; Activation weight calculation; Parameter optimization; Interpretability; Accuracy

1. Introduction

Among various systems and models, the rule-based system, a grey-box modeling type [12], has the system structure constructed by a set of rules and the system parameters determined based on observational data. While several rule-based systems, including the fuzzy rule-based (FRB) system [20][35] and the belief rule-based (BRB) system [30][34], have been proved to be efficient in complex system modeling, the enhancement of interpretability and accuracy has been debated to affect modeling performances [19][27]. Considering the BRB system, the parameter optimization [31] and the structure optimization [36] were proposed to determine optimal parameter values for enhancing accuracy and reasonable parameter numbers for improving interpretability, respectively.

Recently, Liu *et al*. [14] developed the extended belief rule-based (EBRB) system, an advanced rule-based system better than the FRB and BRB systems, having the following advantages:

(1) The extended belief rule in the EBRB system has a generic information representation better than the fuzzy rule in the FRB system and the belief rule in the BRB system, because the extended belief rule can express probabilistic, fuzzy,

1

and incomplete uncertainties in both antecedent and consequent attributes.

(2) The EBRB system can be knowledge-driven, or data-driven, or their combination decision model. Consequently, the extended belief rules are automatically generated from sample input-output data pairs, without requiring the time-consuming optimization process.

Furthermore, numerous studies have been conducted to enhance the interpretability and accuracy of the EBRB system in recent years. Calzada *et al*. [3] discussed the problem of incompleteness and inconsistency in the EBRB system. They suggested that the dynamic rule activation (DRA) method could be used to better determine the number of activation rules. Yang *et al*. [33] introduced two types of tree-based databases to construct a multi-attribute search framework (MaSF) for the EBRB system. The MaSF uses the space relationship between rules for activating extended belief rules efficiently. Espinilla *et al*. [11] proposed the adaption of the EBRB system for human activity recognition based on data collected from binary sensors in a smart environment. They suggested using the Hamming distance for discrete binary data than the Euclidean distance in activation weight calculation to improve the accuracy of the EBRB system. Yang *et al*. [32] discussed that the EBRB system lacked methods to evaluate the efficiency of extended belief rules and the data envelopment analysis (DEA) method was introduced to reduce these inefficient rules.

However, these studies have the following two shortcomings and demand improvements to enhance the interpretability and accuracy of the EBRB system.

The first shortcoming is that none of these studies aims to assure the interpretability of intermediate variables produced by the interpretable parameters of the EBRB system. For example, the attribute weight denotes the importance of the antecedent attribute. The more important the antecedent attribute is, the bigger value will be subjectively provided to the attribute weight. However, intermediate variables, including individual matching degrees and activation weights, are probably uninterpretable because they must be objectively calculated according to the interpretable parameters, and no relevant investigations show their interpretability. Moreover, because these studies are based on intermediate variables, they may weaken the interpretability of the EBRB system.

The second shortcoming is that none of these studies focuses on determining optimal values of relevant system parameters. Reviewing the previous literature [14] reveals that the basic parameter values, including the attribute weights and utility values of reference values for antecedent and consequent attributes, of the EBRB system are usually determined by expertise. In essence, the subjective method to determine the value of these basic parameters may not always be available, because lacking of domain knowledge and measurement data is challenging for experts to provide effective parameter values. How to optimally determine the basic parameter values of the EBRB system in order to enhance its accuracy has not yet been investigated.

To overcome aforementioned two shortcomings, we perform sensitivity analysis introduced in [15][16] for the EBRB system to investigate its potential enhancements for interpretability and accuracy. Sensitivity analysis has been proved to be effective for testing the robustness of the results, facilitating understanding of the input-output variables' relationship, and searching for errors in a model, among others.

For assuring interpretability of intermediate variables, we examine the rule inference scheme of the EBRB system to summarize various intermediate variables, including the Euclidean, standardized Euclidean, and Hamming-based individual matching degrees and activation weights. By distinguishing continuous and discrete data types, we perform sensitivity analysis associated with diverse test input data, extended belief rules, rule weights, or attribute weights to study the interpretability of individual matching degrees and activation weights. Based on the analyses, we propose a novel activation weight calculation method to enhance interpretability of the EBRB system.

For determining optimal values of relevant system parameters, we examine the rule generation scheme of the EBRB system to summarize the influence of the basic parameters on accuracy. By distinguishing discrete and continuous data types, we perform sensitivity analysis for studying the relationship between the basic parameters and the accuracy of the EBRB system. Accordingly, we propose a novel parameter optimization method to obtain optimal values of relevant basic parameters for enhancing accuracy.

In order to elucidate the effectiveness of the proposed methods, we use three case studies based on two prediction problems and five classification problems, along with meticulous comparison between the conventional EBRB system and other existing studies in these three studies, to demonstrate their ability in enhancing interpretability and accuracy of the EBRB system.

The remainder of this paper is organized as follows: Section 2 briefly reviews the EBRB system. Section 3 investigates the interpretability of the EBRB system via sensitivity analysis and proposes a new activation weight calculation method. Section 4 investigates the accuracy of the EBRB system via sensitivity analysis and proposes a new parameter optimization method. Section 5 provides three case studies to demonstrate the effectiveness of the proposed methods for the EBRB system in terms of interpretability and accuracy, and the paper is concluded in Section 6.

2. A review of EBRB systems

The EBRB system, as a data-driven rule-based system, comprises three components shown in Fig. 1: EBRB that is the knowledge base to store uncertain information in the form of extended belief rules, rule generation scheme providing an efficient procedure to generate EBRB from a set of sample input-output data, and rule inference scheme having an effective procedure to reply test input data based on the EBRB.
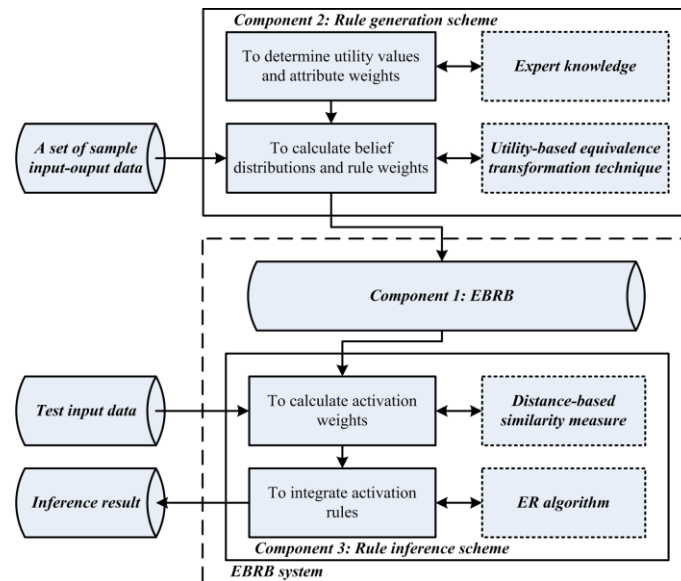

Fig. 1. An EBRB system

2.1. Basic concepts of EBRB systems

An EBRB system comprises M antecedent attributes and one consequent attribute. The antecedent attribute U_i ($i=1,\ldots, M$) is described by J_i reference values $A_{i,j}$ ($j=1,\ldots, J_i$), and the consequent attribute D is described by N reference values D_n ($n=1,\ldots, N$). Thus, an extended belief rule of the EBRB can be written as [14]:

$$R_k : IF\ U_1\ is\ \{(A_{1,j},\alpha_{1,j}^k); j=1,\ldots,J_1\} \wedge \cdots \wedge U_M\ is\ \{(A_{M,j},\alpha_{M,j}^k); j=1,\ldots,J_M\},$$
$$THEN\ D\ is\ \{(D_n,\beta_n^k); n=1,\ldots,N\}, with\ rule\ weight\ \theta_k\ and\ attribute\ weights\ \{\delta_1,\ldots,\delta_M\} \tag{1}$$

where $\alpha_{i,j}^k$ ($0 \le \alpha_{i,j}^k \le 1$) and β_n^k ($0 \le \beta_n^k \le 1$) are the belief degrees to the reference value $A_{i,j}$ and D_n in the kth ($k=1,\ldots, L$) rule with $\sum_{j=1}^{J_i}\alpha_{i,j}^k \le 1$ and $\sum_{n=1}^{N}\beta_n^k \le 1$, δ_i ($0 < \delta_i \le 1$), and θ_k ($0 < \theta_k \le 1$) are the weights of the ith antecedent attribute and the kth rule.

We consider the leak detection of an oil pipeline [28] to instantiate the extended belief rule as follows:

$$R_k : IF\ Pressure\ Difference\ is\ \{(Low,0),(Medium,0.1),(High,0.9)\}$$
$$\wedge\ Flow\ Difference\ is\ \{(Low,0.1),(Medium,0.8),(High,0)\}$$
$$THEN\ Leak\ Size\ is\ \{(Small,0),(Medium,0.5),(Large,0.5)\} \tag{2}$$
$$with\ rule\ weight\ 0.9\ and\ attribute\ weights\ \{0.8,0.1\}$$

The antecedent attributes are *Pressure Difference* and *Flow Difference* with reference values *Low*, *Medium*, and *High*; the consequent attribute is *Leak Size* with reference values *Small*, *Medium*, and *Large*. While 10% sure that *Pressure Difference* is *Medium* and 90% that it is *High*, and 10% sure that *Flow Difference* is *Low* and 80% it is *Medium*, then 50% sure that *Leak Size* is *Medium* and 50% that it is *Large*. As the total belief degree of *Flow Difference* is 0.1+0.8=0.9<1.0, the extended belief rule contains the incomplete uncertainty. Furthermore, the attribute weight of *Pressure Difference* and *Flow Difference* is 0.8 and 0.1, respectively, which reflect the different importance of two antecedent attributes. The rule weight of the kth rule is 0.9, and it represents the importance/reliability of the rule over other rules.

2.2. Rule generation scheme

Fig. 1 illustrates that the rule generation scheme is a vital method for the EBRB system to generate an EBRB. According to [14], the following two steps are summarized to describe the rule generation scheme.

Step 1: To determine utility values and attribute weights using expert knowledge.

Prior to generating extended belief rules from sample input-output data, pre-determining relevant system parameters of an EBRB is essential. These parameters include reference values required for each antecedent and consequent attributes, utility values assigned for these reference values, and attribute weights that are appropriate for antecedent attributes. Hence, in this step, expert knowledge is usually utilized to pre-determine these parameters.

Step 2: To calculate belief distributions and rule weights using the utility-based equivalence transformation technique.

For generating a complete extended belief rule, we suppose that $\{u(A_{i,j}); j=1,\ldots, J_i\}$ is a set of given utility values used for the ith ($i=1,\ldots, M$) antecedent attribute, and $x_{k,i}$ is the kth ($k=1,\ldots, L$) sample input data of the ith antecedent attribute. Thus, the belief distribution of the ith antecedent attribute generated using the utility-based equivalence transformation technique [29] is:

$$S(x_{k,i}) = \{(A_{i,j},\alpha_{i,j}^k); j=1,\ldots, J_i\} \tag{3}$$

where

$$\alpha_{i,j}^{k} = \frac{u(A_{i,j+1}) - x_{k,i}}{u(A_{i,j+1}) - u(A_{i,j})} \text{ and } \alpha_{i,j+1}^{k} = 1 - \alpha_{i,j}^{k}, \text{if } u(A_{i,j}) \leq x_{k,i} \leq u(A_{i,j+1}),$$ (4)

$$\alpha_{i,t}^{k} = 0, \text{ for } t = 1,..., J_{i} \text{ and } t \neq j, j+1,$$ (5)

where $A_{i,j}$ represents the jth reference value of the ith antecedent attribute, $\alpha_{i,j}^{k}$ is the belief degree of $A_{i,j}$ in the kth rule obtained from the sample input data $x_{k,i}$.

Similarly, when the kth sample output data is y_k and the given utility values attached to the consequent attribute D are $\{u(D_n);$ $n=1,\ldots, N\}$, the belief distribution of the consequent attribute is:

$$S(y_k) = \{(D_n, \beta_n^k); n = 1,...,N\}$$ (6)

According to the belief distribution of antecedent and consequent attributes, the rule weight of each extended belief rule is generated by using the similarity measure (please see [14] for details). Finally, these belief distributions and weights are used to generate an extended belief rule shown in Eq. (1).

2.3. Rule inference scheme

Fig. 1 indicates that the rule inference scheme is another vital method for the EBRB system to reply test input data. According to [14], the rule inference scheme consists of the following steps:

Step 1: To calculate the activation weight of each extended belief rule based on the distance-based similarity measure.

We consider a test input data vector $\boldsymbol{x}=(x_1,\ldots, x_M)$ provided for an EBRB system; each input x_i will be transformed or assessed into the belief distribution of the reference values of the ith antecedent attribute using Eqs. (4) and (5):

$$S(x_i) = \{(A_{i,j}, \alpha_{i,j}); j = 1,...,J_i\}$$ (7)

where $\alpha_{i,j}$ is the likelihood to which the test input x_i belongs to the reference value $A_{i,j}$

Next, the individual matching degree of x_i to the ith antecedent attribute in the kth rule, denoted as $S^k(x_i, U_i)$, is calculated using the Euclidean distance:

$$S^k(x_i, U_i) = 1 - d^k(x_i, U_i) = 1 - \sqrt{\sum_{j=1}^{Ji} (\alpha_{i,j} - \alpha_{i,j}^k)^2},$$ (8)

where $\alpha_{i,j}^{k}$ ($j=1,\ldots, J_i$) are the belief degrees generated from the kth ($k=1,\ldots, L$) sample input data $x_{k,i}$ for U_i using Eq. (3), $d^k(x_i, U_i)$ is the distance measurement, and $S^k(x_i, U_i)$ measures the matching degree between belief distribution generated from x_i and the one generated from $x_{k,i}$ for the ith antecedent attribute U_i in the kth rule.

Thus, the activation weight for the kth rule, denoted as w_k, is calculated as:

$$w_k = \frac{\theta_k \prod_{i=1}^{M} \left(S^k(x_i, U_i)\right)^{\bar{\delta}_i}}{\sum_{l=1}^{L} \left(\theta_l \prod_{i=1}^{M} \left(S^l(x_i, U_i)\right)^{\bar{\delta}_i}\right)}, \bar{\delta}_i = \frac{\delta_i}{\max_{i=1,...,M} \{\delta_i\}}$$ (9)

where θ_k is the weight of the kth rule; δ_i is the weight of the ith antecedent attribute.

Step 2: To integrate all activation rules using the ER algorithm.

After calculating activation weights, all activation rules are integrated using the analytical ER algorithm [26]:

$$\beta_n = \frac{\prod_{k=1}^{L}\left(w_k\beta_n^k + 1 - w_k\sum_{i=1}^{N}\beta_i^k\right) - \prod_{k=1}^{L}\left(1 - w_k\sum_{i=1}^{N}\beta_i^k\right)}{\sum_{i=1}^{N}\prod_{k=1}^{L}\left(w_k\beta_i^k + 1 - w_k\sum_{j=1}^{N}\beta_j^k\right) - (N-1)\prod_{k=1}^{L}\left(1 - w_k\sum_{j=1}^{N}\beta_j^k\right) - \prod_{k=1}^{L}(1-w_k)}. \tag{10}$$

For the prediction problem, based on the utility values $\{u(D_n); n=1,\dots, N\}$, the numerical inference result is:

$$f(\boldsymbol{x}) = \sum_{i=1}^{N}\left(u(D_i)\beta_i\right) + \frac{u(D_1)+u(D_N)}{2}\left(1 - \sum_{i=1}^{N}\beta_i\right). \tag{11}$$

For the classification problem, suppose that D_n denotes the nth class, the final inference class is:

$$f(\boldsymbol{x}) = D_n, n = \arg\max_{i=1,\dots,N}\{\beta_i\}. \tag{12}$$

3. Sensitivity analysis on the rule inference scheme for activation weight calculation

In this section, we use the important intermediate variables involved in the rule inference scheme, including individual matching degrees and activation weights, to examine the interpretability of the EBRB system through sensitivity analysis. Based on the analyses, we propose a novel activation weight calculation method to enhance the interpretability of the EBRB system.

3.1. Sensitivity analysis on individual matching degrees

The individual matching degree is one kind of intermediate variables in the rule inference scheme that illustrates the significance of an activation rule for different antecedent attributes. A review of the rule inference scheme shows that calculating individual matching degrees depends on distance measures, including Euclidean, standardized Euclidean, and Hamming distances.

For calculating individual matching degrees, consider two input data $x_{k,i}$ and x_i attached to the ith antecedent attribute, where the former is a sample data to generate the kth extended belief rule and the latter is a test data likely to activate extended belief rules. Both input data can be transformed into belief distributions: $S(x_{k,i})=\{(A_{i,j}, \alpha_{i,j}^k); j=1,\dots, J_i\}$ and $S(x_i)=\{(A_{i,j}, \alpha_{i,j}); j=1,\dots, J_i\}$. Therefore, in addition to the Euclidean distance-based individual matching degree shown in Eq. (8), other individual matching degrees based on different distance measures are listed as follows:

The standardized Euclidean distance-based individual matching degree [33] is:

$$S^k(x_i, U_i) = 1 - \sqrt{\frac{\sum_{j=1}^{J_i}(\alpha_{i,j} - \alpha_{i,j}^k)^2}{2}} \tag{13}$$

The Hamming distance-based individual matching degree [11] is:

$$S^k(x_i, U_i) = 1 - \frac{\sum_{j=1}^{J_i}d_H(\alpha_{i,j},\alpha_{i,j}^k)}{J_i}, d_H(\alpha_{i,j},\alpha_{i,j}^k) = \begin{cases}0, if\ \alpha_{i,j} = \alpha_{i,j}^k \\ 1, otherwise\end{cases} \tag{14}$$

To instantiate these individual matching degrees, we assume that the utility values of reference values for the ith antecedent attribute are assessed by:

$$\{u(A_{i,j}); j = 1, 2, 3\} = \{0,\ 0.5,\ 1\} \tag{15}$$

The input data $x_{k,i}$ and x_i can be grouped into two categories: continuous and discrete. The continuous data are those that take any value having a range, such as human height, flow difference, and pressure difference, whereas the discrete data are those that only take certain discrete values, such as human sex, output of binary sensors, and number of classes. Hence, if the input data is

6

continuous, assume the range of both $x_{k,i}$ and x_i is the interval [0, 1]; otherwise, assume the range is the set {0, 0.5, 1}. Figs. 2 and 3 present the curves of three individual matching degrees.
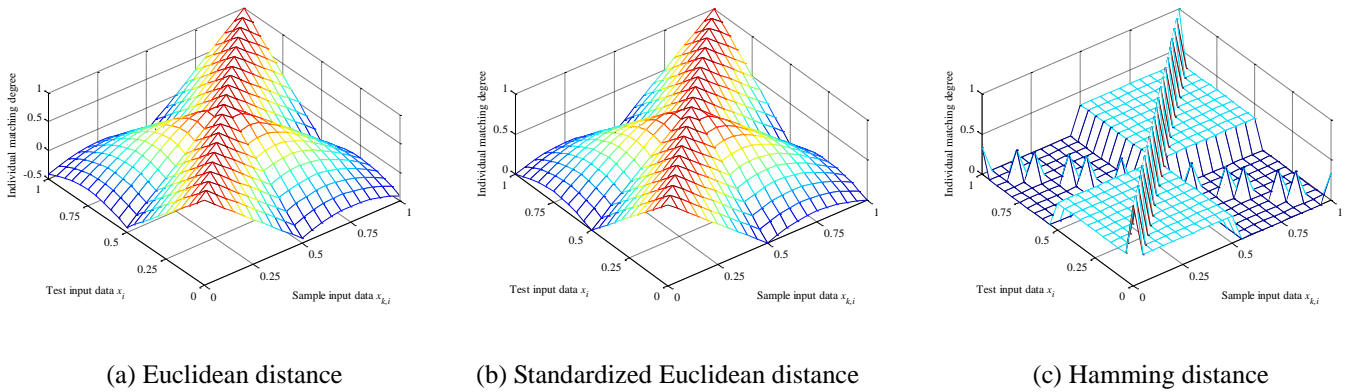

(a) Euclidean distance   (b) Standardized Euclidean distance   (c) Hamming distance

Fig. 2. Sensitivity analysis on individual matching degrees for continuous data

(a) Euclidean distance   (b) Standardized Euclidean distance   (c) Hamming distance
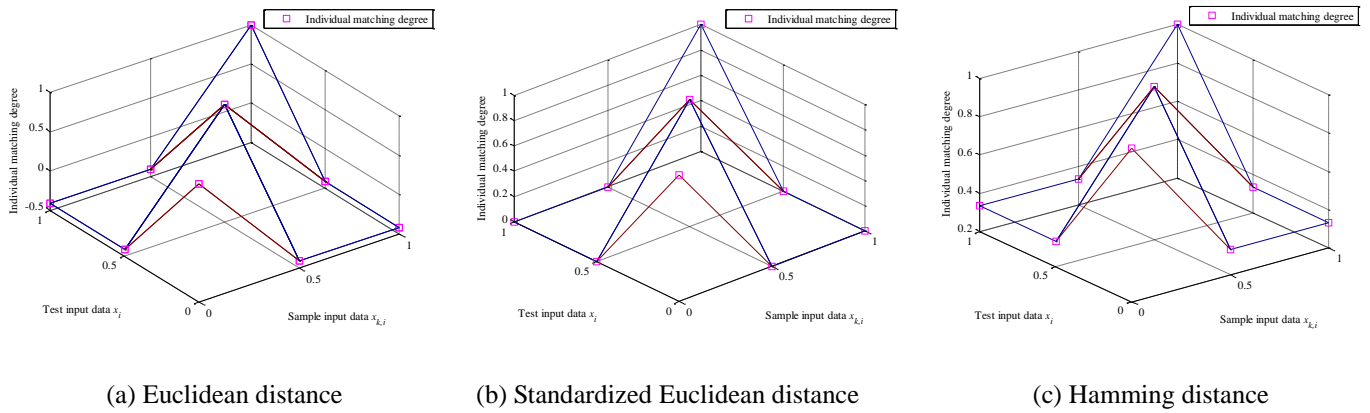
Fig. 3. Sensitivity analysis on individual matching degrees for discrete data

Based on Figs. 2 and 3, two problems are observed in studying the interpretability of individual matching degrees.

Problem 1: The individual matching degree may be counterintuitive.

A significant situation is observed in the Euclidean distance-based individual matching degree for continuous and discrete data. Figs. 2(a) and 3(a) show numerous individual matching degrees as negative, which in fact are counterintuitive. For example, for test input data $x_i=0$ and sample input data $x_{k,i}=0.5$ are provided for the EBRB system, the individual matching degree is -0.414 for continuous and discrete data.

Another significant situation is observed in the standardized Euclidean and Hamming distance-based individual matching degrees for continuous data. Figs. 2(b) and 3(b) demonstrate that, with fixed test input data x_i or sample input data $x_{k,i}$, several minimal and maximal extreme points are crossed. These extreme points counterintuitively reflect the similarity between test input data and extended belief rules. For example, when the sample input data $x_{k,i}$ is 0 for the standardized Euclidean distance-based individual matching degree, a maximal extreme point $x_i=0.75$ is observed in the interval between two minimal extreme points $x_i=0.5$ and $x_i=1$; however, the minimal extreme point $x_i=0.5$ is in another interval between two maximal extreme points $x_i=0.75$ and $x_i=0$.

Problem 2: The individual matching degree may be insensitive to test input data.

A significant situation is observed in three individual matching degrees for continuous and discrete data. Figs. 2 and 3

7

indicate that several individual matching degrees to be same for different test input data that actually are insensitive to particular test input data. For example, when sample input data $x_{k,i}=0$ is used to calculate the standardized Euclidean distance-based individual matching degree, individual matching degrees are same for the different test input data $x_i=0.5$ and $x_i=1$ for continuous and discrete data.

Another significant situation is observed in the Hamming distance-based individual matching degree for discrete data. Fig. 3(c) illustrates that individual matching degrees are not equal to 0, indicating that the individual matching degree is in a local value domain while expressing the similarity between the test and sample input data. For example, although the extreme sample input data $x_{k,i}=0$ and the test input data $x_i=1$ are provided for the EBRB system, the individual matching degree is 0.33, not 0.

3.2. Sensitivity analysis on activation weights

In this subsection, we perform sensitivity analysis on activation weights for investigating the interpretability of another intermediate variable observed in the rule inference scheme. Considering extended belief rules consisting only one antecedent attribute, according to Eq. (9), the activation weight of the kth extended belief rule is simplified as the following unnormalized formula:

$$w_k = \theta_k \left(S^k(x_i, U_i) \right)^{\delta_i} \tag{16}$$

Fig. 4 presents two curves of activation weights relative to individual matching degrees, attribute weights, and rule weights, assuming $\theta_k = 1$ in Fig. 4(a) and $\delta_i = 1$ in Fig. 4(b).

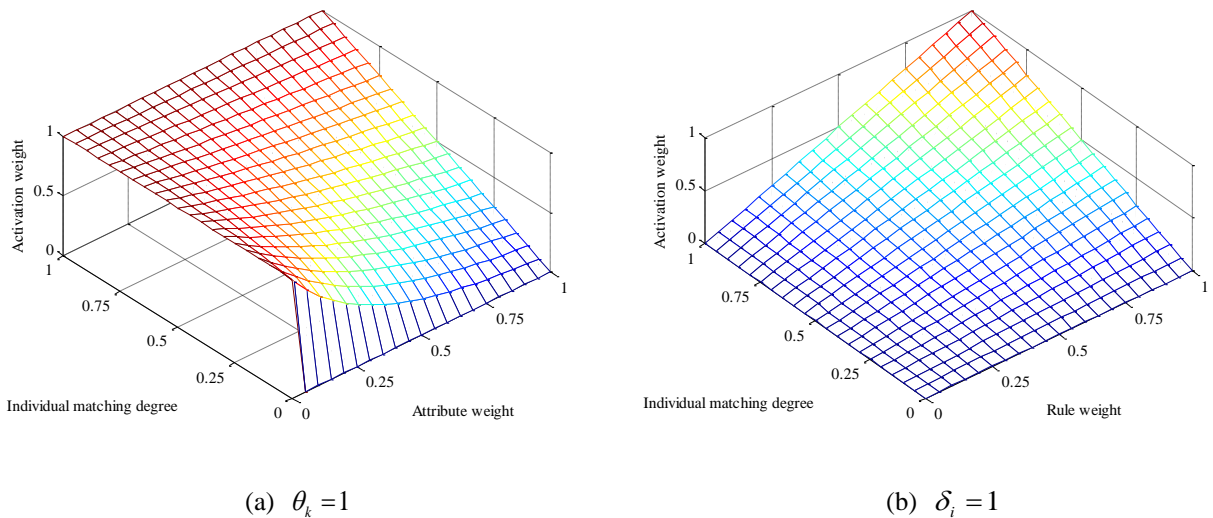(a) $\theta_k = 1$ (b) $\delta_i = 1$

Fig. 4. Sensitivity analysis on activation weights

According to Fig. 4, one problem is observed in investigating the interpretability of activation weights.

Problem 3: The attribute weight is negatively correlated with the activation weight.

Fig. 4 illustrates that both the rule weight and individual matching degree are positively correlated with activation weights. With an extended belief rule having bigger rule weight or individual matching degree indicating this rule is more important than others, producing a bigger activation weight is a high possibility. However, Fig. 4(a) indicates the relationship between the attribute weight and the activation weight as contradictory. It is clearly counterintuitive for the EBRB system, because a bigger

attribute weight produces a smaller activation weight. For example, when the antecedent attribute is ineffective for the EBRB system and its attribute weight is therefore assumed 0, the resulting activation weight has the biggest value.

For investigating the interpretability of activation weights, we consider the ith antecedent attribute with three reference values, the utility values of its reference values are shown in Eq. (15), and the belief distribution of the ith antecedent attribute for the kth extended belief rule is:

$$\{(A_{i,j}, \alpha_{i,j}^k); j = 1,2,3\} = \{(A_{i,1},1),(A_{i,2},0),(A_{i,3},0)\} \tag{17}$$

For continuous and discrete data, Figs. 5 and 6 show the curves of three types of distance-based activation weights with respect to test input data x_i and attribute weights.
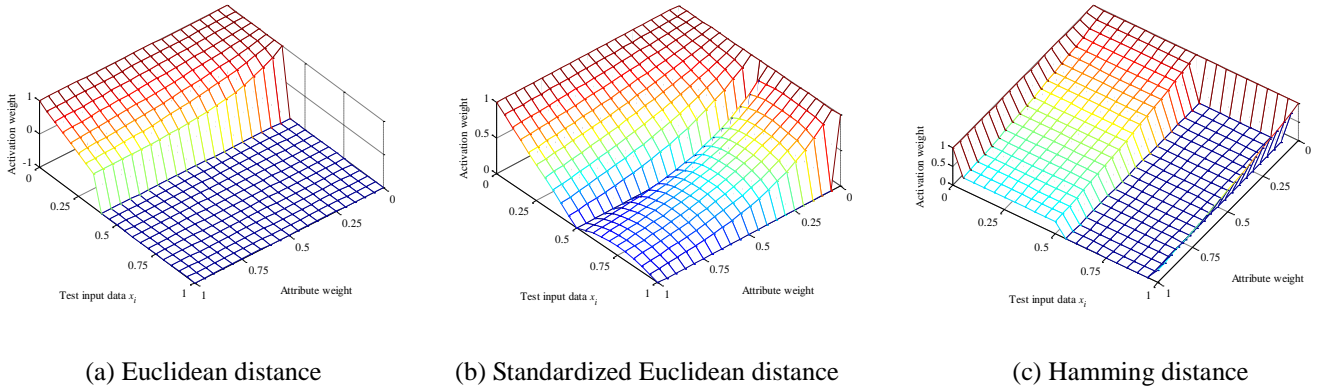

| (a) Euclidean distance | (b) Standardized Euclidean distance | (c) Hamming distance |

Fig. 5. Sensitivity analysis on activation weights based on different distance measures for continuous data

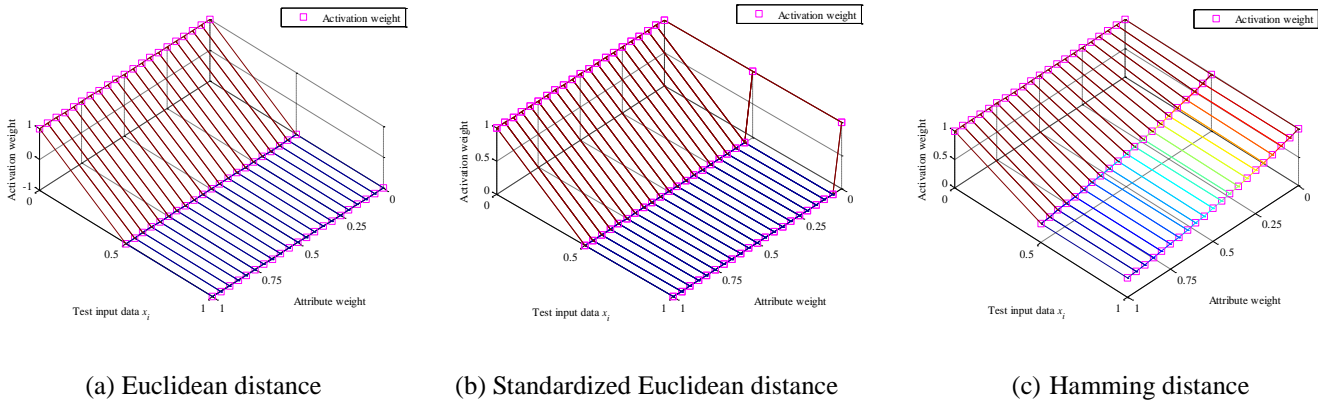| (a) Euclidean distance | (b) Standardized Euclidean distance | (c) Hamming distance |

Fig. 6. Sensitivity analysis on activation weights based on different distance measures for discrete data type

Based on Figs. 5 and 6, another problem is observed in investigating the interpretability of activation weights.

Problem 4: The activation weight may be counterintuitive and insensitive to test input data.

Counterintuitive is observed in three types of activation weights for both continuous and discrete data. For the Euclidean distance-based activation weight, Figs. 5(a) and 6(a) illustrate several activation weights to be negative due to negative individual matching degrees shown in Figs. 2(a) and 3(a). In other words, when the test input data x_i=0.5 is provided for the EBRB system, the individual matching degree is -0.414 and further produces a negative activation weight. For the standardized Euclidean and Hamming distance-based activation weights for continuous data, Figs. 5(b) and 5(c) show that many minimal and maximal extreme points are crossed when the attribute weight is fixed. In other words, when the attribute weight is 0 for the standardized Euclidean distance-based activation weight, a maximal extreme point x_i=0.75 is observed in the interval between two minimal

extreme points x_i=0.5 and x_i=1, but the minimal extreme point x_i=0.5 is observed in the interval between two maximal extreme points x_i=0.75 and x_i=0.

Insensitivity to test input data is observed in three types of activation weights for both continuous and discrete data. For the standardized Euclidean distance-based activation weight, Figs. 5(b) and 6(b) indicate several activation weights to be same for different test input data. In other words, when the test input data x_i=0.5 and x_i =1 are provided for the EBRB system, the activation weights must be 0. Furthermore, for the Hamming distance-based activation weight, Fig. 6(c) shows that none of the activation weights is 0. This is similar to the individual matching degrees shown in Fig. 3(c).

3.3. Proposed activation weight calculation method

To overcome four problems identified from sensitivity analysis in Sections 3.1 and 3.2, we propose a novel activation weight calculation method for the EBRB system; Fig. 7 illustrates the general process.
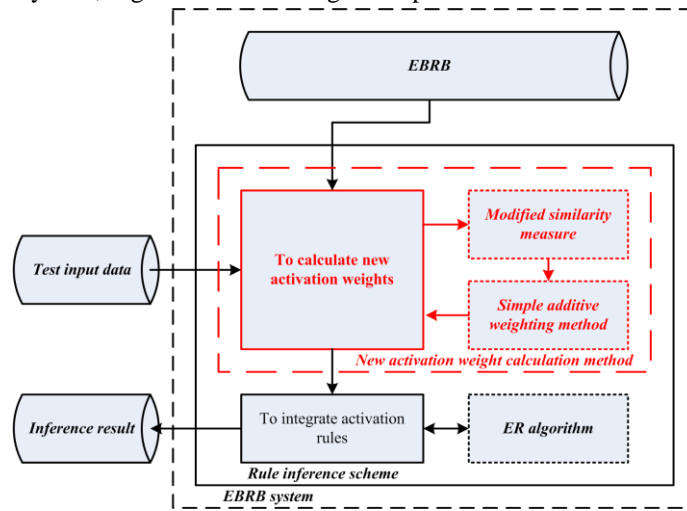

Fig. 7. Illustration of new activation weight calculation method for the EBRB system

The proposed activation weight calculation method works as follows:

Step 1: To transform the test input data into the distributed likelihood of the reference values using the modified similarity measure.

Consider a test input data vector x=(x_1,…,x_M), where M is the number of antecedent attributes. Each input x_i (i=1,…, M) is transformed into the distributed representation of the reference values of the ith antecedent attribute:

$$S(x_i) = \{(A_{i,j}, \gamma_{i,j}); j = 1,...,J_i\}, \tag{18}$$

where

$$\gamma_{i,j} = 1.0 - \frac{|u(A_{i,j}) - x_i|}{\max_{k=1,...,J_i}\{u(A_{i,k})\} - \min_{k=1,...,J_i}\{u(A_{i,k})\}} \tag{19}$$

Rather than using Eqs. (4) and (5), we use Eq. (19) to calculate $\gamma_{i,j}$ as a new modification, where $u(A_{i,j})$ is the utility value of the reference value $A_{i,j}$; $\gamma_{i,j}$ is the likelihood to which the test input x_i belongs to the reference value $A_{i,j}$; and J_i is the number of reference values referred to the ith antecedent attribute.

Step 2: To calculate individual matching degrees for each antecedent attribute using the simple additive weighting method.

Unlike distance-based measures given in Eqs. (8), (13), or (14), we calculate the individual matching degree of x_i to the ith antecedent attribute in the kth rule as follows:

$$S^k(x_i, U_i) = \sum_{j=1}^{J_i} \gamma_{i,j} \alpha_{i,j}^k \tag{20}$$

where $\alpha_{i,j}^k$ ($j=1,\ldots,J_i$) is the belief degrees generated from the kth ($k=1,\ldots,L$) sample input data $x_{k,i}$ for U_i using Eq. (3).

Step 3: To calculate new activation weights for each extended belief rule.

We calculate the new activation weight w_k for the kth extended belief rule as follows:

$$w_k = \frac{\theta_k \prod_{i=1}^M \left(S^k(x_i, U_i)\right)^{1/\delta_i}}{\sum_{l=1}^L (\theta_l \prod_{i=1}^M \left(S^k(x_i, U_i)\right)^{1/\delta_i})} \tag{21}$$

where θ_k is the weight of the kth rule; δ_i is the weight of the ith antecedent attribute.

In order to investigate the interpretability of the EBRB system with new activation weight calculation method, we perform sensitivity analysis on individual matching degrees. Fig. 8 illustrates the relationship between individual matching degree of the test input data x_i to the kth rule generated from the sample input data $x_{k,i}$ for U_i for continuous and discrete data.
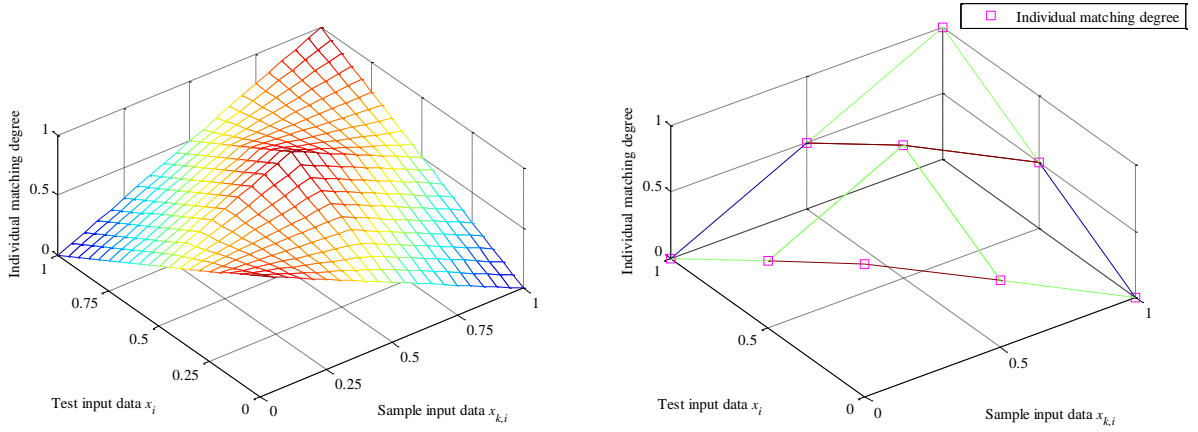
(a) Individual matching degree for continuous data　　　(b) Individual matching degree for discrete data

Fig. 8. Sensitivity analysis on new individual matching degrees

Remark 1: Fig. 8 shows several advantages comparing to the existing distance-based individual matching degrees: first, the new individual matching degree belongs to the global value range [0, 1] for continuous and discrete data, and second, only one maximal and two minimal extreme points are available for any given test input data. Hence, it is believed that the new activation weight calculation method can overcome Problems 1 and 2.

For sensitivity analysis on new activation weights, we consider the extended belief rules consisting one antecedent attribute for illustration; its unnormalized calculation is:

$$w_k = \theta_k \left(S^k(x_i, U_i)\right)^{1/\delta_i} \tag{22}$$

Fig. 9, by assuming that $\theta_k = 1$ and $\delta_i = 1$, shows two curves of new activation weights.

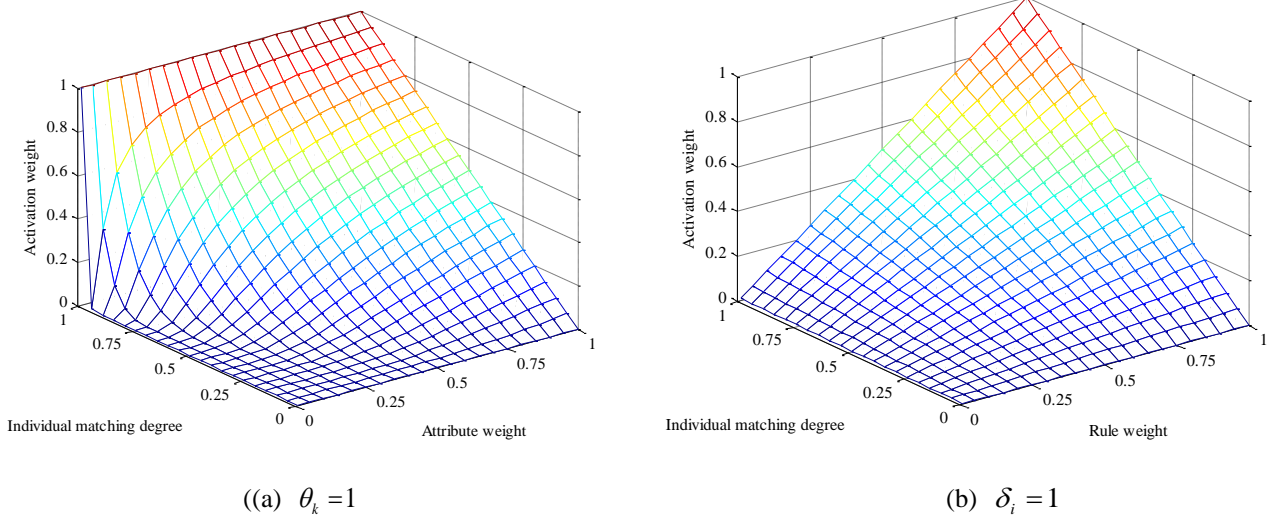((a) $\theta_k = 1$ (b) $\delta_i = 1$

Fig. 9. Sensitivity analysis on new activation weights in terms of attribute and rule weights

Remark 2: Fig. 9 shows that the rule weights, attribute weights, and individual matching degrees are positively correlated with activation weights. Hence, it is believed that the new activation weight calculation method can overcome Problem 3.

To perform sensitivity analysis on new activation weights, we consider utility values of reference values for the ith antecedent attribute shown in Eq. (15) and the belief distribution of the ith antecedent attribute for the kth extended belief rule given in Eq. (17). Fig. 10 illustrates the curves of new activation weights with respect to test input data x_i and attribute weights.
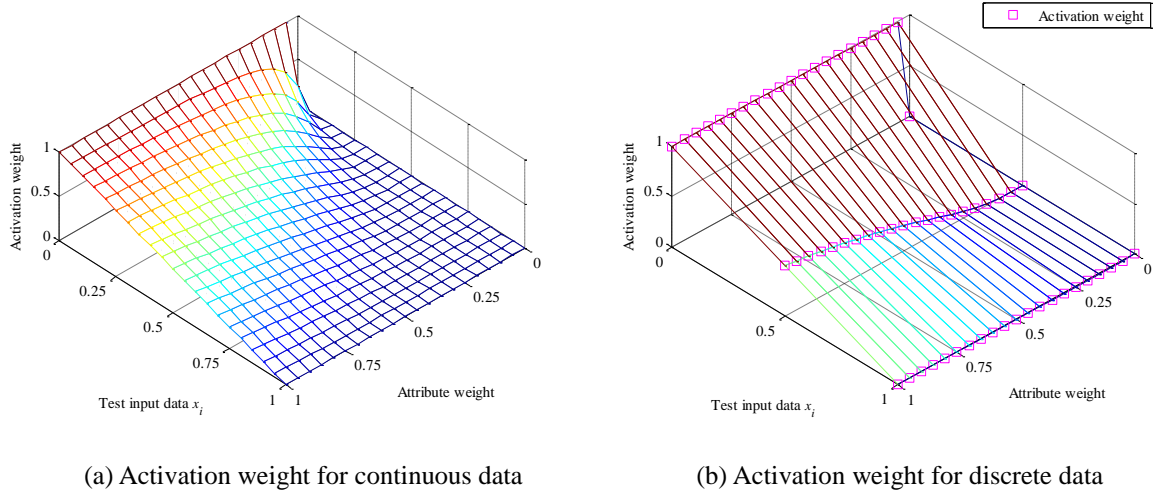

(a) Activation weight for continuous data   (b) Activation weight for discrete data

Fig. 10. Sensitivity analysis on new activation weights in terms of test input data and attribute weight

Remark 3: Fig. 10 proves that several advantages over existing distance-based activation weights: first, new activation weights belong to the global value range [0, 1] for continuous and discrete data, and second, only one maximal and one minimal extreme point are provided for any given test input data. Hence, it is believed that the new activation weight calculation method can overcome Problem 4.

4. Sensitivity analysis on the rule generation scheme with new activation weight calculation for parameter optimization

In this section, we use the basic parameters in the rule generation scheme, including utility values of the reference values for the antecedent/consequent attribute and the attribute weights, for examining their effect on the accuracy of the EBRB system with

new activation weight calculation via sensitivity analysis. Based on the analyses, we propose a novel parameter optimization method to determine optimal values of these basic parameters for enhancing the accuracy of the EBRB system.

4.1. Sensitivity analysis on utility values and attribute weights used for antecedent attributes

The utility values and attribute weights used for antecedent attributes are the basic parameters for generating extended belief rules. However, they are normally determined by expertise, making them unreliable for complex system modeling. For studying their effect on accuracy of the EBRB system, we perform sensitivity analysis in term of activation weights.

We consider that extended belief rules contain only one antecedent attribute (e.g., the ith antecedent attribute) and three utility values for the ith antecedent attribute whose utility values are:

$$\{u(A_{i,j}); j = 1,2,3\} = \{0, 0 < u(A_{i,2}) < 1, 1\} \tag{23}$$

Furthermore, we consider the input data $x_{k,i}$ and x_i as sample data to generate the kth extended belief rule and a test data to activate extended belief rules, respectively. According to Eq. (21), the activation weight of the kth extended belief rule is (the detailed formal derivation is provided in Appendix A):

$$w_k = \begin{cases} \theta_k \left(1 - x_i - x_{k,i} + \dfrac{2x_i x_{k,i}}{u(A_{i,2})}\right)^{1/\delta_i}, & if \ 0 \le x_i, x_{k,i} \le u(A_{i,2}) \\ \theta_k \left(1 - x_i + x_{k,i}\right)^{1/\delta_i}, & if \ 0 \le x_{k,i} \le u(A_{i,2}), u(A_{i,2}) < x_i \le 1 \\ \theta_k \left(1 + x_i - x_{k,i}\right)^{1/\delta_i}, & if \ u(A_{i,2}) < x_{k,i} \le 1, 0 \le x_i \le u(A_{i,2}) \\ \theta_k \left(1 + x_{k,i} + x_i + \dfrac{2x_i x_{k,i} - 2x_{k,i} - 2x_i}{1 - u(A_{i,2})}\right)^{1/\delta_i}, & if \ u(A_{i,2}) < x_i, x_{k,i} \le 1 \end{cases} \tag{24}$$

where θ_k is the weight of the kth extended belief rule, and δ_i is the weight of the ith antecedent attribute.

Eq. (24) explains that the activation weight of the kth extended belief rule can be affected by the utility values and attribute weights used for antecedent attributes. However, the utility value $u(A_{i,2})$ affects activation weights when the test input data x_i and sample input data $x_{k,i}$ belong to the same interval of utility values. For studying the effect of utility values and attribute weights on accuracy of the EBRB system, we suppose that the test input data x_i is 0.5 and the rule weight is 1. Figs. 11 and 12 illustrate the curves of activation weights for various attribute weights for continuous and discrete data.
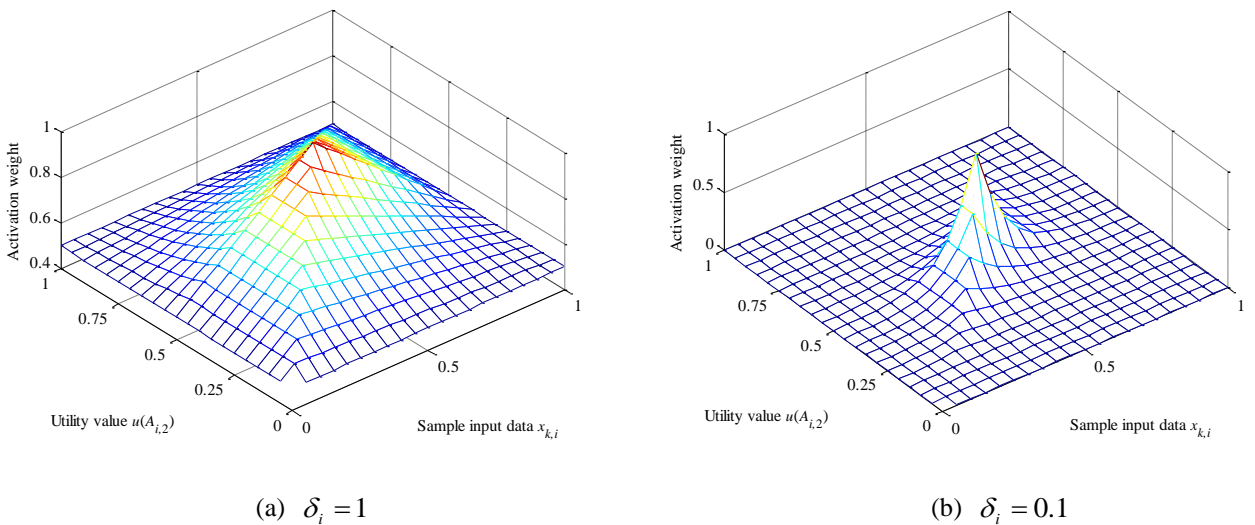

(a) $\delta_i = 1$ (b) $\delta_i = 0.1$

Fig. 11. Sensitivity analysis on utility values used for antecedent attributes for continuous data

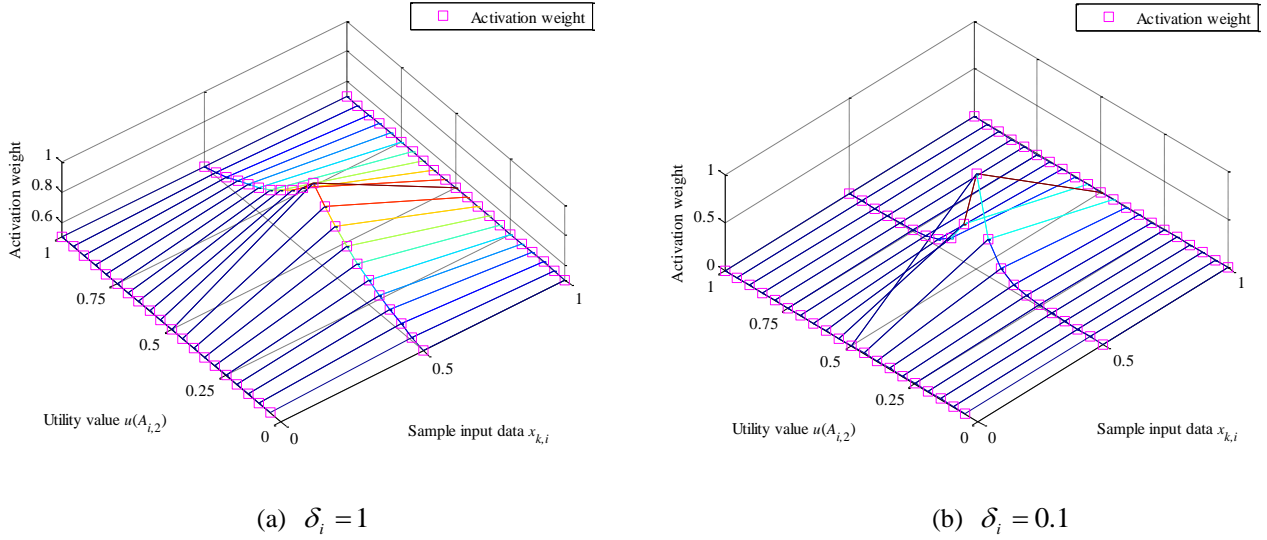(a) $\delta_i = 1$ (b) $\delta_i = 0.1$

Fig. 12. Sensitivity analysis on utility values used for antecedent attributes for discrete data

Figs. 11 and 12 summarize one conclusion while investigating the influence of utility values and attribute weights used for antecedent attributes on the accuracy of the EBRB system.

Conclusion 1: Optimal utility values and attribute weights used for antecedent attributes facilitate activation of the desired extended belief rules for the EBRB system in response to the given test input data.

Figs. 11 and 12 clarify a significant situation that when the sample input data $x_{k,i}$ is 0.5, which is equal to the test input data x_i, the utility value $u(A_{i,2})=0.5$ or $u(A_{i,2})=0$ makes the kth extended belief rule to have a maximum or minimum activation weight. Here, the kth extended belief rule is an anticipated rule to reply the test input data x_i because the sample input data $x_{k,i}$ for generating the kth extended belief rule equal the test input data x_i. Furthermore, the maximum or minimum activation weight of the kth extended belief rule proves the influence of utility values on the accuracy of the EBRB system because activating the desired rule with a bigger activation weight is advantageous. In addition, Figs. 8(b) and 9(b) demonstrate that when the attribute weight is 0.1, a minimum activation weight for the kth extended belief rule is produced. In other words, the attribute weight affects the accuracy of the EBRB system.

4.2. Sensitivity analysis on utility values used for consequent attribute

In the rule generation scheme, utility values used for the consequent attribute are similar to those for antecedent attributes; although they are used to generate extended belief rules, they must be pre-determined based on expertise. Hence, we perform sensitivity analysis on utility values used for the consequent attribute.

We suppose the output data y_k and y_l as sample data to generate the kth and lth extended belief rules; three utility values used for the consequent attribute are:

$$\{u(D_n); n=1,2,3\} = \{0, 0 < u(D_2) < 1, 1\} \tag{25}$$

According to the formal formula derivation provided in Appendix B, we express the integrated belief degree of the reference value D_2 as:

$$\beta_2 = \begin{cases} \dfrac{w_k w_l y_k y_l}{\left(u(D_2)\right)^2} + \dfrac{(w_k)^2 y_k + (w_l)^2 y_l}{u(D_2)}, & if\ 0 \le y_k, y_l \le u(D_2) \\[2ex] \dfrac{w_k w_l y_k (1-y_l)}{u(D_2)\left(1-u(D_2)\right)} + \dfrac{(w_k)^2 y_k}{u(D_2)} + \dfrac{(w_l)^2 (1-y_l)}{1-u(D_2)}, & if\ 0 \le y_k \le u(D_2),\ u(D_2) < y_l \le 1 \\[2ex] \dfrac{w_k w_l (1-y_k) y_l}{\left(1-u(D_2)\right)u(D_2)} + \dfrac{(w_k)^2 (1-y_k)}{1-u(D_2)} + \dfrac{(w_l)^2 y_l}{u(D_2)}, & if\ u(D_2) < y_k \le 1,\ 0 \le y_l \le u(D_2) \\[2ex] \dfrac{w_k w_l (1-y_k)(1-y_l)}{\left(1-u(D_2)\right)^2} + \dfrac{(w_k)^2 (1-y_k) + (w_l)^2 (1-y_l)}{1-u(D_2)}, & if\ u(D_2) < y_k, y_l \le 1 \end{cases} \quad (26)$$

where w_k and w_l are the activation weight of the kth and lth extended belief rules, respectively.

Eq. (26) demonstrates that utility values used for the consequent attribute affect all integrated belief degrees associated with the reference value D_2. Furthermore, we suppose the sample output data y_l as 0.5, the range of the sample output data y_k is the interval [0, 1.0] for continuous data or the set {0, 0.5, 1} for discrete data. Fig. 13 and 14 illustrate the curves of integrated belief degrees for different activation weights for continuous and discrete data.
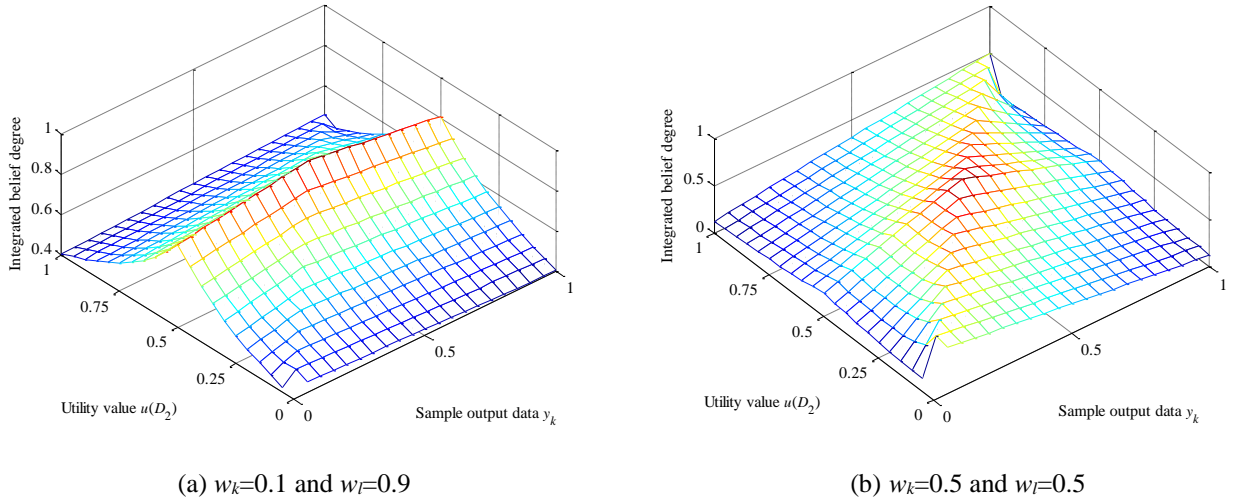

(a) w_k=0.1 and w_l=0.9   (b) w_k=0.5 and w_l=0.5

Fig. 13. Sensitivity analysis on utility values used for consequent attribute for continuous data

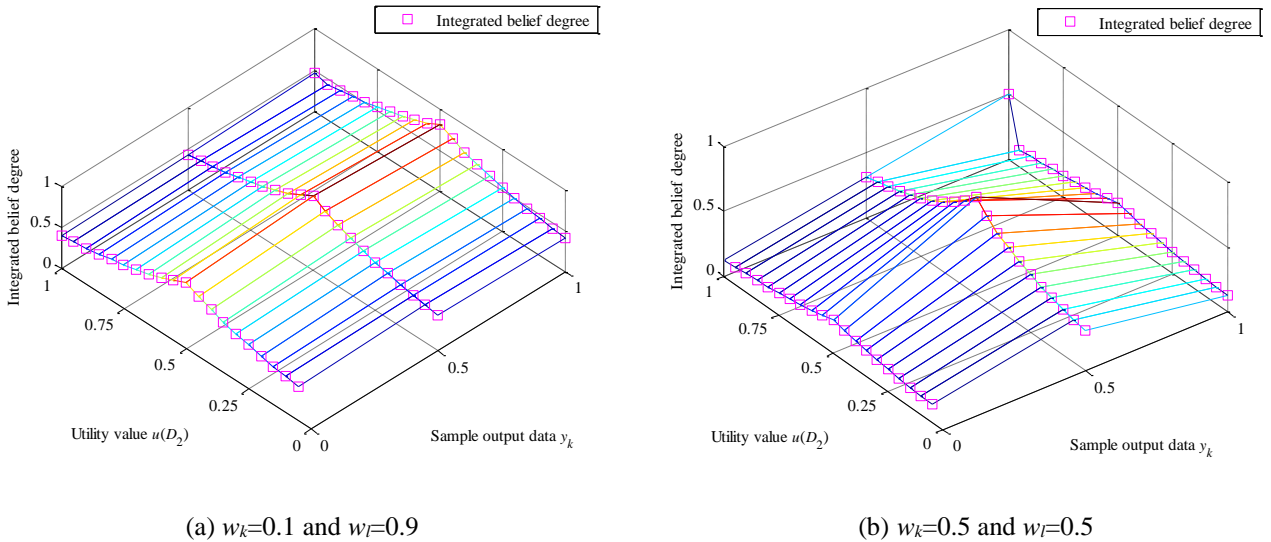(a) w_k=0.1 and w_l=0.9   (b) w_k=0.5 and w_l=0.5

Fig. 14. Sensitivity analysis on utility values used for consequent attribute for discrete data

Figs. 13 and 14 summarize one conclusion while investigating the influence of utility values used for the consequent attribute on the accuracy of the EBRB system.

Conclusion 2: Optimal utility values used for the consequent attribute produce the required inference results for the EBRB system in response to the given test input data.

For different activation weights, such as w_k=0.1 and w_l=0.9, the lth extended belief rule is influential on the inference result of the EBRB system comparing with the kth extended belief rule. Figs. 13(a) and 14(a) show that when the utility value $u(D_2)$ and sample output data y_l are same, it is useful to ensure the inference result of the EBRB system to approximate the output of the lth extended belief rule without the influence of the kth extended belief rule.

For same activation weights, such as w_k=0.5 and w_l=0.5, the lth and kth extended belief rules is equally important for the inference result of the EBRB system. Figs. 13(b) and 14(b) illustrate that when the utility value $u(D_2)$ and sample output data y_k are same, the inference result of the EBRB system approximates the output data of the kth extended belief rule. The same result is observed for the output of the lth extended belief rule for the sample output data y_l.

4.3. New parameter optimization model and algorithm

Conclusions 1 and 2 verified that, with the new activation weight calculation, the optimal utility values in both antecedent and consequent attributes and the attribute weights are important in enhancing the accuracy of the EBRB system. However, these utility values and attribute weights are determined using expertise from the literature on the EBRB system, thereby making it fundamental to propose an appropriate method. In this subsection, we propose a new parameter optimization method and its methodological framework is illustrated in Fig. 15.
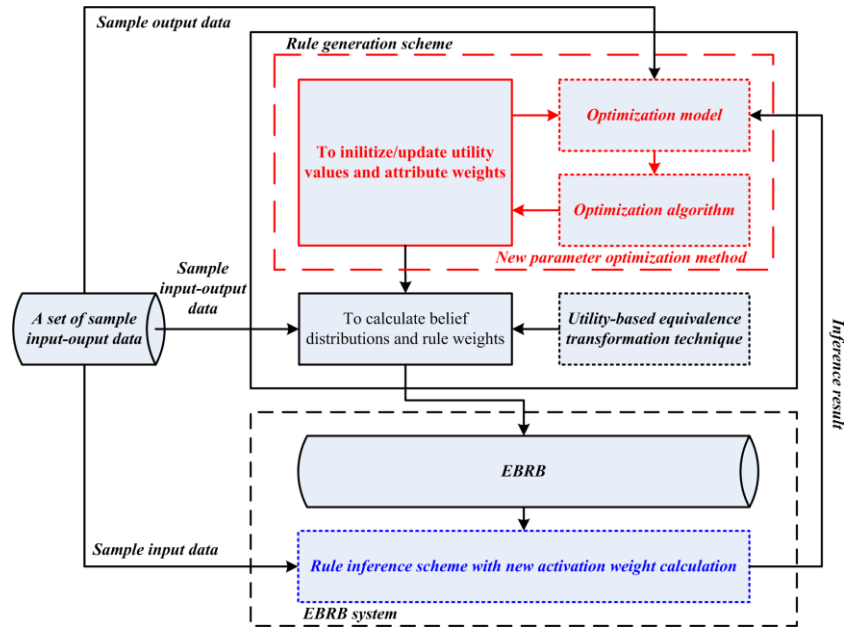
Fig. 15. Illustration of parameter optimization involved in the rule generation scheme

Fig. 15 shows that the new parameter optimization method includes two components: optimization model and optimization algorithm. The following constraints are are required for the optimization model:

(1) Each attribute weight should have the the following constraint:

$$0 < \delta_i \leq 1; i = 1,...,M \tag{27}$$

(2) The utility value $u(A_{i,j})$ of reference value $A_{i,j}$ should have the following constraints:

$$u(A_{i,j}) \le u(A_{i,j+1}); i = 1,...,M; j = 1,...,J_1 - 1 \tag{28a}$$

$$u(A_{i,1}) = lb_i; i = 1,...,M \tag{28b}$$

$$u(A_{i,J_i}) = ub_i; i = 1,..., M \tag{28c}$$

where lb_i and ub_i are the lower and upper bounds of the ith antecedent attribute, respectively.

(3) The utility value $u(D_n)$ of reference value D_n should have the following constraints:

$$u(D_n) \le u(D_{n+1}); n = 1,...,N-1 \tag{29a}$$

$$u(D_1) = lb \tag{29b}$$

$$u(D_N) = ub \tag{29c}$$

where lb and ub are the lower and upper bounds of the consequent attribute, respectively.

Furthermore, when T sample input-output data (x_t, y_t) $(t=1,..., T)$ are provided for the EBRB system, the target function of optimization model is expressed as the mean absolute error (MAE):

(1) For the prediction problem, the target function is :

$$\min MAE(\{\delta_i, u(A_{i,j}), u(D_n)\}) = \sum_{t=1}^{T} |y_t - f(x_t)| \tag{30}$$

where $f(x_t)$ is the inference result of the EBRB system for the sample input data x_t.

(2) For the classification problem, the target function is:

$$\min MAE(\{\delta_i, u(A_{i,j}), u(D_n)\}) = \sum_{t=1}^{T} E_t \tag{31}$$

where

$$E_t = \begin{cases} 0; if\ f(x_t)\ is\ equal\ to\ y_t \\ 1; otherwise \end{cases} \tag{32}$$

where $f(x_t)$ is the final inference class of the EBRB system for the sample input data x_t, y_t is the real class of x_t.

For this optimization model, we observe the following remarks:

Remark 4: As indicated in the rule generation scheme, the rule weights and belief distributions of antecedent and consequent attributes are generated from a set of sample input-output data. Thus, training these parameters in the optimization model is inessential. However, for each obtained optimal utility value and attribute weight, it is required to update rule weights and belief distribution of each extended belief rule.

Remark 5: A comparison with existing optimization models [4][5][6] indicates that the proposed optimization model includes fewer parameters to be trained and the number of these parameters does not increase with the increasing number of rules.

In order to obtain the minimal MAE of the optimization model, we introduce the optimization algorithm based on differential evolution (DE) algorithm [13][18]. The DE algorithm, a cutting-edge evolutionary algorithm for simple and straightforward framework, has been proved to outperform other evolutionary algorithms in a series of performance contests [17][21]. The proposed optimization algorithm comprises the following steps:

Step 1 (Initialization): We suppose that the optimization algorithm comprises C individuals and S generations to train the basic parameters of an EBRB system, where the cth individual in the sth generation is expressed as:

$$P_{s,c} = \{p_k^{s,c}; k = 1,..., K\} = \{\delta_i^{s,c}, u(A_{i,j})^{s,c}, u(D_n)^{s,c}\}; s = 1,..., S; c = 1,..., C \qquad (33)$$

where $\delta_i^{s,c}$ is the weight of the ith antecedent attribute associated with the cth individual in the dth generation, $u(A_{i,j})^{s,c}$ and $u(D_n)^{s,c}$ are the utility values of the reference value $A_{i,j}$ and D_n related to the cth individual in the dth generation, $p_k^{s,c}$ is the kth parameter related to the cth individual in the sth generation, and K is the total number of these utility values and attribute weights.

For initializing individuals based on the constraints of Eqs. (27) to (29), we consider ub_k and lb_k as the upper and lower bounds of $p_k^{s,c}$, respectively. The initial value of parameters is generated by using a random value between 0 and 1:

$$p_k^{0,c} = lb_k + (ub_k - lb_k) * random(0,1); c = 1,...,C; k = 1,...,K \qquad (34)$$

Step 2 (Evolution): In the sth generation, for the cth individual, we generate a new individual P_{s,c_0} by using three different individuals randomly selected from the C individuals, in which the three individuals are signed as P_{s,c_1}, P_{s,c_2}, and P_{s,c_3}. The parameter value of P_{s,c_0} is assigned as follows:

$$p_k^{s,c_0} = \begin{cases} p_k^{s,c}, & if\ random(0,1) > CR \\ p_k^{s,c_1} + F * (p_k^{s,c_2} - p_k^{s,c_3}), otherwise \end{cases}, c = 1,...,C; k = 1,...,K \qquad (35)$$

where F is the mutation operator and CR is the crossover operator.

Step 3 (Selection): In the sth generation, when the parameter value of P_{s,c_0} is out of the constraints shown in Eqs. (27) to (29), we produce a new parameter value using Eq. (34). Then, based on the rule inference scheme with new activation weight calculation, we generate the MAE of P_{s,c_0} by using T sample input-output data and update the parameter values of $P_{s,c}$:

$$P_{s,c} = \begin{cases} P_{s,c_0}; if\ MAE(P_{s,c_0}) < MAE(P_{s,c}) \\ P_{s,c}; \ otherwise \end{cases} \qquad (36)$$

Step 4 (Termination): When the current generation is equal to S, the individual with the minimum MAE is selected as the best individual, and its parameter values are regarded as the optimal attribute weights and the optimal utility values used for antecedent and consequent attributes.

For this optimization algorithm, we provide the following remarks:

Remark 6: On basis of optimal attribute weights and utility values, optimal rule weights and belief distributions of antecedent and consequent attributes are generated from sample input-output data. Finally, all these optimal parameters construct a trained EBRB system.

Remark 7: In the optimization algorithm, the number of individuals C and generations S are set according to the desired accuracy of users. Large C and S can improve a better performance of the trained EBRB system, but compromising on the running time of the parameter optimization. According to the literature [5], the mutation operator F and crossover operator CR are set as 0.5 and 0.9, respectively.

Remark 8: In addition to the DE algorithm, Chen *et al*. [7][8][9] and Tsai *et al*. [22][23] proposed optimization techniques that can be further studies for obtaining the minimal MAE of the proposed parameter optimization method.

5. Case studies

In this section, we consider two prediction problems, namely oil pipeline leak detection and bridge risk assessment, and five classification problems to demonstrate the effectiveness of the proposed activation weight calculation method and parameter optimization method in enhancing the interpretability and accuracy of the EBRB system.

5.1. Datasets and experiment conditions

The oil pipeline leak detection belongs to the continuous data because its original dataset was collected from the real-time parameter of an oil pipeline; the bridge risk assessment belongs to the discrete data because of discrete rating used for assessing bridge structures. For the five classification problems selected from the KEEL dataset repository [1], Table 1 summarizes the features of all datasets, indicating the number of data (#Data), number of attributes (#Attributes), number of continuous attributes (#Continuous) and discrete attributes (#Discrete), and number of classes (#classes). We use these datasets to observe the performance of the EBRB system with the proposed activation weight calculation and parameter optimization methods for continuous and discrete data.

Table 1. Statistics on classification datasets

Number	Dataset	#Data	#Attributes	#Continuous	#Discrete	#Classes
1	Bupa	345	6	1	5	2
2	Wine	178	13	13	0	3
3	Glass	214	9	9	0	7
4	Ecoli	336	7	7	0	8
5	Hayes-roth	160	4	0	4	3

A comparison of different activation weight calculation methods is performed by constructing four EBRB systems with the Euclidean, standardized Euclidean, and Hamming distance-based activation weight calculations and new activation weight calculation. These four EBRB systems are abbreviated as EBRB-E, EBRB-S, EBRB-H, and EBRB-New.

In order to demonstrate the effectiveness of the proposed parameter optimization method, all EBRB systems are used to achieve the optimization of the four parameters, including attribute weights, utility values in both antecedent and consequent attributes. Additionally, since the proposed optimization model comprises fewer parameters to be trained than in the previous optimization models used in the BRB system, we set the algorithm parameter to a small value for obtaining desired accuracies. Based on the point of view, the number of individuals and generations are set to 20 and 40, respectively, in the case study of prediction and classification problems. Furthermore, we implement the EBRB systems and the proposed optimization algorithm through the Microsoft Visual C++ on Intel (R) Core (TM) i5-4300U CPU at 1.90GHz and 4GB RAM with Windows 7.

5.2. EBRB systems for continuous prediction problem

5.2.1. Oil pipeline leak detection using the EBRB system with the proposed methods

The leak detection of the oil pipeline installed in Great Britain [28] is a common benchmark used for testing the BRB system and its improvements, including the local training BRB system [28], adaptive training BRB system [10], and optimal training

BRB system [6]. Totally, 2008 input-output data $\{(x_{t,1}, x_{t,2}, y_t); t=1,\ldots, 2008\}$ are available, where $x_{t,1}$ is the continuous input data of the flow difference (FD), $x_{t,2}$ is the continuous input data of the pressure difference (PD), and y_t is the output data of the leak size (LS) of an oil pipeline at the sampling time t.

In order to construct an EBRB system, we randomly generate 500 input-output data from the original 2008 input-output data with eight and seven reference values applied to describe the FD and PD, respectively. These reference values include negative large (NL), negative medium (NM), negative small (NS), negative very small (NVS), zero (Z), positive small (PS), positive medium (PM), and positive large (PL). For the LS, we use five reference values, including zero (Z), very small (VS), medium (M), high (H), and very high (VH). Based on the previous study [28], the initial utility values of these reference values and the attribute weights are:

$$(A_{FD}, u(A_{FD})) \in \{(NL,-10),(NM,-5),(NS,-3),(NVS,-1),(Z,0)(PS,1),(PM,2),(PL,3)\} \tag{37a}$$

$$(A_{PD}, u(A_{PD})) \in \{(NL,-0.02),(NM,-0.005),(NS,-0.002),(Z,0)(PS,0.002),(PM,0.005),(PL,0.02)\} \tag{37b}$$

$$(D_{LS}, u(D_{LS})) \in \{(Z,0),(VS,2),(M,4),(H,6),(VH,8)\} \tag{37c}$$

$$\{\delta_{FD}, \delta_{PD}\} = \{1, 1\} \tag{37d}$$

According to the rule generation scheme together with the basic parameters shown in Eqs. (37a)-(37d), 500 input-output data are used to generate the initial rule weight and the initial belief distributions for each extended belief rule. Finally, all these initial parameter values construct an initial EBRB system.

For obtaining the optimal utility values and attribute weights, we use the 500 input-output data together with the proposed optimization model and algorithm to train the basic parameters. The trained parameters are shown as follows:

$$(A_{FD}, u(A_{FD})) \in \{(NL,-10),(NM,-7.8422),(NS,-6.5429),(NVS,-3.0714),(Z,-1.8962)(PS,0.1149),(PM,1.588),(PL,3)\} \tag{38a}$$

$$(A_{PD}, u(A_{PD})) \in \{(NL,-0.02),(NM,-0.0152),(NS,-0.0054),(Z,-0.0002)(PS,0.0065),(PM,0.0129),(PL,0.02)\} \tag{38b}$$

$$(D_{LS}, u(D_{LS})) \in \{(Z,1),(VS,2.3801),(M,5.7010),(H,6.5345),(VH,8)\} \tag{38c}$$

$$\{\delta_{FD}, \delta_{PD}\} = \{0.0111, 0.2470\} \tag{38d}$$

Accordingly, the optimal rule weight and the optimal belief distributions are generated based on the optimal utility values and attribute weights shown in Eqs. (38a) - (38d). Finally, all these optimal parameter values construct a trained EBRB system.

Afterwards, we use the 2008 input-output data to compare the initial and trained EBRB systems. Fig. 16 shows that the initial and trained EBRB systems clearly detect the leaks which happened at around $t=850$ and ended at around $t=1400$. However, the estimated leak sizes of the initial EBRB system are different from the observed ones, but the trained EBRB system matches the observed leak sizes accurately.
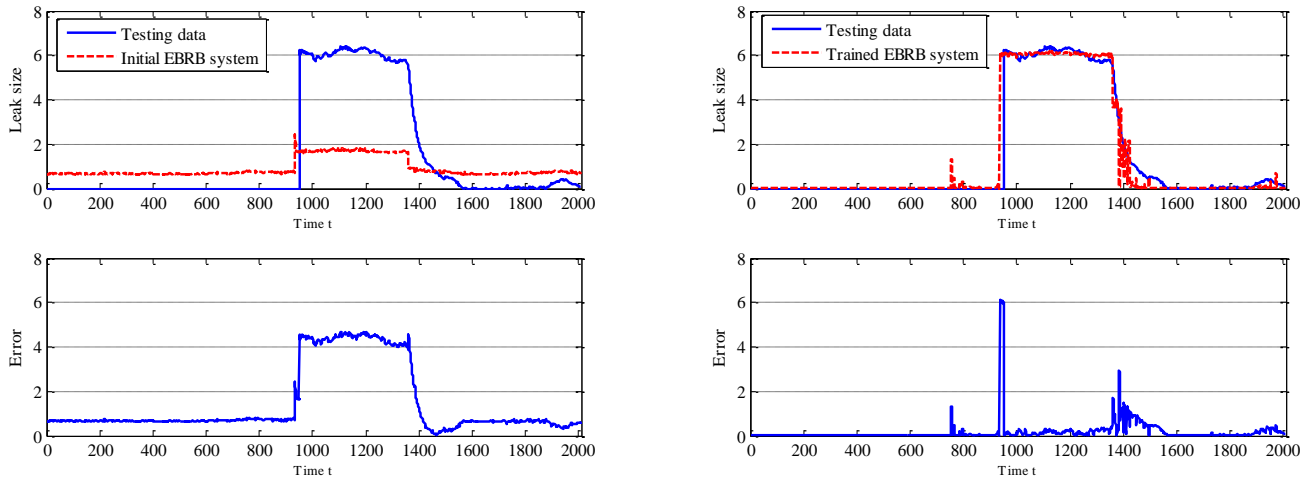
(a) Estimated leak size and error of initial EBRB system   (b) Estimated leak size and error of trained EBRB system

Fig. 16. Comparison of initial and trained EBRB systems for oil pipeline leak detection

5.2.2. Performance comparisons with conventional EBRB systems and existing studies

After performing 10 cycles of the parameter optimization of the EBRB system with different 500 input-output data, we compare the average results obtained from EBRB-E, EBRB-S, EBRB-H, and EBRB-New with the following measures: 1) interpretability of the EBRB system, which evaluates the number of continuous test input data leading to the counterintuitive execution (CE) and insensitive execution (IE) in the rule inference scheme; 2) accuracy of the EBRB system, which evaluates the MAE and the mean squared error (MSE) of the EBRB system; and 3) the time of the EBRB system, which indicates the training time of the proposed parameter optimization method.

Table 2 presents the CE, IE, MAE, MSE, and time after comparing the EBRB-E, EBRB-S, EBRB-H, and EBRB-New using 2008 input-output data. The EBRB-New obtains the minimum MAE and MSE by using the trained parameters and prevents the counterintuitive and insensitive situations for continuous input data in a relatively few training time. These are the common problems of the EBRB-E, EBRB-S, and EBRB-H, which are investigated in Section 3, so that the CE and IE of these EBRB systems are greater than 0. Moreover, the comparison of four EBRB systems with the initial and trained parameters indicates that the optimal utility values and attribute weights are indeed helpful to enhance the accuracy of the EBRB systems.

Table 2. Comparisons of the initial and trained EBRB systems

EBRB type	Initial parameters				Time (s)	Trained parameters			
	CE	IE	MAE	MSE		CE	IE	MAE	MSE
EBRB-E	2008	2008	0.2388	0.5072	150.4	2008	199	0.1983	0.3818
EBRB-S	2008	2008	0.8210	1.5276	216.7	2008	72	0.4406	0.7140
EBRB-H	1893	2008	1.5846	4.6389	173.1	0	2008	0.9107	2.3176
EBRB-New	0	0	1.4478	4.2644	165.5	0	0	**0.1671**	**0.3540**

To demonstrate the effectiveness of the proposed activation weight calculation and parameter optimization methods, we summarize existing studies in the same oil pipeline leak detection application. Table 3 presents the comparison results and clearly demonstrates that the proposed methods significantly facilitate enhancing the performance of the EBRB system. First, the average

MAE of this study is better than other improved BRB systems. Second, the average MSE of this study is the second-best result. Third, the number of parameters to be optimized is 22, including 15 utility values used for antecedent attributes, 5 utility values used for consequent attribute, and 2 attribute weights, whereas other improved BRB systems require more parameters, such as 353 parameters in the local training and adaptive training.

Table 3. Comparison of EBRB system and existing BRB systems

No.	Year	Description	Size (train)	Size (test)	Parameters	MAE	MSE
1	2007 [28]	Local training	500	2008	353	0.2223	0.4049
2	2011 [10]	Adaptive training	500	2008	353	0.2064	0.3990
3	2015 [6]	Approximate causal inference	500	2008	107	0.2014	0.3709
4	2017 [4]	Bi-level BRB	500	2008	40	0.1941	0.2917
5	**The proposed method**	**Parameter optimization**	**500**	**2008**	**22**	**0.1671**	**0.3540**

In summary, for the oil pipeline leak detection, the comparison results have proved that the proposed activation weight calculation method enhances the interpretability of the EBRB system better than distance-based activation weight calculation methods. Moreover, the proposed parameter optimization method enhances the accuracy of the EBRB system, which shows the best MAE in the oil pipeline leak detection applications.

5.3. EBRB systems for discrete prediction problem

5.3.1. Bridge risk assessment using the EBRB system with the proposed method

The bridge risk assessment is a well-known benchmark to demonstrate varied conventional methods, including artificial neural network (ANN), evidential reasoning with learning (ERL), multiple regression analysis (MRA), and adaptive neuro-fuzzy inference system (ANFIS). Based on the British Highways Agency [2], there are 506 input-output data $\{(x_{t,1}, x_{t,2}, x_{t,3}, x_{t,4}, y_t);$ $t=1,\ldots,506\}$, where $x_{t,1}$ is the discrete input data of the safety (SA), $x_{t,2}$ is the discrete input data of the functionality (FU), $x_{t,3}$ is the discrete input data of the sustainability (SU), $x_{t,4}$ is the discrete input data of the environment (EN), and y_t is the output data of the risk score (RS) of the tth bridge project.

For constructing the EBRB system, we use the data preprocessing strategy [24] to generate 66 input-output data from the original 506 input-output data. Five reference values are applied to describe the antecedent and consequent attributes. These reference values include none (N), low (L), medium (M), high (H), and very high (VH). Based on expert knowledge, the initial utility values of these reference values and attribute weights are:

$$(A_{SA}, u(A_{SA})) \in \{(N,0),(L,0.5),(M,1.5),(H,2.5),(VH,3)\} \tag{39a}$$

$$(A_{FU}, u(A_{FU})) \in \{(N,0),(L,0.5),(M,1.5),(H,2.5),(VH,3)\} \tag{39b}$$

$$(A_{SU}, u(A_{SU})) \in \{(N,0),(L,0.5),(M,1.5),(H,2.5),(VH,3)\} \tag{39c}$$

$$(A_{EN}, u(A_{EN})) \in \{(N,0),(L,0.5),(M,1.5),(H,2.5),(VH,3)\} \tag{39d}$$

$$(D_{RS}, u(A_{RS})) \in \{(N,0),(L,25),(M,50),(H,75),(VH,100)\} \tag{39e}$$

$$\{\delta_{SA}, \delta_{FU}, \delta_{SU}, \delta_{EN}\} = \{1, 1, 1, 1\} \tag{39f}$$

Next, we use a combination of the rule generation scheme and basic parameters shown in Eqs. (39a) - (39f) to generate the initial EBRB system from the 66 input-output data, which are also used to train these basic parameters to obtain the optimal utility values and attribute weights as follows:

$$(A_{SA}, u(A_{SA})) \in \{(N,0),(L,0.8842),(M,1.3767),(H,2.2522),(VH,3)\} \tag{40a}$$

$$(A_{FU}, u(A_{FU})) \in \{(N,0),(L,1.8046),(M,1.9661),(H,2.6716),(VH,3)\} \tag{40b}$$

$$(A_{SU}, u(A_{SU})) \in \{(N,0),(L,0.6150),(M,1.7311),(H,2.3384),(VH,3)\} \tag{40c}$$

$$(A_{EN}, u(A_{EN})) \in \{(N,0),(L,0.9902),(M,1.7208),(H,2.6032),(VH,3)\} \tag{40d}$$

$$(D_{RS}, u(A_{RS})) \in \{(N,0),(L,14.3546),(M,46.3227),(H,79.0271),(VH,100)\} \tag{40e}$$

$$\{\delta_{SA}, \delta_{FU}, \delta_{SU}, \delta_{EN}\} = \{0.0526, 0.0540, 0.1128, 0.8306\} \tag{40f}$$

Accordingly, the 66 input-output data and obtained optimal basic parameters are used to generate optimal rule weights and belief distributions, and all these optimal parameters construct a trained EBRB system. Fig. 17 shows the comparisons of the initial and trained EBRB systems while testing 506 input-output data, where the estimated RS of the initial EBRB system fails to match the observed one, thereby causing large estimation errors at all testing data. Correspondingly, the training EBRB system assesses the RS of bridges accurately.
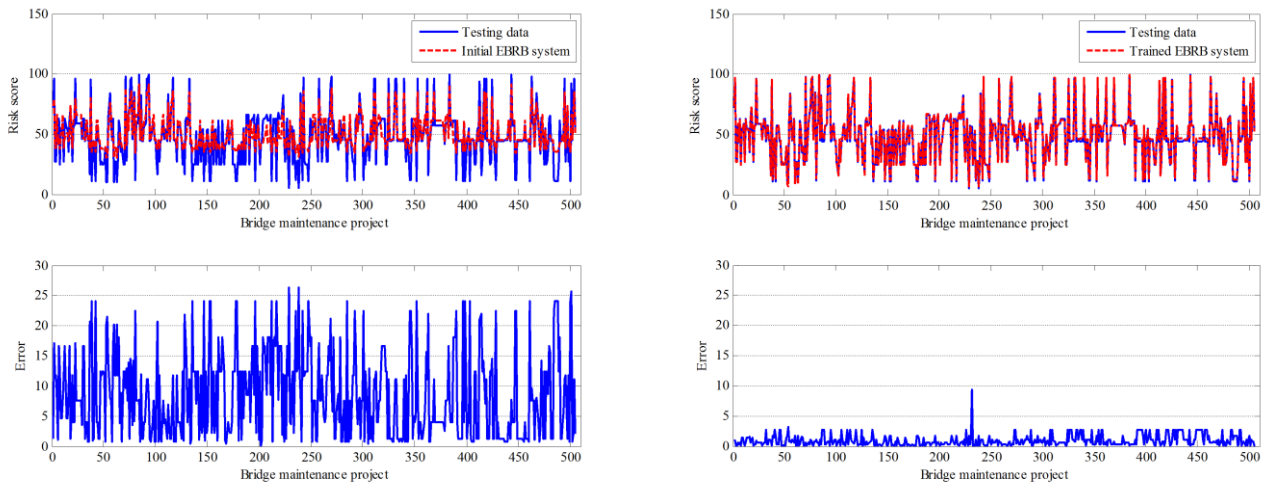
(a) Estimated risk scores and errors of initial EBRB system   (b) Estimated risk scores and errors of trained EBRB system

Fig. 17. Comparison of initial and trained EBRB systems for bridge risk assessment

5.3.2. Performance comparisons of conventional EBRB systems and existing studies

After performing 10 cycles of the parameter optimization of the EBRB system with the 66 input-output data, Table 4 shows the average results for the EBRB-E, EBRB-S, EBRB-H, and EBRB-New, which are compared by the following measures: 1) CE and IE, which indicate the interpretability of the EBRB system; 2) the mean absolute percentage error (MAPE), root mean squared error (RMSE), and correlation coefficient (R), which indicate the accuracy of the EBRB system; and 3) the time, which shows the training time of the proposed parameter optimization method.

Table 4. Comparisons of the initial and trained EBRB systems

EBRB type	Initial parameters					Time (s)	Trained parameters				
	CE	IE	MAPE	RMSE	R		CE	IE	MAPE	RMSE	R
EBRB-E	506	411	5.5959	2.6430	0.9926	3.3	506	0	**2.2996**	2.1102	0.9955
EBRB-S	0	411	14.6091	4.5384	0.9859	4.8	227	0	4.1887	2.8595	0.9925
EBRB-H	411	411	48.0239	15.4245	0.9327	4.5	497	506	4.0278	2.4886	0.9945
EBRB-New	0	0	33.6011	11.0746	0.8862	3.6	0	0	2.9785	**1.5941**	**0.9976**

Table 4 indicates that the EBRB-New achieves the minimum RMSE and R by using the trained parameters and presents the being counterintuitive and insensitive to discrete input data, namely the values of both CE and IE are 0. Although the EBRB-E has the minimum MAPE, 506 CEs are observed in the rule inference scheme. In other words, all discrete input data cause the EBRB-E to produce the counterintuitive situation, such as negative individual matching degrees or activation weights. For the EBRB-S, insensitivity to all discrete input data is prevented; however, the resulting MAPE, RMSE, and R are inferior to other EBRB systems.

To prove the effectiveness of the proposed activation weight calculation and parameter optimization methods, ten alternative models developed by using ANN, ERL, MRA, and ANFIS for bridge risk assessments are summarized here for comparison, denoted by BP-ANN, ER1, ER2, MRA3, MRA7, MRA8, and MRA9 based on [24], and ANFIS1, ANFIS2, and ANFIS3 while using triangular, trapezoidal, and Gaussian membership functions [25]. Table 5 presents these comparison results and the trained EBRB system with the proposed activation weight calculation method. The trained EBRB system produces satisfactory results in the MAPE, RMSE, and R, which are 2.9785, 1.5941, and 0.9976, respectively, and these results are better than the ten models.

Table 5. Comparisons of EBRB system and conventional approaches

Performance criterion	Models										The proposed method
	BP-ANN	ER1	MRA3	MRA8	ER2	MRA7	MRA9	ANFIS1	ANFIS2	ANFIS3	
MAPE	9.6294	22.4544	18.5775	23.9799	18.7808	24.1941	19.1456	6.7280	19.6389	15.7431	**2.9785**
RMSE	4.1871	8.9255	10.9527	10.4653	11.2736	10.3510	11.3653	3.4643	8.4179	6.5046	**1.5941**
R	0.9834	0.9077	0.8687	0.8794	0.8918	0.8796	0.8904	0.9876	0.9254	0.9567	**0.9976**

In summary, for the bridge risk assessment, the comparison results have demonstrated that the proposed activation weight calculation and parameter optimization methods enhance both the interpretability and accuracy of the EBRB system; the accuracy of the EBRB system is superior to some existing studies in bridge risk assessment.

5.4. EBRB systems for continuous and discrete classification datasets

We further demonstrate the effectiveness of the proposed activation weight calculation and parameter optimization methods by using the classification datasets shown in Table 1. First, for the basic parameters, we consider that the attribute weights are 1:

$$\delta_i = 1; i = 1,...,M \tag{41}$$

where M is the number of antecedent attributes in a classification dataset. We assume the number of reference values as three for

each antecedent attribute, and define the utility value of these reference values as:

$$\{u(A_{i,j}); j=1,2,3\} = \left\{ lb_i, \frac{lb_i + ub_i}{2}, ub_i \right\}; i=1,...,M \tag{42}$$

where $u(A_{i,j})$ denotes the utility value of the reference value $A_{i,j}$, lb_i and ub_i are the lower and upper bounds of the ith antecedent attribute, respectively.

Table 6 illustrates the accuracy of initial and trained EBRB systems and the training time of the proposed parameter optimization method. In term of accuracy, the EBRB-New shows significant improvement over five classification datasets than other three EBRB systems, which are based on distance measures. For example, 33.74% and 36.87% improvements are observed for the EBRB-New while testing the Ecoli and Hayes-roth datasets, respectively. However, the proposed parameter optimization method is not always effective for the EBRB systems with distance-based activation weight calculation because of over-fitting. As presented in Table 6, the accuracy of trained EBRB-E is slightly worse than that of initial EBRB-E while testing the Wine and Glass datasets. The similar situations are observed in the Wine dataset for EBRB-S and the Bupa, Glass, and Ecoli datasets for EBRB-H. In term of training time, the comparison result indicates that EBRB-H < EBRB-New < EBRB-E < EBRB-S for the five datasets Bupa, Wine, Glass, Ecoli, and Hayes-roth.

Table 6. Comparisons of accuracy (%) and training time (second) for the EBRB systems

Dataset	EBRB-E			EBRB-S			EBRB-H			EBRB-New		
	Initial	Trained	Time	Initial	Trained	Time	Initial	Trained	Time	Initial	Trained	Time
Bupa	60.58	**64.06**	540.2	58.55	**65.2**1	660.3	**60.8**7	56.23	365.9	57.97	**66.38**	479.6
Wine	**95.51**	94.38	277.8	**96.62**	96.07	332.9	12.92	**13.48**	156.5	97.19	**97.75**	210.3
Glass	**65.42**	61.22	294.6	64.49	**65.89**	369.7	**56.54**	46.73	187.0	55.14	**66.82**	233.6
Ecoli	75.30	**80.65**	551.3	66.96	**79.46**	734.6	**66.96**	50.60	328.9	48.21	**81.85**	501.9
Hayes-roth	65.63	**71.25**	63.0	61.88	**71.88**	92.6	58.13	**70.63**	50.9	41.25	**78.12**	57.5

We further verify the validity of the proposed methods by introducing the conventional classifiers to compare with the EBRB-New, including the k nearest neighbor (KNN), Naïve Bayes, C4.5 decision tree, support vector machine (SVM), and ANN; these classifiers are implemented by using the WEKA software with the method setup that 20% number of training data are set as neighbors for the KNN classifier, 5% as the minimum number of data per leaf for the C4.5 classifier, half of total number of antecedent attributes and classes are set as the number of hidden layers for the ANN.

Table 7 provides a comparison of the results of selected classifiers and the EBRB-New, thus validating that the proposed activation weight calculation and parameter optimization methods produce satisfactory results, that is, for the Wine, Ecoli, and Hayes-roth datasets, the 97.75%, 81.85% and 78.12% accuracies have outperformed all listed classifiers. For the Bupa and Glass datasets, the 66.38% and 66.82% accuracies reach the second-best result in comparison with the conventional classifiers KNN, Naïve Bayes, C4.5, SVM, and ANN.

Table 7. Comparisons of accuracy (%) for EBRB system and conventional classifiers

Dataset	KNN	Naïve Bayes	C4.5	SVM	ANN	EBRB-New
Bupa	61.74	55.07	66.09	59.42	**68.99**	66.38
Wine	97.19	97.19	92.70	43.26	**97.75**	**97.75**
Glass	60.28	50.93	63.08	66.82	**67.29**	66.82
Ecoli	64.88	80.65	77.38	42.55	79.76	**81.85**
Hayes-roth	35.62	75.00	70.63	81.25	68.75	**78.12**

In summary, the comparison results obtained from classification datasets have demonstrated the effectiveness of the proposed activation weight calculation and parameter optimization methods in enhancing the accuracy of the EBRB system. Compared with the conventional classifiers, the enhanced EBRB system shows a absolute competitiveness for handling classification problems.

5.5. Discussion on new activation weight calculation and parameter optimization

In this subsection, we provide some intuitive comparison results based on the case studies of oil pipeline leak detection and bridge risk assessment to demonstrate the advantages of the proposed activation weight calculation and parameter optimization methods.

Regarding interpretability, Table 8 shows the average CE and IE of the EBRB system with the proposed activation weight calculation and the distance-based activation weight calculation for continuous and discrete data. Although the proposed activation weight calculation method effectively prevents counterintuitive and insensitive situations, these situations are observed in the existing distance-based activation weight calculation methods.
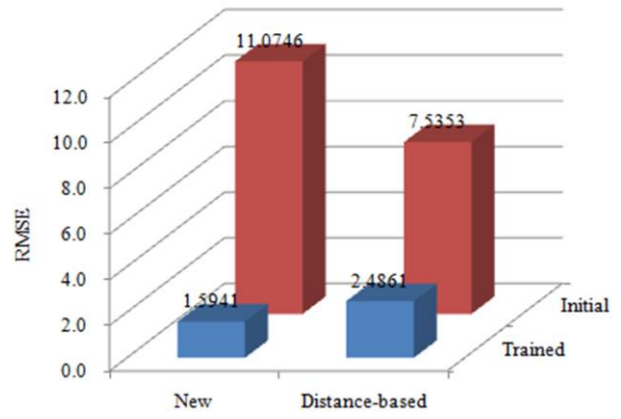
Table 8. Comparisons of activation weight calculation for continuous and discrete data

Data type	Activation weight calculation	Initial parameters		Trained parameters	
		CE	IE	CE	IE
Continuous data	Distance-based	1970	2008	1339	760
	This study	0	0	0	0
Discrete data	Distance-based	306	411	410	169
	This study	0	0	0	0

Regarding accuracy, we use the MSE obtained from continuous data and the RMSE obtained from discrete data for comparing different activation weight calculation methods with and without using the proposed parameter optimization method. Fig. 18 illustrates that the proposed parameter optimization method improves the accuracy of the EBRB system with both the distance-based and proposed activation weight calculation methods. Furthermore, the best accuracy is obtained from the trained EBRB system using the proposed activation weight calculation method.
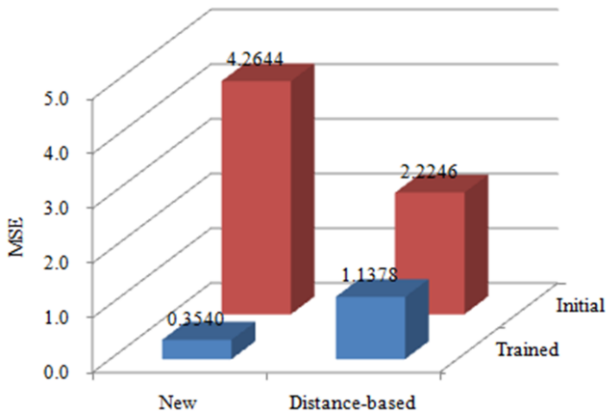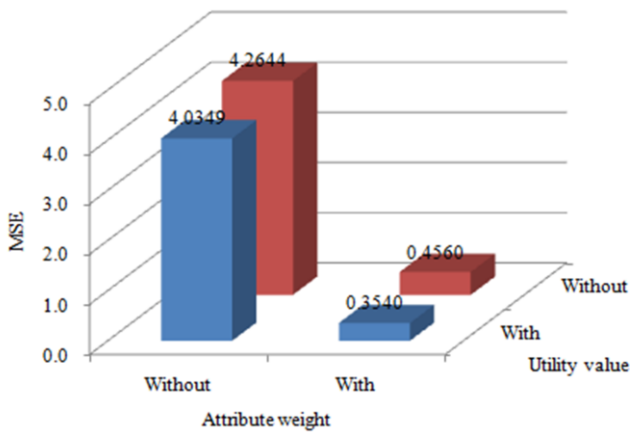
(a) Comparison of MSE on continuous data   (b) Comparison of RMSE on discrete data

Fig. 18. Comparison of EBRB systems for continuous and discrete data

Fig. 19 shows MSE and RMSE of the EBRB system with the new activation weight calculation while including/excluding the optimization of attribute weights and utility values to confer the function of basic parameters in the proposed parameter optimization method. As shown in Fig. 19(b), the improvements of the RMSE are 0.2603, 8.9953, and 9.4805 while the proposed parameter optimization method considers utility values, attribute weights, and considering both, respectively. The optimization of attribute weights has more improvement, but the optimization of both attribute weights and utility values assure the significant improvement in the performance of the trained EBRB system.
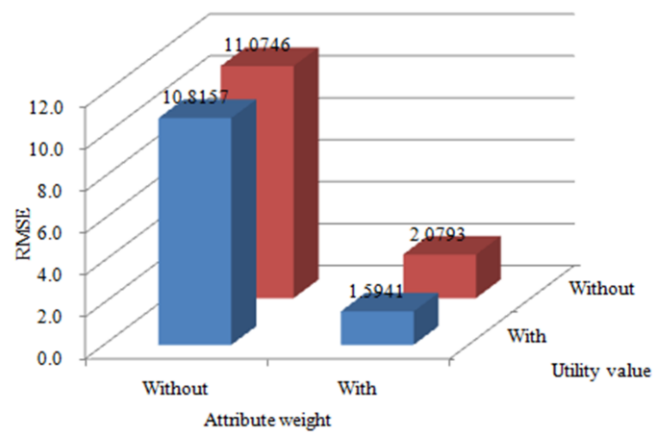
(a) Comparison of MSE on continuous data   (b) Comparison of RMSE on discrete data

Fig. 19. Comparison of EBRB systems with/without attribute weight and utility value

In addition, the importance of parameter optimization needs consideration because the EBRB system is a data-driven model and no study on the EBRB system involves parameter optimization, which is a time-consuming process. However, the following advantages should be emphasized:

First, the existing studies of the rule generation scheme in the EBRB system fail to provide an effective method to determine the basic parameter values of the EBRB system. However, Figs. 18 and 19 clarify that significant improvements are observed between the accuracy of the initial and trained EBRB systems. Hence, the proposed parameter optimization method is essential for the EBRB system.

Second, the previous study [37] has proved that the time complexity of the DE algorithm can be represented as $O(K, S, C)$, in which K is the number of parameters needed to be trained, S is the number of generations, and C is the number of individuals. Hence, for the same value of S and C, the proposed parameter optimization method is relatively efficient for the EBRB system compared with the existing parameter optimization models in the BRB system, because the proposed method just includes a very few number of parameters shown in Table 3.

Third, although the optimization model of the EBRB system is a nonlinear optimization problem that must be solved by using iterative algorithms, it can be implemented before the EBRB system is used to predict or classify given input data. Hence, the proposed parameter optimization method has no effect on the time complexity of the EBRB system in term of prediction and classification purposes.

6. Conclusions and future research

This study focused on an advanced rule-based system, called the EBRB system, which has attracted much attention in the last few years. We investigated the interpretability and accuracy of the EBRB system through sensitivity analyses and proposed a new activation weight calculation method and parameter optimization method for enhancing the EBRB system. The proposed methods were validated using three case studies with different data sets, including discrete data situation, continuous data situation, or combination of both. The detailed contributions are summarized into three aspects:

(1) We summarized six intermediate variables, including the Euclidean, standardized Euclidean, and Hamming-based individual matching degrees and activation weights by reviewing the rule inference scheme of the EBRB system. To investigate the interpretability of these intermediate variables, we performed a sensitivity analysis associated with different test input data, sample input data, rule weights, and attribute weights. Furthermore, we proposed a new activation weight calculation method for enhancing the interpretability of the EBRB system by modifying the individual matching degree calculation and revising the function between activation weights and attribute weights.

(2) We deduced eight intuitive formulas, including calculating activation weights and integrating belief degrees by reviewing the rule generation scheme of the EBRB system. These derivations explained the composition relationship between inference results and basic parameters, including attribute weights and utility values used for antecedent and consequent attributes. Furthermore, we performed a sensitivity analysis based on these derivations and proposed a new parameter optimization method was proposed to enhance the accuracy of the EBRB system by determining the optimal value of these basic parameters.

(3) Two prediction problems, namely oil pipeline leak detection and bridge risk assessment, and five classification problems were used to validate the performance of the proposed activation weight calculation method and the proposed parameter optimization method. The comparison revealed that the proposed activation weight calculation method enhanced the interpretability of the EBRB system, and no counterintuitive and insensitive situations were observed in the rule inference scheme of the EBRB system, and the proposed parameter optimization method enhanced the accuracy

of the EBRB system by outperforming other existing studies.

For future research, the application of the proposed EBRB system and to make it more efficient is a challenge to address the practical problem with big data. Moreover, we recommend availability of the proposed EBRB system for comparing several datasets using different algorithmic approaches.

Appendix A. Formula derivation of activation weights

Assuming that three utility values used for the ith antecedent attribute are shown in Eq. (23), $x_{k,i}$ is the sample input data to generate the kth extended belief rule, x_i is the test input data to activate extended belief rules. Hence, when one antecedent attribute is in extended belief rules, the belief distribution of $x_{k,i}$ and x_i can be simplified as follows:

For the sample input data $x_{k,i}$, the following belief distributions are obtained:

$$S(x_{k,i}) = \{(A_{i,j}, \alpha_{i,j}^k); j=1,...,J_i\} = \begin{cases} \{(A_{i,1}, 1-\dfrac{x_{k,i}}{u(A_{i,2})}),(A_{i,2},\dfrac{x_{k,i}}{u(A_{i,2})}),(A_{i,3},0)\}; & if\ 0 \le x_{k,i} \le u(A_{i,2}) \\ \{(A_{i,1},0),(A_{i,2},\dfrac{1-x_{k,i}}{1-u(A_{i,2})}),(A_{i,3},\dfrac{x_{k,i}-u(A_{i,2})}{1-u(A_{i,2})})\}; & otherwise \end{cases} \tag{A1}$$

For the test input data x_i, the following belief distributions are obtained:

$$S(x_i) = \{(A_{i,j}, \gamma_{i,j}); j=1,...,J_i\} = \begin{cases} \{(A_{i,1},1-x_i),(A_{i,2},1-u(A_{i,2})+x_i),(A_{i,3},x_i)\}; & if\ 0 \le x_i \le u(A_{i,2}) \\ \{(A_{i,1},1-x_i),(A_{i,2},1+u(A_{i,2})-x_i),(A_{i,3},x_i)\}; & otherwise \end{cases} \tag{A2}$$

Next, based on Eqs. (A1) and (A2), the activation weight can be grouped into four situations.

(1) When $u(A_{i,1}) \le x_{k,i} \le u(A_{i,2})$ and $u(A_{i,1}) \le x_i \le u(A_{i,2})$, the activation weight is:

$$\begin{aligned} w_k &= \theta_k \left(\sum_{j=1}^{J_i} \gamma_{i,j} \alpha_{i,j}^k \right)^{1/\delta_i} \\ &= \theta_k \left(\left(1-\frac{x_{k,i}}{u(A_{i,2})}\right) \cdot (1-x_i) + \frac{x_{k,i}}{u(A_{i,2})} \cdot (1-u(A_{i,2})+x_i) + 0 \cdot x_i \right)^{1/\delta_i} \\ &= \theta_k \left(1-\frac{x_{k,i}}{u(A_{i,2})} - x_i + \frac{x_i x_{k,i}}{u(A_{i,2})} + \frac{x_{k,i}}{u(A_{i,2})} - x_{k,i} + \frac{x_i x_{k,i}}{u(A_{i,2})} \right)^{1/\delta_i} \\ &= \theta_k \left(1 - x_i - x_{k,i} + \frac{2x_i x_{k,i}}{u(A_{i,2})} \right)^{1/\delta_i} \end{aligned} \tag{A3}$$

(2) When $u(A_{i,1}) \le x_{k,i} \le u(A_{i,2})$ and $u(A_{i,2}) < x_i \le u(A_{i,3})$, the activation weight is:

$$w_k = \theta_k \left(\sum_{j=1}^{J_i} \gamma_{i,j} \alpha_{i,j}^k \right)^{1/\delta_i}$$

$$= \theta_k \left(\left(1 - \frac{x_{k,i}}{u(A_{i,2})} \right) \cdot \left(1 - x_i \right) + \frac{x_{k,i}}{u(A_{i,2})} \cdot \left(1 + u(A_{i,2}) - x_i \right) + 0 \cdot x_i \right)^{1/\delta_i}$$

$$= \theta_k \left(1 - \frac{x_{k,i}}{u(A_{i,2})} - x_i + \frac{x_i x_{k,i}}{u(A_{i,2})} + \frac{x_{k,i}}{u(A_{i,2})} + x_{k,i} - \frac{x_i x_{k,i}}{u(A_{i,2})} \right)^{1/\delta_i}$$

$$= \theta_k \left(1 - x_i + x_{k,i} \right)^{1/\delta_i}$$

(A4)

(3) When $u(A_{i,2}) < x_{k,i} \le u(A_{i,3})$ and $u(A_{i,1}) \le x_i \le u(A_{i,2})$, the activation weight is:

$$w_k = \theta_k \left(\sum_{j=1}^{J_i} \gamma_{i,j} \alpha_{i,j}^k \right)^{1/\delta_i}$$

$$= \theta_k \left(0 \cdot \left(1 - x_i \right) + \frac{1 - x_{k,i}}{1 - u(A_{i,2})} \cdot \left(1 - u(A_{i,2}) + x_i \right) + \frac{x_{k,i} - u(A_{i,2})}{1 - u(A_{i,2})} \cdot x_i \right)^{1/\delta_i}$$

$$= \theta_k \left(\frac{1 - x_{k,i} - u(A_{i,2}) + x_{k,i} u(A_{i,2}) + x_i - x_{k,i} x_i + x_{k,i} x_i - u(A_{i,2}) x_i}{1 - u(A_{i,2})} \right)^{1/\delta_i}$$

$$= \theta_k \left(1 + x_i - x_{k,i} \right)^{1/\delta_i}$$

(A5)

(4) When $u(A_{i,2}) < x_{k,i} \le u(A_{i,3})$ and $u(A_{i,2}) < x_i \le u(A_{i,3})$, the activation weight is:

$$w_k = \theta_k \left(\sum_{j=1}^{J_i} \gamma_{i,j} \alpha_{i,j}^k \right)^{1/\delta_i}$$

$$= \theta_k \left(0 \cdot \left(1 - x_i \right) + \frac{1 - x_{k,i}}{1 - u(A_{i,2})} \cdot \left(1 + u(A_{i,2}) - x_i \right) + \frac{x_{k,i} - u(A_{i,2})}{1 - u(A_{i,2})} \cdot x_i \right)^{1/\delta_i}$$

$$= \theta_k \left(\frac{1 - x_{k,i} + u(A_{i,2}) - x_{k,i} u(A_{i,2}) - x_i + x_{k,i} x_i + x_{k,i} x_i - u(A_{i,2}) x_i}{1 - u(A_{i,2})} \right)^{1/\delta_i}$$

$$= \theta_k \left(1 + x_{k,i} + x_i + \frac{2 x_i x_{k,i} - 2 x_{k,i} - 2 x_i}{1 - u(A_{i,2})} \right)^{1/\delta_i}$$

(A6)

Appendix B. Formula derivation of integrated belief degrees

Assuming three utility values used for the consequent attribute are showed in Eq. (25), y_k is the sample output data to generate the kth extended belief rule, y_l is the sample output data to generate the lth extended belief rule. Hence, based on the ER algorithm, the integrated belief degree without normalization is expressed as follows:

$$\beta_n = \prod_{k=1}^{2} \left(w_k \beta_n^k + 1 - w_k \sum_{i=1}^{N} \beta_i^k \right) - \prod_{k=1}^{2} \left(1 - w_k \sum_{i=1}^{N} \beta_i^k \right)$$

$$= \left(w_k \beta_n^k + 1 - w_k \right) \left(w_l \beta_n^l + 1 - w_l \right) - \left(1 - w_k \right) \left(1 - w_l \right)$$

$$= \left(w_k \beta_n^k + w_l \right) \left(w_l \beta_n^l + w_k \right) - w_l w_k$$

$$= w_k w_l \beta_n^k \beta_n^l + w_k w_k \beta_n^k + w_l w_l \beta_n^l + w_l w_k - w_l w_k$$

$$= w_k w_l \beta_n^k \beta_n^l + w_k w_k \beta_n^k + w_l w_l \beta_n^l$$

(B1)

In addition, for the sample output data y_k, the following belief distributions are obtained:

$$S(y_k) = \{ (D_n, \beta_n^k); n = 1, \ldots, N \} = \begin{cases} \{ (D_1, 1 - \frac{y_k}{u(D_2)}), (D_1, \frac{y_k}{u(D_2)}), (D_1, 0) \}; if \ 0 \le y_k \le u(D_2) \\ \{ (D_1, 0), (D_1, \frac{1 - y_k}{1 - u(D_2)}), (D_1, \frac{y_k - u(D_2)}{1 - u(D_2)}) \}; otherwise \end{cases}$$

(B2)

Hence, based on Eqs. (B1) and (B2), the integrated belief degree can be grouped into four situations.

(1) When $u(D_1) \leq y_k \leq u(D_2)$ and $u(D_1) \leq y_l \leq u(D_2)$, the integrated belief degree of D_2 is:

$$
\begin{aligned}
\beta_2 &= w_k w_l \beta_2^k \beta_2^l + w_k w_k \beta_2^k + w_l w_l \beta_2^l \\
&= w_k w_l \frac{y_k y_l}{u(D_2)u(D_2)} + w_k w_k \frac{y_k}{u(D_2)} + w_l w_l \frac{y_l}{u(D_2)} \\
&= \frac{w_k w_l y_k y_l}{\left(u(D_2)\right)^2} + \frac{(w_k)^2 y_k + (w_l)^2 y_l}{u(D_2)}
\end{aligned}
\tag{B3}
$$

(2) When $u(D_1) \leq y_k \leq u(D_2)$ and $u(D_2) < y_l \leq u(D_3)$, the integrated belief degree of D_2 is:

$$
\begin{aligned}
\beta_2 &= w_k w_l \beta_2^k \beta_2^l + w_k w_k \beta_2^k + w_l w_l \beta_2^l \\
&= w_k w_l \frac{y_k(1-y_l)}{u(D_2)\left(1-u(D_2)\right)} + w_k w_k \frac{y_k}{u(D_2)} + w_l w_l \frac{1-y_l}{1-u(D_2)} \\
&= \frac{w_k w_l y_k(1-y_l)}{u(D_2)\left(1-u(D_2)\right)} + \frac{(w_k)^2 y_k}{u(D_2)} + \frac{(w_l)^2(1-y_l)}{1-u(D_2)}
\end{aligned}
\tag{B4}
$$

(3) When $u(D_2) < y_k \leq u(D_3)$ and $u(D_1) \leq y_l \leq u(D_2)$, the integrated belief degree of D_2 is:

$$
\begin{aligned}
\beta_2 &= w_k w_l \beta_2^k \beta_2^l + w_k w_k \beta_2^k + w_l w_l \beta_2^l \\
&= w_k w_l \frac{(1-y_k)y_l}{\left(1-u(D_2)\right)u(D_2)} + w_k w_k \frac{1-y_k}{1-u(D_2)} + w_l w_l \frac{y_l}{u(D_2)} \\
&= \frac{w_k w_l(1-y_k)y_l}{\left(1-u(D_2)\right)u(D_2)} + \frac{(w_k)^2(1-y_k)}{1-u(D_2)} + \frac{(w_l)^2 y_l}{u(D_2)}
\end{aligned}
\tag{B5}
$$

(4) When $u(D_2) < y_k \leq u(D_3)$ and $u(D_2) < y_l \leq u(D_3)$, the integrated belief degree of D_2 is:

$$
\begin{aligned}
\beta_2 &= w_k w_l \beta_2^k \beta_2^l + w_k w_k \beta_2^k + w_l w_l \beta_2^l \\
&= w_k w_l \frac{(1-y_k)(1-y_l)}{\left(1-u(D_2)\right)\left(1-u(D_2)\right)} + w_k w_k \frac{1-y_k}{1-u(D_2)} + w_l w_l \frac{1-y_l}{1-u(D_2)} \\
&= \frac{w_k w_l(1-y_k)(1-y_l)}{\left(1-u(D_2)\right)^2} + \frac{(w_k)^2(1-y_k) + (w_l)^2(1-y_l)}{1-u(D_2)}
\end{aligned}
\tag{B6}
$$

References

[1] Alcala-Fdez J, Fernández A, Luengo J, Derrac J, García S, Sanchez L, Herrera F (2011) KEEL Data-Mining Software Tool-Data Set Repository, Integration of Algorithms and Experimental Analysis Framework, Journal of Multiple-Valued Logic and Soft Computing 17: 255-287.

[2] British Highways Agency (2004) Value Management of the Structures Renewal Programme (Version 2.2).

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

[4] Chang LL, Zhou ZJ, Liao TJ, Chen YW, Yu H, Yang LH (2017) Belief rule base structure and parameter joint learning under disjunctive assumption for nonlinear complex system modeling. IEEE Transaction on Systems, Man, and Cybernetics-Part A: Systems, In Press.

[5] Chang LL, Zhou ZJ, You Y, Yang LH, Zhou ZG (2016) Belief rule based expert system for classification problems with new

rule activation and weight calculation procedures. Information Sciences 336: 75-91.

[6] Chen Y, Chen YW, Xu XB, Pan CC, Yang JB, Yang GK (2015) A data-driven approximate causal inference model using the evidential reasoning rule. Knowledge-Based Systems 88: 264-272.

[7] Chen SM, Chien CY (2011) Parallelized genetic ant colony systems for solving the traveling salesman problem. Expert Systems with Applications 38(4): 3873-3883.

[8] Chen SM, Chung NY (2006) Forecasting enrollments of students by using fuzzy time series and genetic algorithms. International Journal of Information and Management Sciences 17(3): 1-17.

[9] Chen SM, Huang CM (2003) Generating weighted fuzzy rules from relational database systems for estimating null values using genetic algorithms. IEEE Transactions on Fuzzy Systems 11(4): 495-506.

[10] Chen YW, Yang JB, Xu DL, Zhou ZJ, Tang DW (2011) Inference analysis and adaptive training for belief rule based system. Expert System with Applications 38:12845-12860.

[11] Espinilla M, Medina J, Calzada A, Liu J, Martinez L, Nugent C (2017) Optimizing the configuration of an heterogeneous architecture of sensors for activity recognition, using the extended belief rule-based inference methodology. Microprocessors and Microsystems 52: 381-390.

[12] Jorgensen SB, Hangos KM (1995) Grey box modelling for control: qualitative models as a unifying framework. International Journal of Adaptive Control & Signal Processing 9(6): 547-562.

[13] Khan A, Jaffar MA, Shao L (2015) A modified adaptive differential evolution algorithm for color image segmentation. Knowledge and Information Systems 43(3): 583-797.

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

[15] Nikolic V, Mitic VV, Kocic L, Petkovic D (2017) Wind speed parameters sensitivity analysis based on fractals and neuro-fuzzy selection technique. Knowledge and Information Systems 52: 255-265.

[16] Pannell DJ (1997) Sensitivity Analysis of Normative Economic Models: Theoretical Framework and Practical Strategies. Agricultural Economics 16(2): 139-152.

[17] Price K (1997) Differential evolution vs. the functions of the 2nd ICEO, in: Proceedings of the 1997 IEEE International Conference of Evolutionary Computation pp.153-157.

[18] Price K, Storn RM, Lampinen JA (2005) Differential Evolution: A Practical Approach to Global Optimization (Natural Computing Series). Springer-Verlag New York, Inc.

[19] Rey MI, Galende M, Fuente MJ, Sainz-Palmero CI (2017) Multi-objective based Fuzzy Rule Based Systems (FRBSs) for trade-off improvement in accuracy and interpretability: A rule relevance point of view. Knowledge-Based Systems 127, 67-84.

[20] Soua B, Borgi A, Tagina M (2013) An ensemble method for fuzzy rule-based classification systems. Knowledge and Information Systems 36: 385-410.

[21] Storm R, Price K (1996) Minimizing the real functions of the ICEC 1996 contest by differential evolution, in: Proceedings of the 1996 IEEE International Converence of Evolutionary Computation pp. 842-844.

[22] Tsai PW, Pan JS, Chen SM, Liao BY (2012) Enhanced parallel cat swarm optimization based on the Taguchi method. Expert Systems with Applications 39(7): 6309-6319.

[23] Tsai PW, Pan JS, Chen SM, Liao BY, Hao SP (2008) Parallel cat swarm optimization. Proceeding of the 2008 International Conference on Machine Learning and Cybernetics, Kunming, China, 6: 3328-3333.

[24] Wang YM, Elhag TMS (2007) A comparison of neural network, evidential reasoning and multiple regression analysis in modelling bridge risks. Expert Systems with Applications 32(2): 336-348.

[25] Wang YM, Elhag TMS (2008) An adaptive neuro-fuzzy inference system for bridge risk assessment, Expert Systems with Applications 34(4): 3099-3106.

[26] Wang YM, Yang JB, Xu DL (2006) Environmental impact assessment using the evidential reasoning approach. European Journal of Operational Research 174(3): 1885-1913.

[27] Wang YM, Yang LH, Fu YG, Chang LL, Chin KS (2016) Dynamic rule adjustment approach for optimizing belief rule-base expert system. Knowledge-Based Systems 96: 40-60.

[28] Xu DL, Liu J, Yang JB, Liu GP, Wang J, Jenkinson I, Ren J (2007) Inference and learning methodology of belief-rule-based expert system for pipeline leak detection. Expert Systems with Applications 32(1): 103-113.

[29] Yang JB (2001) Rule and utility based evidential reasoning approach for multiattribute decision analysis under uncertainties. European Journal of Operational Research 131(1): 31-61.

[30] Yang JB, Liu J, Wang J, Sii HS, Wang HW (2006) Belief rule-base inference methodology using the evidential reasoning approach - RIMER. IEEE Transactions on Systems Man and Cybernetics-Part A: Systems and Humans 36(2): 266-285.

[31] Yang JB, Liu J, Xu DL, Wang J, Wang HW (2007) Optimization models for training belief-rule-based systems. IEEE Transactions on Systems Man and Cybernetics-Part A: Systems and Humans 37(4): 569-585.

[32] Yang LH, Wang YM, Lan YX, Chen L, Fu YG (2017) A data envelopment analysis (DEA)-based method for rule reduction in extended belief-rule-based systems. Knowledge-Based Systems 123: 174-187.

[33] Yang LH, Wang YM, Su Q, Fu YG, Chin KS (2016) Multi-attribute search framework for optimizing extended belief rule-based systems. Information Sciences 370-371: 159-183.

[34] Yang LH, Wang YM, Chang LL, Fu YG (2017) A disjunctive belief rule-based expert system for bridge risk assessment with dynamic parameter optimization model. Computers & Industrial Engineering 113: 459-474.

[35] Zadeh LA (1999) Fuzzy logic and the calculi of fuzzy rules and fuzzy graphs, and fuzzy probabilities. Computers & Mathematics with Applications 37(11-12): 35.

[36] Zhou ZJ, Hu CH, Yang JB, Xu DL, Chen MY, Zhou DH (2010) A sequential learning algorithm for online constructing belief rule based systems. Expert Systems with Applications 37(2): 1790-1799.

[37] Zielinski K, Peters D, Laur R (2005) Run time analysis regarding stopping criteria for differential evolution and particle

swarm optimization, in: Proceeding of the 1st International Conference on Experiments/Process/System/Modelling/Simulation/Optimization.