Ensemble extended belief rule-based systems with different similarity measures for classification problems

Fei Gao[a,*], Weikai He[a], Wenhao Bi[b]

[a]*School of Aeronautics, Shandong Jiaotong University, Jinan, China*
[b]*School of Aeronautics, Northwestern Polytechnical University, Xi'an, China*

Abstract

The extended belief rule-based (EBRB) system is shown to have the potential to handle both quantitative and qualitative information under uncertainty, and it has been used as an effective tool for decision support and classification problems. However, despite these advances, several drawbacks have emerged recently, and the most significant one is caused by its similarity measure using Euclidean distance, which could lead to counterintuitive individual matching degrees, while other widely used similarity measures have not been studied for their application in the EBRB system. To this end, seventeen similarity measures are investigated and applied in the EBRB system in this paper, and based on the analysis, an ensemble method for EBRB system with different similarity measures is proposed. Firstly, the problem of the similarity measure of the conventional EBRB system is investigated. Then, a variety of similarity measures are analyzed and their application in the EBRB system is studied. Next, the ensemble method for EBRB systems with different similarity measures is proposed, which consists of two parts, the adaptive weight learning method for determining the weight of each EBRB system with different similarity measures, and the evidential reasoning (ER)-based combination method for combining the inferential results of different methods. Finally, 25 classification datasets are studied to test the performance of EBRB systems with different similarity measures as well as the proposed method, and the results are compared with existing works. The comparison results show that the proposed method could not only achieve better results than any other EBRB systems with different similarity measures, but

*Corresponding author
Email address: `gaofei1995@hotmail.com` (Fei Gao)

also outperform other conventional classifiers on some classification datasets, especially small-scale datasets.

Keywords: Extended belief rule-based system, Similarity measure, Classification

1. Introduction

Among various knowledge representation schemes, the rule-based system has been significant for its ability to model various kinds of human cognition under the same framework using simple "IF-THEN" rules [1], and it has become one of the fastest-growing methods in the field of artificial intelligence and decision support system [2–4]. In conventional rule-based system, simple IF-THEN rules such as "IF failure rate is high, THEN risk is high" are used to construct the rule base, where both the antecedent and consequent terms are believed to be 100% certain, however, such strict knowledge representation scheme may not be sufficient in expressing information with uncertainty. To this end, many novel rule-based systems have been developed [5–9]. Among these methods, the belief rule-based (BRB) system proposed by Yang et al. [10] has received extensive attention. Based on Dempster-Shafer (D-S) theory of evidence, fuzzy set theory and rule-based system, the BRB system uses belief structure in the consequent terms to capture uncertain information, and provides a more flexible way to represent various kinds of knowledge under uncertainty, such as fuzziness, ignorance, and incompleteness. However, the uncertainty in the antecedent terms has yet to be addressed, to this end, Liu et al. [11] proposed a more general belief rule-based system, where belief structures are embedded in the antecedent attributes of each rule, called the extended belief rule-based (EBRB) system. With belief structures embedded in both the antecedent and consequent terms of each rule, the EBRB system can more effectively deal with different kinds of uncertainty and present more accurate results. Compared to conventional rule-based system and BRB system, the EBRB system has the following advantages [12–14]:

(1) Belief structures are embedded in both the consequents and antecedent attributes of the extended belief rule. Hence, both conventional IF-THEN rules and belief rules can be regarded as special cases of extended belief rules.

(2) The EBRB system can be either a knowledge-driven or a data-driven, or combined decision model, which makes it possible to directly generate rules from input-output data pairs.

Therefore, the EBRB system has been used in various decision-making problems, such as health estimation [15], environmental mentoring [16, 17], activity recognition [18, 19], and many others [20–24]. However, there are several drawbacks in the process of determining activated rules and their activation weight in the conventional EBRB system, and several studies have been conducted tackling these problems [25–28]. For instance, Zhang et al. [29] proposed a new rule reduction method based on DBSCAN algorithm for EBRB system, where similar rules are searched and fused to reduce the size of the EBRB and remove the impact of noise and redundant rules. Yang et al. [30] proposed a new activation rule determination and weight calculation method, where the activation region of extended belief rules is constructed to remove the impact of inconsistent rules and improve the conventional EBRB system. Yang et al. [31] introduced data envelopment analysis (DEA) to the EBRB system to evaluate the efficiency of each rule for rule reduction, and used the classic CCR (Charnes, Cooper, and Rhodes) model to calculate the efficiency value of the extended belief rule and achieve the compact structure of an EBRB. Zhu et al. [32] proposed a minimum centre distance rule activation (MCDRA) method, which requires no subjective information or iteration procedure while unrelated simples are eliminated and related simples are selected and activated.

However, despite these advances, little attention has been paid to the activation weight calculation formula itself, where the similarity between the input and each rule as well as the weight of each rule are used to determine to what extent each rule is activated. In conventional EBRB systems, the similarity between the extended belief rule and the input is obtained by simply calculating the Euclidean distance between two belief structures, which could limit its application in some situations, as different problems may require different methods. The similarity calculation requires further research because: (1) Due to the limitation of the Euclidean distance, the calculated activated weights of the rules could be inconsistent; (2) There has been numerous research on the similarity measure for belief functions, such as Jousselme distance, Chebyshev distance, and Tanimoto similarity, which could be used to calculate the similarity between the input and the extended belief rules.

As different similarity measures have different applicability and effectiveness, it is possible to develop a more effective measure that could combine the results of different similarity measures. Therefore, motivated by these challenges and advancements, seventeen different similarity measures are investigated in this paper for their potential application in the EBRB system. Furthermore, on the basis of that, an adaptive weight learning method for EBRB system with different similarity measures is proposed. Firstly, seventeen different measures are investigated, including Jousselme distance, Manhattan distance, Intersection similarity, and many others. Secondly, an adaptive weight learning method for similarity measures based on the differential evolution (DE) algorithm is introduced, and an evidential reasoning (ER)-based combination method for EBRB systems with different similarity measures is subsequently proposed. Finally, an experiment on classification problems is conducted to illustrate the effectiveness and efficiency of the proposed method.

The remainder of this paper is organized as follows. Section 2 briefly describes the basics of extended belief rule-based system. Section 3 reviews the problem in current similarity measure of the EBRB system, and investigates different kinds of similarity measures. The adaptive weight learning method for EBRB systems with different similarity measures is proposed in Section 4. An experiment on classification problem is illustrated in Section 5 to demonstrate the effectiveness of the proposed method, and Section 6 concludes the paper.

2. Preliminaries

Based on the belief rule-based system, Liu et al. [11] proposed the extended belief rule-based system, where the belief degrees are embedded in both the consequents and the antecedent attributes of each rule. The extended belief rule-based system is composed of two parts, the extended belief rule base (EBRB) which stores various kinds of information under uncertainty, and the ER-based inference approach for aggregating activated extended belief rules.

2.1. Extended belief rule base

In the EBRB system, the EBRB is used to store various kinds of information under uncertainty, such as qualitative and quantitative, complete and incomplete, linguistic and numerical information. An EBRB is comprised of a series of extended belief rules, and the kth rule in the EBRB is expressed as:

$$
\begin{aligned}
R_k :& \text{IF } U_1 \text{ is } \{(A^k_{1,j}, \alpha^k_{1,j}), j = 1, \ldots, J_1\} \wedge \cdots \wedge \\
& U_M \text{ is } \{(A^k_{M,j}, \alpha^k_{M,j}), j = 1, \ldots, J_M\}, \\
& \text{THEN } D \text{ is } \{(D_n, \beta^k_n), n = 1, \ldots, N\}, \\
& \text{with rule weight } \theta_k \text{ and attribute weights } \{\delta_1, \ldots, \delta_M\}
\end{aligned}
\tag{1}
$$

where $\alpha^k_{i,j}$ is the belief degree to which U_i is evaluated to be the referential value $A^k_{i,j}$ in the kth rule with $0 \leq \alpha^k_{i,j} \leq 1$ and $\sum_{j=1}^{J_i} \alpha^k_{i,j} \leq 1$ $(i = 1, \ldots, M)$. β^k_n is the belief degree to which D is evaluated to be the referential value D_n in the kth rule with $0 \leq \beta^k_n \leq 1$ and $\sum_{n=1}^{N} \beta^k_n \leq 1$. The consequent is complete when $\sum_{n=1}^{N} \beta^k_n = 1$, otherwise, it is incomplete. θ_k $(0 \leq \theta_k \leq 1)$ and δ_i $(0 \leq \delta_i \leq 1)$ denotes the rule weight and attribute weight respectively. $k = 1, \ldots, L$, and L is the number of the extended belief rules in the EBRB.

2.2. Inference approach using the ER algorithm

Based on the ER algorithm, the inference process of the EBRB system comprises the calculation of the individual matching degree, calculation of the activation weight and combination of the activated rules. Firstly, suppose x_i represents the input of the ith antecedent attribute U_i, then it can be transformed into belief structure using utility-based information transformation technique as follows [10]:

$$
S(x_i) = \{(A_{i,j}, \alpha_{i,j}), j = 1, \ldots, J_i\} \tag{2}
$$

with

$$
\begin{aligned}
\alpha_{i,j} =& \frac{u(A_{i,j+1}) - x_i}{u(A_{i,j+1}) - u(A_{i,j})}, \alpha_{i,j+1} = 1 - \alpha_{i,j}, \\
& if \;\; u(A_{i,j}) \leq x_i \leq u(A_{i,j+1}) \\
\alpha_{i,m} =& 0, if \;\; m = \{1, \ldots, J_i\} \; and \; m \neq j, j+1
\end{aligned}
\tag{3}
$$

where $A_{i,j}$ represents the jth referential value in the ith antecedent attribute, $\alpha_{i,j}$ is the belief degree to which the input x_i is assessed to the referential value $A_{i,j}$, $u(A_{ij})$ is the utility value of $A_{i,j}$.

2.2.1. Individual matching degree

Suppose the input is obtained as $S(x_i) = \{(A_{i,j}, \alpha_{i,j})\}$, then the individual matching degree S^k_i of x_i to U_i of the kth rule, which represents how *close* these two belief structures are, is defined as:

$$
S_{k,i} = 1 - d_{k,i} \tag{4}
$$

with

$$d_{k,i} = d_k(x_i, U_i) = \sqrt{\sum_{j=1}^{J_i} (\alpha_{i,j} - \alpha_{i,j}^k)^2} \tag{5}$$

2.2.2. Activation weight

Once the individual matching degree of each rule is obtained, the activation weight of each rule can be calculated, and the activation weight of the kth extended belief rule is calculated as:

$$\omega_k = \frac{\theta_k \prod_{i=1}^{M} (S_i^k)^{\bar{\delta}_i}}{\sum_{l=1}^{L} (\theta_l \prod_{i=1}^{M} (S_i^l)^{\bar{\delta}_i})} \tag{6}$$

with

$$\bar{\delta}_i = \frac{\delta_i}{\max_{i=1,\ldots,M}\{\delta_i\}} \tag{7}$$

where θ_k represents the rule weight of the kth rule, δ_i represents the weight of the ith attribute in the kth rule. It should be noted that

$$0 \leq \omega_k \leq 1 \ (k = 1, 2, \ldots, L), \quad \sum_{k=1}^{L} \omega_k = 1 \tag{8}$$

with $\omega_k = 0$ denoting that the kth rule is not activated.

2.2.3. Rule aggregation

After calculating the activation rule weight, the ER algorithm is applied for aggregating activated extend belief rules. Firstly, the consequent of each activated rule is transformed into the basic probability mass:

$$\begin{aligned} m_{n,k} &= \omega_k \beta_{n,k} \\ m_{D,k} &= 1 - \omega_k \sum_{n=1}^{N} \beta_{n,k} \\ \bar{m}_{D,k} &= 1 - \omega_k \\ \tilde{m}_{D,k} &= 1 - \sum_{n=1}^{N} \beta_{n,k} \end{aligned} \tag{9}$$

where $m_{n,k}$ represents the basic probability mass assigned to the nth grade, and $m_{D,k}$ represents the basic probability mass unassigned to any grades.

$m_{D,k} = \bar{m}_{D,k} + \tilde{m}_{D,k}$, with $\bar{m}_{D,k}$ representing the incompleteness caused by the activation weight of the kth rule and $\tilde{m}_{D,k}$ representing the incomplete of the kth rule.

Then, the analytical ER process is introduced as follows to aggregate L rules:

$$\begin{aligned} m_n &= \mu \left[\prod_{k=1}^{L} (m_{n,k} + \bar{m}_{D,k} + \tilde{m}_{D,k}) - \prod_{k=1}^{L} (\bar{m}_{D,k} + \tilde{m}_{D,k}) \right] \\ \tilde{m}_D &= \mu \left[\prod_{k=1}^{K} (\bar{m}_{D,k} + \tilde{m}_{D,k}) - \prod_{k=1}^{L} \bar{m}_{D,k} \right] \\ \bar{m}_D &= \mu \left[\prod_{k=1}^{L} \bar{m}_{D,k} \right] \end{aligned} \tag{10}$$

with

$$\mu = \left[\sum_{n=1}^{N} \prod_{k=1}^{L} (m_{n,k} + \bar{m}_{D,k} + \tilde{m}_{D,k}) - (N-1) \prod_{k=1}^{L} (\bar{m}_{D,k} + \tilde{m}_{D,k}) \right]^{-1} \tag{11}$$

Next, the belief degree of each grade can be calculated as:

$$\begin{aligned} \beta_n &= \frac{m_n}{1 - \bar{m}_D} \\ \beta_D &= \frac{m_D}{1 - \bar{m}_D} \end{aligned} \tag{12}$$

where β_n represents the belief degree of the nth grade, and β_D represents the belief degree of the incomplete information. Hence, the aggregated result can be described as:

$$S = \{(D_n, \beta_n), n = 1, 2, \ldots, N\} \tag{13}$$

2.2.4. Inferential result calculation

Finally, the inferential result of the EBRB system could be obtained based on the combined belief structure. For regression problem, suppose $u(D_n)$ is the utility value of the nth consequent D_n, then the inferential result can be obtained as:

$$f = \sum_{n=1}^{N} u(D_n)\beta_n + \frac{u(D_1) + u(D_N)}{2} \left(1 - \sum_{n=1}^{N} \beta_n \right) \tag{14}$$

For classification problem, suppose D_n represents the nth class, and the inferential result can be expressed as follows:

$$f = D_n,\ n = \text{argmax}_{n=1,\dots,N}(\beta_n) \tag{15}$$

3. Similarity measures

In this section, the limitation of current similarity measure of the EBRB system is analyzed at first. Then, seventeen other similarity measures for EBRB systems are introduced.

Let m_1 and m_2 be two evidences, where $m_{1,j}$ and $m_{2,j}$ denote the basic probability mass assigned to the jth grade in m_1 and m_2, respectively. The definitions of the similarity measures are given as follows.

3.1. Similarity measure of the EBRB system

In the inference approach of the EBRB system, similarity measure is used to calculate the individual matching degree between the input and the extended belief rules, as shown in Eq. (4). However, there are several problems in the similarity measure. According to Eqs. (4)-(6), the calculation formula of individual matching degree can be deduced as follows:

$$\begin{aligned} S_i^k &= 1 - \sqrt{\sum_{j=1}^{J_i} (\alpha_{i,j} - \alpha_{i,j}^k)^2} \geq 1 - \sqrt{\sum_{j=1}^{J_i} |\alpha_{i,j} - \alpha_{i,j}^k|} \\ &\geq 1 - \sqrt{\sum_{j=1}^{J_i} (|\alpha_{i,j}| + |\alpha_{i,j}^k|)} = 1 - \sqrt{\sum_{j=1}^{J_i} \alpha_{i,j} + \sum_{j=1}^{J_i} \alpha_{i,j}^k} \\ &\geq 1 - \sqrt{2} \end{aligned} \tag{16}$$

It is obvious from Eq. (16) that a necessary normalization is neglected in Eq. (4), and the lower bound of individual matching degrees is $1-\sqrt{2}$, which could lead to counterintuitive individual matching degrees. To this end, there have been several modifications. For example, Zhang et al. [29] limited the values of individual matching degrees by defining $S_i^k = 0$ for cases where $d_i^k > 1$. Yang et al. [31] used the Jousselme distance instead of the Euclidean distance to calculate the distance between the two belief structures. However, though these modifications have provided promising results, there lack of comprehensive investigation of possible similarity measures of the EBRB system to enhance its performance. As the belief structure used in the EBRB system can be regarded as a piece of evidence with single hypotheses, evidence similarity measures could be applied in the EBRB system. Therefore, several kinds of similarity measures are investigated for the application in the EBRB system.

3.2. Similarity measures

In this subsection, 17 similarity measures are introduced, such as Manhattan distance, Chebyshev distance, and Lorentzian distance. It should be noted that since distance measure and similarity measure are interchangeable [33], both distance measures and similarity measures are introduced in this subsection.

3.2.1. Minkowski family

The Minkowski family of distance has been widely applied in evidence theory, which includes some of the most popular similarity measures such as Euclidean distance, Manttan distance, and Chebyshev distance. The Minkowski family of distance between two belief structures can be expressed under the following general form [34]:

$$d_w(m_1, m_2) = \left(\left[(Um_1 - Um_2)^{\frac{p}{2}} \right]' \left[(Um_1 - Um_2)^{\frac{p}{2}} \right] \right)^{\frac{1}{p}} \tag{17}$$

where U is the upper triangular matrix of Cholesky decomposition of the matrix W, that is $W = U'U$, and p is an integer greater than 1. Typical cases are $p = 1$, $p = 2$ and $p = \infty$, which lead to the Manhattan, Euclidean and Chebyshev distances, respectively.

When $p = 1$, Eq. (17) becomes the Manhattan distance as follows [35]:

$$d_{Man}(m_1, m_2) = \sum_{j=1}^{J} |m_{1,j} - m_{2,j}| \tag{18}$$

When $p = 2$, Eq. (17) becomes the Euclidean distance, as shown in Eq. (6), and it has been predominantly used as the distance measure in the EBRB system.

When $p = \infty$, it becomes Chebyshev distance, which relies on a max operator, and it can be expressed as [36]:

$$d_{Cheb}(m_1, m_2) = \max_j |m_{1,j} - m_{2,j}| \tag{19}$$

3.2.2. Manhattan family

There have been several new modifications based on the Manhattan distance, and most of them can be used to measure the distance between the input and the extended belief rules in the EBRB system. As these distance measures facilitate the Manhattan distance, they may be able to reach better results. These distance measures include Sorensen distance and Lorentzian distance.

Sorensen distance uses the sum of m_1 and m_2 as the denominator, and can be expressed as [37]:

$$d_{Sor}(m_1, m_2) = \frac{\sum_{j=1}^{J} |m_{1,j} - m_{2,j}|}{\sum_{j=1}^{J} (m_{1,j} + m_{2,j})} \tag{20}$$

Obviously, since both m_1 and m_2 are complete belief structures, it is the Manhattan distance divided by 2.

Lorentzian distance applies both the absolute difference and the natural logarithm, and is expressed as follows [38]:

$$d_{Lor}(m_1, m_2) = \sum_{j=1}^{J} \ln(1 + |m_{1,j} - m_{2,j}|) \tag{21}$$

where 1 is added to maintain non-negativity while avoiding the log of 0.

3.2.3. Intersection similarity

The intersection similarity uses the minimum value to calculate the similarity between two probability density functions, as belief structures can be regarded as evidence with single hypotheses, the intersection similarity could be applied as the similarity measure. It should be noted that the Intersection similarity directly calculates the similarity instead of distance, and it is calculated using the sum of minimum belief degrees of m_1 and m_2 as follows [38]:

$$S_{IS}(m_1, m_2) = \sum_{j=1}^{J} \min(m_{1,j}, m_{2,j}) \tag{22}$$

3.2.4. Inner family

The Inner family of similarity measures is frequently used in the fields of information retrieval and biological taxonomy for the binary feature vector comparison, and can be expressed as the following general form [34]:

$$S_W(m_1, m_2) = m_1' W m_2 \tag{23}$$

where the weighting matrix W can be qualified as a similarity matrix, and different similarity measures can be obtained. These similarity measures include Inner product similarity, Tanimoto similarity, Cosine similarity, and Dice similarity.

Inner product similarity directly calculates the product of two belief structures, and is expressed as follows [38]:

$$S_{IP}(m_1, m_2) = \sum_{j=1}^{J} m_{1,j} m_{2,j} \tag{24}$$

Tanimoto similarity is calculated as follows [39]:

$$S_{Tani}(m_1, m_2) = \frac{\sum_{j=1}^{J} m_{1,j} m_{2,j}}{\sum_{j=1}^{J} (m_{1,j})^2 + \sum_{j=1}^{J} (m_{2,j})^2 - \sum_{j=1}^{J} m_{1,j} m_{2,j}} \tag{25}$$

Cosine similarity has been widely used to calculate the similarity between two belief structures, and is defined as [38]:

$$S_{Cos} = \frac{\sum_{j=1}^{J} m_{1,j} m_{2,j}}{\sqrt{\sum_{j=1}^{J} (m_{1,j})^2} \sqrt{\sum_{j=1}^{J} (m_{2,j})^2}} \tag{26}$$

Based on both the inner product and the sum of the squares of two belief structures, Dice similarity is defined as [40]:

$$S_{Dice}(m_1, m_2) = \frac{2 \sum_{j=1}^{J} m_{1,j} m_{2,j}}{\sum_{j=1}^{J} (m_{1,j})^2 + \sum_{j=1}^{J} (m_{2,j})^2} \tag{27}$$

3.2.5. Fidelity family

Based on the square root of probability distributions, the Fidelity family is a popular measure of distance in quantum theory, and includes Fidelity similarity, Bhattacharyya distance, Hellinger distance, Matusita distance, and Squared-Chord similarity.

Fidelity similarity directly uses the square root to calculate similarity between two belief structures as follows [38]:

$$S_{Fid}(m_1, m_2) = \sum_{j=1}^{J} \sqrt{m_{1,j} m_{2,j}} \tag{28}$$

Bhattacharyya distance combines the square root and the natural logarithm, and is expressed as follows [41]:

$$d_B(m_1, m_2) = -\ln \sum_{j=1}^{J} \sqrt{m_{1,j} m_{2,j}} \tag{29}$$

Hellinger distance is calculated as follows :

$$d_{Hel}(m_1, m_2) = \sqrt{2 \sum_{j=1}^{J} \left(\sqrt{m_{1,j}} - \sqrt{m_{2,j}} \right)^2} = 2\sqrt{1 - \sum_{j=1}^{J} \sqrt{m_{1,j} m_{2,j}}} \tag{30}$$

Matusita distance is represented as follows [38]:

$$d_{Mat}(m_1, m_2) = \sqrt{\sum_{j=1}^{J} \left(\sqrt{m_{1,j}} - \sqrt{m_{2,j}} \right)^2} = \sqrt{2 - 2\sum_{j=1}^{J} \sqrt{m_{1,j} m_{2,j}}} \tag{31}$$

Squared-Chord similarity is represented as follows [42]:

$$S_{Sq} = 2 \sum_{j=1}^{J} \sqrt{m_{1,j} m_{2,j}} - 1 \tag{32}$$

3.2.6. Squared Euclidean distance

Based on the Euclidean distance, the Squared Euclidean distance is calculated as follows [33]:

$$d_{sqe}(m_1, m_2) = \sum_{j=1}^{J} \left(m_{1,j} - m_{2,j} \right)^2 \tag{33}$$

3.2.7. Jousselme distance measure

Proposed by Jousselme et al. [43], Jousselme distance has been widely used to calculate the similarity between two belief structures. It combines the inner product and the Jaccard index, and is defined as:

$$d_{Jou}(m_1, m_2) = \sqrt{\frac{1}{2}(m_1 - m_2)' \underline{D}(m_1 - m_2)} \tag{34}$$

where D is the Jaccard matrix, and it represented as

$$\underline{D}(A, B) = \frac{|A \cap B|}{|A \cup B|},\ A, B \in 2^X \tag{35}$$

It should be noted that when Jousselme distance is used in calculating distance between evidence, both single and composite hypotheses are considered, i.e., $2^X - 1$ elements. However, as only single hypotheses are considered in the EBRB system, Jousselme distance would be simplified as:

$$d_{Jac}(m_1, m_2) = \sqrt{\frac{1}{2}(m_1 - m_2)'(m_1 - m_2)} = \sqrt{\frac{1}{2}\sum_{j=1}^{J}\left(m_{1,j} - m_{2,j}\right)^2} \tag{36}$$

Obviously, there is $0 \leq d_{Jac} \leq 1$, hence, the similarity can be calculated as $S = 1 - d_{Jac}$.

Jousselme distance has been applied to calculate the similarity between the input and the antecedent of the extended belief rule in several studies. However, though it could effectively model the difference between two belief structures, there still are some problems, and the most significant one is that the belief structure itself is not well-considered, and simply calculating the average value could neglect the characteristics of the belief structures. Hence, a modified Jousselme distance is introduced in this paper, where the sum of belief structures is used as the denominator, and is defined as follows:

$$d_{New}(m_1, m_2) = \sqrt{\frac{\sum_{j=1}^{J}\left(m_{1,j} - m_{2,j}\right)^2}{\sum_{j=1}^{J}\left(m_{1,j}\right)^2 + \sum_{j=1}^{J}\left(m_{2,j}\right)^2}} \tag{37}$$

In order to better demonstrate the performance and characteristics of different similarity measures, consider the following example.

Example 1. Let m_1 and m_2 be two evidence, where $m_1 = \{(H_1, 0.4), (H_2, 0.3), (H_3, 0.3)\}$ and $m_2 = \{(H_1, 0.2), (H_2, 0.7), (H_3, 0.1)\}$. Then, by using the above-similarity

Table 1: Characteristics and results of different similarity measures for Example 1

Similarity measure	Calculation formula	Similarity
Euclidean distance	$d_{Euc}=\sqrt{\sum_{j=1}^{J}(m_{i,j}-m_{2,j})^2}$	0.5101
Manhattan distance	$d_{Man}(m_1,m_2)=\sum_{j=1}^{J}\lvert m_{1,j}-m_{2,j}\rvert$	0.2000
Chebyshev distance	$d_{Cheb}(m_1,m_2)=\max_j \lvert m_{1,j}-m_{2,j}\rvert$	0.6000
Sorensen distance	$d_{Sor}(m_1,m_2)=\frac{\sum_{j=1}^{J}\lvert m_{1,j}-m_{2,j}\rvert}{\sum_{j=1}^{J}(m_{1,j}+m_{2,j})}$	0.4000
Lorentzian distance	$d_{Lor}(m_1,m_2)=\sum_{j=1}^{J}\ln(1+\lvert m_{1,j}-m_{2,j}\rvert)$	0.2989
Intersection similarity	$S_{IS}(m_1,m_2)=\sum_{j=1}^{J}\min(m_{1,j},m_{2,j})$	0.6000
Inner product similarity	$S_{IP}(m_1,m_2)=\sum_{j=1}^{J}m_{1,j}m_{2,j}$	0.3200
Tanimoto similarity	$S_{Tani}(m_1,m_2)=\frac{\sum_{j=1}^{J}m_{1,j}m_{2,j}}{\sum_{j=1}^{J}(m_{1,j})^2+\sum_{j=1}^{J}(m_{2,j})^2-\sum_{j=1}^{J}m_{1,j}m_{2,j}}$	0.5714
Cosine similarity	$S_{Cos}=\frac{\sum_{j=1}^{J}m_{1,j}m_{2,j}}{\sqrt{\sum_{j=1}^{J}(m_{1,j})^2}\sqrt{\sum_{j=1}^{J}(m_{2,j})^2}}$	0.7468
Dice similarity	$S_{Dice}(m_1,m_2)=\frac{2\sum_{j=1}^{J}m_{1,j}m_{2,j}}{\sum_{j=1}^{J}(m_{1,j})^2+\sum_{j=1}^{J}(m_{2,j})^2}$	0.7273
Fidelity similarity	$S_{Fid}(m_1,m_2)=\sum_{j=1}^{J}\sqrt{m_{1,j}m_{2,j}}$	0.9143
Bhattacharyya distance	$d_B(m_1,m_2)=-\ln\sum_{j=1}^{J}\sqrt{m_{1,j}m_{2,j}}$	0.9104
Hellinger distance	$d_{Hel}(m_1,m_2)=2\sqrt{1-\sum_{j=1}^{J}\sqrt{m_{1,j}m_{2,j}}}$	0.4145
Matusita distance	$d_{Mat}(m_1,m_2)=\sqrt{2-2\sum_{j=1}^{J}\sqrt{m_{1,j}m_{2,j}}}$	0.5860
Squared-Chord similarity	$S_{Sq}=2\sum_{j=1}^{J}\sqrt{m_{1,j}m_{2,j}}-1$	0.8286
Squared Euclidean distance	$d_{sqe}(m_1,m_2)=\sum_{j=1}^{J}(m_{1,j}-m_{2,j})^2$	0.7600
Jousselme distance	$d_{Jac}(m_1,m_2)=\sqrt{\frac{1}{2}\sum_{j=1}^{J}(m_{1,j}-m_{2,j})^2}$	0.6536
Modified Jousselme distance	$d_{New}(m_1,m_2)=\sqrt{\frac{\sum_{j=1}^{J}(m_{1,j}-m_{2,j})^2}{\sum_{j=1}^{J}(m_{1,j})^2+\sum_{j=1}^{J}(m_{2,j})^2}}$	0.4778

measures, the similarity between m_1 and m_2 is calculated, as summarized in Table 1.

From Table 1, it can be found that for different similarity measures, the obtained results for Example 1 are clearly different. More importantly, the differences among the results of different similarity measures could be quite significant, as the similarity calculated by the Manhattan distance is 0.2000, whereas the similarity calculated by the Fidelity similarity is 0.9143. Therefore, it is possible to combine the results using these different similarity measures to obtain more balanced and reliable results.

4. Ensemble method for EBRB systems with different similarity measures

Due to the different characteristics of different problems, different similarity measures may be needed in the EBRB system to achieve the best results, and it is necessary to combine the results of EBRB systems with different similarity measures to achieve higher accuracy. Hence, in this section, ensemble method for EBRB systems with different similarity measures is proposed, and it is comprised of two main components: an adaptive weight learning method for EBRB systems with different similarity measures, and an ER-based combination method for aggregating inferential results of EBRB systems with different similarity measures.

4.1. Framework of the EBRB ensemble method

In this section, the framework of the ensemble method for EBRB systems is illustrated, and as shown in Fig 1, the proposed method mainly includes three main phases.

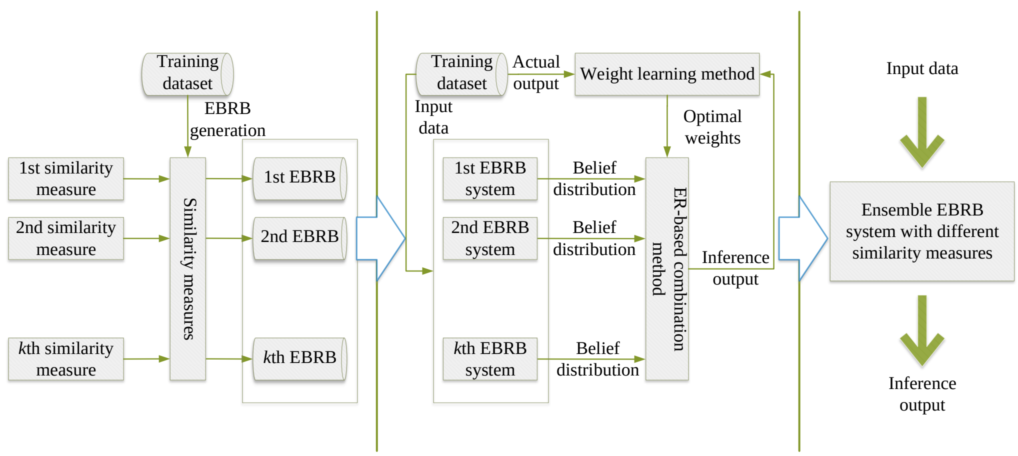

Figure 1: Framework of the proposed method.

First, EBRB systems with different similarity measures are constructed using training data based on the EBRB generation method [11], where the input is used to generate the antecedent attributes and the output is used to generate the consequents of the extended belief rules. Different similarity measures are applied in the inference approach, and a number of EBRB systems can be obtained.

Second, since each EBRB system with different similarity measures can produce an inferential result for the same input, it is necessary to combine the inferential results of all EBRB systems to achieve better results, which usually have different accuracy and importance. Hence, a weight learning method is applied to determine the importance of each EBRB system with different similarity measures.

Third, in order to combine the inferential results of all EBRB systems with different similarity measures, the ER algorithm is applied, and an ER-based combination method for aggregating results of different systems is introduced.

From the framework of the EBRB ensemble method, it is clear that the generation of EBRB systems with different similarity measures can be conducted based on the conventional generation method directly using training data since the EBRB system can be regarded as a data-driven system, and each EBRB system would have a different similarity measure, corresponding to the similarity measures introduced in Section 3.

4.2. The adaptive weight learning method for EBRB systems with different similarity measures

For different problems, there is no universal similarity measure that could achieve the best result due to their different characteristics, hence, when combining the inferential results of EBRB systems with different similarity measures, different EBRB systems could have different weights since they may have different performance. Suppose M weights $\{\omega_1, \omega_2, \ldots, \omega_M\}$ are used to represent the importance of each EBRB system with different similarity measures. Normally, it would be difficult to determine the optimal values of these weights by simply using expert knowledge as a small change may significantly influence the result of EBRB system. In recent years, the differential evolution (DE) algorithm has been widely used in optimization problems. Hence, since the weight learning process can be regarded as a typical optimization problem, a weight learning method based on the DE algorithm is proposed to determine the optimal value of these weights.

Proposed by Storn and Price [44], the DE algorithm has been widely used in optimization problems in the past decades. The basic idea of DE algorithm is to generate new candidate solutions by mutation and crossover operations according to the evolution strategy until the optimal solution is achieved or the ending criteria are reached. Based on the DE algorithm, the detailed steps of the weight learning method are introduced as follows.

Step 1: Initialize a set of candidate solutions, suppose that there are C sets of M solutions and the cth candidate is expressed as follows:

$$id_c = \{id_{c,m}; m = 1, 2, \ldots, M\}; c = 1, 2, \ldots, C \tag{38}$$

where $id_{c,m}$ is the importance degree of the mth system of the c set, and their initial value is determined using a random value between 0 and 1, i.e.,

$$id_{c,m} = random(0, 1); c = 1, 2, \ldots, C; m = 1, 2, \ldots, M \tag{39}$$

Step 2: Generate new candidate solutions. For each candidate set id_c in the C candidate sets, a new candidate set id_c^0 is generated by using three different candidate sets randomly selected from the C candidate sets, denoted as id_c^1, id_c^2, and id_c^3, respectively. The weights in the candidate set id_c^0 are assigned based on the evolution strategy as follows [45]:

$$id_{c,m}^0 = \begin{cases} id_{c,m}, & \text{if } random(0, 1) > CR \\ id_{c,m}^1 + F \times (id_{c,m}^2 - id_{c,m}^3), & \text{otherwise} \end{cases} \tag{40}$$

where F and CR are mutation and crossover constants, and are usually set as 0.9 and 0.5, respectively.

Step 3: Update all candidate solutions. When the importance degree in the new candidate set id_c^0 exceeds the range $[0, 1]$, a new value should be generated using Eq. (39). Then, according to the ER-based combination method, the error of the candidate set id_c^0, e.g., MSE or classification error, denoted as $\epsilon(id_c^0)$, can be calculated using training data, and the importance degrees of the candidate set id_c should be updated as follows:

$$id_c = \begin{cases} id_c^0, & \text{if } \epsilon(id_c^0) < \epsilon(id_c) \\ id_c, & \text{otherwise} \end{cases} \tag{41}$$

with

$$\epsilon(id_c) = \sum_{t=1}^{T} E_t, \ E_t = \begin{cases} 1, & \text{if } G(x_t) \neq y_t \\ 0, & \text{otherwise} \end{cases} \tag{42}$$

where $G(x_t)$ represents the inferential results of the EBRB system for the tth input x_t, and y_t is the actual output of the tth input.

Step 4: Select the best candidate solution. When the number of iterations reaches the maximum number of iterations S, the candidate set with the minimum error is selected as the best one, and its importance degrees are regarded as optimal and can be used to obtain the optimal weights for combining M EBRB systems with different similarity measures. The weights can be calculated based on the optimal importance degrees $id_{optimal}$ as follows:

$$\omega_{optimal} = \left\{ \frac{id_{optimal,m}}{\sum_{m=1}^{M} id_{optimal,m}}; m = 1, 2, \ldots, M \right\} \tag{43}$$

Remark 1. Step 3 guarantees that all candidate sets can only get better results or remain unchanged, hence, the weight learning process can converge to an optimal set of weights after performing S iterations.

Remark 2. The optimal results are directly affected by the maximum number of iterations S and the number of candidate sets C, and the inferential accuracy would become better if the C and S are larger, however, the weight learning process would also be more time-consuming. Hence, it is necessary to carefully choose these two parameters based on the actual system and situation to balance between inferential performance and time cost.

4.3. The ER-based combination method to aggregate inferential results of different EBRB systems

In order to make full use of different similarity measures for EBRB systems, it is necessary to develop an effective combination method to combine the inferential results of these EBRB systems with different similarity measures. As the results of all the EBRB systems are independent, the ER algorithm can be applied to propose a combination method for combining these EBRB systems.

Consider the importance of different EBRB systems with different similarity measures, together with the analytical ER algorithm, an ER-based combination method is proposed to aggregate the inferential results of EBRB systems with different similarity measures, and the process of the proposed method is detailed as follows:

Step 1: Suppose there are M EBRB systems with different similarity measures, and the inferential result of the mth EBRB system is denoted as $f_m(x) = \{(D_n, \beta_n^m)\}$ while x being the input. Therefore, the inferential results of all EBRB systems with different similarity measures $g(x)$ can be represented as follows:

$$g(x) = \begin{bmatrix} \beta_1^1 & \beta_2^1 & \dots & \beta_N^1 \\ \beta_1^2 & \beta_2^2 & \dots & \beta_N^2 \\ \vdots & \vdots & \dots & \vdots \\ \beta_1^M & \beta_2^M & \dots & \beta_N^M \end{bmatrix} \tag{44}$$

Step 2: Using the adaptive weight learning method detailed in Section 4.2, the weight of each EBRB system should be obtained. Then, using ω_m to represent the weight of the mth EBRB system, the analytical ER algorithm is applied to aggregate the inferential results of all EBRB systems as follows:

$$\beta_n = \frac{d\left[\prod_{m=1}^{M}\left(\omega_m\beta_n^m + 1 - \omega_m\sum_{k=1}^{N}\beta_k^m\right) - \prod_{m=1}^{M}\left(1 - \omega_m\sum_{k=1}^{N}\beta_k^m\right)\right]}{1 - d\left[\prod_{m=1}^{M}(1-\omega_m)\right]} \tag{45}$$

with

$$d = \left[\sum_{n=1}^{N}\prod_{m=1}^{M}\left(\omega_m\beta_n^m + 1 - \omega_m\sum_{k=1}^{N}\beta_k^m\right) - (N-1)\prod_{m=1}^{M}\left(1 - \omega_m\sum_{n=1}^{N}\beta_n^m\right)\right]^{-1} \tag{46}$$

where β_n denotes the aggregated belief degree on the nth grade.

Step 3: To obtain final inferential results, for regression problems, the final result can be produced using the referential value $u(D_n)$ of the nth grade as follows:

$$G(x) = \sum_{n=1}^{N} u(D_n)\beta_n \tag{47}$$

For classification problems, the final result can be obtained according to the biggest value from N aggregated belief degrees as follows

$$G(x) = D_t,\ t = \arg\max_{n=1,\ldots,N}\{\beta_n\} \tag{48}$$

5. Case study

In order to validate the performance of the above-mentioned similarity measures as well as the proposed ensemble method for EBRB with different similarity measures, a case study on classification problem using datasets from the well-known UCI machine-learning repository [46] is conducted.

5.1. Datasets and experimental setting

In order to test the performance of different similarity measures and the effectiveness and efficiency of the proposed method, 25 classification datasets from the UCI machine learning repository are chosen, and the main characteristics of these datasets are summarized in Table 2.

To construct the EBRB, suppose that each antecedent attribute has three antecedent grades and the consequent has the same number of grades as the number of classes. Furthermore, 10-fold cross-validation (10-CV) is used in

Table 2: Statistics on classification datasets

Dataset	# instances	# attributes	# classes
Avila	20867	10	12
Banana	5300	2	2
Banknote	1372	4	2
Breast	106	9	6
Cancer	699	9	2
Car	1728	6	4
Diabetes	768	8	2
Ecoli	336	7	8
Glass	214	9	7
Haberman	306	3	2
Iris	150	4	3
Knowledge	403	5	4
Letter	20000	16	26
Liver	345	6	2
Mammographic	830	5	2
Pageblocks	5472	10	5
Red wine	1599	11	10
Seeds	210	7	3
Transfusion	748	4	2
Vehicle	846	18	4
Vertebral	310	6	3
Vowel	990	10	11
Wine	178	13	3
Winconsin	683	9	2
Yeast	1484	8	10

this study, where each dataset is divided evenly into 10 blocks, with 9 blocks as training data and the remaining one as testing data.

Taking Iris dataset as an example, firstly, all four antecedent attributes are defined using three referential grades, and these grades are chosen evenly from each antecedent attribute data. The corresponding referential values for each antecedent attribute are expressed as follows:

$$U_1 \in \{L, M, H\} = \{4.3, 6.1, 7.9\}$$
$$U_2 \in \{L, M, H\} = \{2, 3.2, 4.4\}$$
$$U_3 \in \{L, M, H\} = \{1, 3.95, 6.9\}$$
$$U_4 \in \{L, M, H\} = \{0.1, 1.3, 2.5\}$$

Three consequent referential grades are used to described three classes as follows:

$$D \in \{\text{Iris Setosa}, \text{Iris Versicolour}, \text{Iris Virginica}\} = \{1, 2, 3\}$$

150 Iris data are equally divided into 10 blocks, in each iteration, 9 of the 10 blocks are used as training data to construct the EBRB, and the inferential result of each EBRB system can be obtained using different similarity measures based on the ER algorithm. The inferential results of all the EBRB systems are combined based on the proposed ensemble method for EBRB system with different similarity measures, where the weight of each EBRB system is determined based on the adaptive weight learning method using the DE algorithm. Finally, the accuracy of different EBRB systems is verified by the remaining one-block testing data. In order to more precisely and effectively compare the results, 5 independent runs of 10-CV are conducted.

5.2. Comparative analysis of different similarity measures

Firstly, as EBRB systems with different similarity measures could have different performance, in order to verify the applicability and performance of these similarity measures as well as the proposed method, the classification result of these EBRB systems and the EBRB system using the proposed method, denoted as ESM, are compared, and the classification accuracy of different EBRB systems are shown in Table 3 and Fig 2.

As shown in Table 3, though the Euclidean distance could achieve generally satisfying accuracy, the results are not necessarily the best, in fact, for datasets such as Car and Wine, its accuracy is clearly lower than other similarity measures. From the classification results of different EBRB systems with different similarity measures, it can be concluded that there is no universal similarity measure that can achieve the best accuracy for all datasets, as many factors such as internal structure and noise data are likely

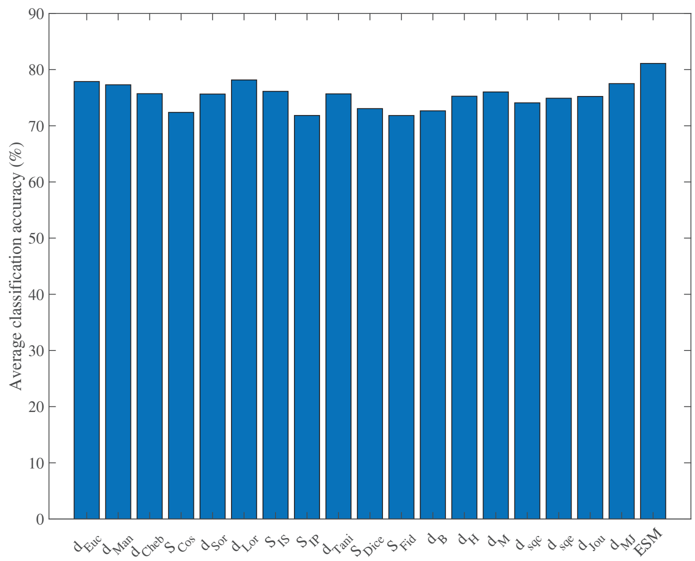

Figure 2: Average classification accuracy (%) of EBRB systems with different similarity measures and the proposed method.

to affect the classification results. For example, one of the worst similarity measures, S_{Fid} could still reach the second-best accuracy for Car dataset. However, it is also worth noting that though they cannot always achieve the best accuracy, several similarity measures stand out for being able to obtain relatively satisfactory results, such as the Lorentzian distance measure, as it could reach one of the highest accuracies for most datasets. Therefore, it is worth considering replacing the conventional Euclidean-based similarity in the EBRB system with novel similarity measures such as the Lorentzian distance.

More importantly, as demonstrated in Table 3 and Fig 2, the EBRB system using the proposed method always reaches the best accuracy. Furthermore, because of the weight learning process, the impact of noise data could be reduced, and its accuracy is even better than the best result from other EBRB systems. For example, for Diabetes dataset, the proposed method reaches 77.14% accuracy, which is significantly higher than the best results from other EBRB systems (75.15%), and it shows that the proposed method can improve the performance of the EBRB system. Therefore, it can be concluded that the proposed ensemble method for EBRB systems with different

Table 3: Comparison of classification accuracy (%) of EBRB systems with different similarity measures and the proposed method

Dataset	d_{Euc}	d_{Man}	d_{Cheb}	S_{Cos}	d_{Sor}	d_{Lor}	S_{IS}	S_{IP}	d_{Tani}	S_{Dice}	S_{Fid}	d_B	d_H	d_M	d_{sqc}	d_{sqe}	d_{Jou}	d_{MJ}	ESM
Avila	41.29 (3)	40.95 (5)	38.77 (14)	39.23 (11)	38.83 (13)	40.14 (7)	39.71 (10)	39.98 (8)	40.23 (6)	38.48 (15)	38.22 (16)	39.17 (12)	37.21 (19)	41.04 (4)	37.89 (18)	39.77 (9)	38.01 (17)	42.40 (2)	42.79 (1)
Banana	61.57 (4)	60.74 (6)	59.11 (14)	57.79 (18)	57.66 (19)	61.73 (3)	60.23 (8)	59.58 (11)	60.44 (7)	58.87 (15)	59.18 (13)	58.87 (15)	60.04 (9)	59.31 (12)	58.85 (17)	59.91 (10)	60.83 (5)	61.93 (2)	62.84 (1)
Banknote	98.44 (3)	98.99 (2)	95.22 (10)	94.59 (14)	95.22 (10)	97.04 (5)	95.22 (10)	95.93 (9)	96.63 (7)	93.79 (15)	84.48 (19)	87.65 (18)	96.72 (6)	92.51 (16)	90.16 (17)	96.56 (8)	94.69 (13)	97.51 (4)	99.05 (1)
Breast	69.53 (4)	69.93 (3)	67.10 (9)	56.11 (17)	67.44 (7)	70.73 (2)	67.44 (7)	48.56 (19)	61.96 (11)	58.52 (13)	57.40 (15)	56.68 (16)	57.60 (14)	53.09 (18)	61.91 (12)	63.30 (10)	68.20 (5)	67.99 (6)	71.28 (1)
Cancer	96.35 (3)	95.68 (6)	95.01 (8)	94.88 (11)	95.01 (8)	96.66 (2)	95.01 (8)	94.84 (13)	94.43 (16)	94.70 (14)	90.62 (19)	91.57 (18)	94.66 (15)	95.71 (5)	93.71 (17)	94.87 (12)	95.36 (7)	96.24 (4)	97.01 (1)
Car	83.72 (16)	79.41 (19)	91.35 (9)	91.94 (7)	91.35 (9)	92.74 (4)	91.35 (9)	92.77 (3)	89.93 (15)	91.20 (12)	93.03 (2)	92.73 (5)	80.72 (18)	92.42 (6)	91.88 (8)	83.69 (17)	90.12 (14)	90.98 (13)	93.46 (1)
Diabetes	74.74 (3)	74.32 (4)	70.55 (11)	69.53 (17)	70.55 (11)	75.15 (2)	70.55 (11)	69.01 (18)	72.50 (9)	69.64 (15)	66.85 (19)	70.34 (14)	73.33 (6)	73.25 (7)	72.06 (10)	73.52 (5)	69.61 (16)	72.92 (8)	77.24 (1)
Ecoli	77.48 (6)	81.58 (3)	66.50 (12)	59.93 (18)	66.50 (12)	73.22 (8)	66.50 (12)	58.38 (19)	71.81 (9)	61.30 (17)	62.02 (16)	69.22 (11)	82.47 (2)	81.45 (4)	79.55 (5)	69.45 (10)	62.39 (15)	74.08 (7)	83.72 (1)
Glass	61.79 (7)	62.06 (6)	61.75 (8)	57.92 (14)	61.75 (8)	62.43 (4)	61.75 (8)	57.63 (15)	61.07 (11)	57.45 (16)	49.11 (19)	50.06 (18)	62.47 (3)	62.65 (2)	53.95 (17)	60.28 (13)	60.91 (12)	62.37 (5)	62.82 (1)
Heberman	73.14 (10)	73.59 (2)	73.53 (3)	73.00 (13)	73.53 (3)	73.07 (12)	73.53 (3)	73.27 (9)	72.94 (16)	73.00 (13)	73.53 (3)	73.00 (13)	72.29 (19)	72.94 (16)	73.14 (10)	73.40 (8)	73.53 (3)	72.87 (18)	74.12 (1)
Iris	95.33 (4)	96.67 (2)	95.33 (4)	93.33 (15)	94.67 (8)	96.00 (3)	94.67 (8)	88.00 (19)	94.67 (8)	93.33 (15)	92.67 (17)	92.00 (18)	95.33 (4)	94.00 (13)	94.00 (13)	94.67 (8)	94.67 (8)	95.33 (3)	98.00 (1)
Knowledge	85.65 (3)	88.53 (2)	77.91 (9)	71.49 (16)	77.91 (9)	83.42 (5)	77.91 (9)	71.09 (17)	80.96 (8)	72.64 (15)	67.39 (19)	68.68 (18)	82.58 (6)	81.52 (7)	75.47 (14)	77.11 (12)	76.42 (13)	83.47 (4)	89.03 (1)
Letter	92.92 (6)	92.86 (7)	91.80 (16)	91.93 (14)	92.02 (13)	93.13 (3)	92.77 (8)	92.43 (11)	92.14 (12)	91.85 (15)	91.73 (17)	92.47 (10)	93.01 (4)	92.53 (9)	91.54 (18)	91.38 (19)	92.93 (5)	93.27 (2)	93.53 (1)
Liver	61.18 (4)	62.93 (3)	58.45 (11)	58.39 (14)	58.45 (11)	61.01 (5)	58.45 (11)	58.39 (14)	58.04 (18)	58.56 (10)	58.39 (14)	59.10 (8)	63.10 (2)	60.01 (6)	58.86 (9)	57.70 (19)	58.39 (14)	59.15 (7)	64.08 (1)
Mammographic	79.08 (8)	68.27 (17)	80.14 (4)	78.74 (13)	80.35 (2)	79.37 (6)	80.35 (2)	78.79 (10)	78.59 (14)	78.79 (10)	79.32 (7)	77.77 (15)	71.37 (16)	57.03 (18)	51.06 (19)	79.51 (5)	78.79 (10)	79.00 (9)	80.48 (1)
Pageblocks	93.46 (11)	94.13 (9)	94.20 (8)	93.53 (10)	95.02 (5)	95.21 (4)	94.47 (6)	93.10 (15)	93.27 (13)	93.19 (14)	92.96 (16)	92.87 (17)	92.80 (18)	91.79 (19)	93.33 (12)	94.23 (7)	95.43 (3)	95.77 (2)	96.02 (1)
Red wine	59.21 (4)	55.82 (15)	58.57 (7)	56.85 (14)	58.60 (5)	61.20 (2)	58.60 (5)	56.96 (13)	58.17 (10)	57.16 (12)	52.72 (19)	53.07 (18)	55.42 (16)	57.62 (11)	55.11 (17)	58.38 (9)	58.57 (7)	59.87 (3)	62.15 (1)
Seeds	92.00 (6)	93.05 (3)	91.24 (12)	90.57 (16)	91.24 (12)	91.43 (10)	91.24 (12)	90.19 (18)	92.10 (5)	90.48 (17)	91.62 (9)	91.05 (15)	93.52 (2)	93.05 (3)	91.81 (8)	91.90 (7)	89.90 (19)	91.33 (11)	94.76 (1)
Transfusion	76.58 (5)	77.09 (3)	76.34 (15)	76.47 (11)	76.34 (15)	76.39 (14)	76.34 (15)	76.53 (8)	76.58 (5)	76.45 (13)	76.29 (18)	76.47 (11)	77.14 (2)	76.63 (4)	76.58 (5)	76.53 (8)	76.23 (19)	76.50 (10)	78.67 (1)
Vehicle	70.99 (2)	49.15 (18)	66.93 (7)	62.22 (14)	66.93 (7)	70.57 (3)	66.93 (7)	60.78 (15)	67.26 (5)	63.43 (12)	60.03 (16)	59.78 (17)	35.95 (19)	66.54 (10)	62.52 (13)	64.51 (11)	67.00 (6)	68.49 (4)	74.71 (1)
Vertebral	69.61 (7)	73.35 (4)	67.94 (11)	65.23 (16)	67.94 (11)	70.19 (6)	67.94 (11)	66.13 (18)	67.16 (17)	64.52 (19)	68.32 (10)	69.61 (7)	78.77 (2)	73.94 (3)	71.48 (5)	67.29 (15)	68.77 (9)	67.74 (14)	79.96 (1)
Vowel	94.32 (6)	97.80 (2)	89.05 (8)	54.42 (18)	89.05 (8)	95.56 (5)	89.05 (8)	43.68 (19)	82.57 (13)	66.14 (17)	68.83 (16)	70.00 (15)	97.64 (3)	96.16 (4)	84.69 (12)	74.08 (14)	86.67 (11)	91.64 (7)	98.63 (1)
Wine	95.53 (13)	91.01 (18)	96.88 (4)	95.28 (16)	96.88 (4)	96.28 (8)	96.88 (4)	94.83 (17)	95.64 (12)	96.32 (7)	97.00 (3)	97.55 (2)	88.79 (19)	96.09 (9)	96.09 (9)	95.41 (15)	95.52 (14)	95.86 (11)	97.84 (1)
Wisconsin	93.16 (9)	97.36 (3)	90.42 (12)	89.89 (15)	89.10 (18)	96.34 (5)	97.51 (2)	97.21 (4)	89.39 (16)	90.18 (13)	89.38 (17)	90.03 (14)	84.99 (19)	94.00 (8)	95.01 (7)	93.05 (10)	90.47 (11)	96.16 (6)	97.81 (1)
Yeast	49.81 (4)	57.21 (2)	38.81 (11)	36.41 (18)	38.81 (11)	44.97 (6)	38.81 (11)	37.78 (14)	43.58 (8)	36.42 (17)	34.56 (19)	36.84 (16)	53.63 (3)	45.25 (5)	41.51 (10)	42.13 (9)	37.13 (15)	44.79 (7)	57.48 (1)

similarity measures is effective in assembling the best results of different EBRB systems to achieve the best accuracy.

In order to better demonstrate the effectiveness and efficiency of the proposed method, the nonparametric statistical test is performed, where the proposed method is compared in terms of classification accuracy. The Friedman tests are performed in the widely used KEEL [47], and the results are summarized in Table 4 and Table 5.

Table 4: Average Rankings of different methods (Friedman)

Method	Ranking
d_{Euc}	6.18
d_{Man}	6.58
d_{Cheb}	10.24
S_{Cos}	14.62
d_{Sor}	10.38
d_{Lor}	5.36
S_{IS}	9.02
S_{IP}	13.52
d_{Tani}	10.94
S_{Dice}	14.16
S_{Fid}	14.46
d_B	13.74
d_H	9.9
d_M	8.84
d_{sqc}	12.18
d_{sqe}	10.9
d_{Jou}	11.12
d_{MJ}	6.86
ESM	1

From Table 4, it can be found that for the first step of the Friedman test, the ranking of the ESM-EBRB is 1, which clearly indicates that the proposed ESM-EBRB method could achieve superior performance compared with other methods. From Table 5, it can be found that for the post hoc step, the proposed method is compared pairwise against other methods, where the p-value for most methods is less than 0.05, which further confirms the good performance of the proposed method. Therefore, from the statistical

Table 5: Post Hoc comparison Table for $\alpha = 0.05$ (FRIEDMAN)

i	Method	$z = (R_0 - R_i)/SE$	p
18	S_{Cos}	8.557185	0
17	S_{Fid}	8.45666	0
16	S_{Dice}	8.268176	0
15	d_B	8.004298	0
14	S_{IP}	7.866076	0
13	d_{sqc}	7.02418	0
12	d_{Jou}	6.358202	0
11	d_{Tani}	6.245112	0
10	d_{sqe}	6.219981	0
9	d_{Sor}	5.893274	0
8	d_{Cheb}	5.805315	0
7	d_H	5.5917	0
6	S_{IS}	5.038813	0
5	d_M	4.925722	0.000001
4	d_{MJ}	3.681726	0.000232
3	d_{Man}	3.505807	0.000455
2	d_{Euc}	3.254495	0.001136
1	d_{Lor}	2.739305	0.006157

test results, it can be concluded that the proposed method could provide more reliable and accurate results by combining the results of different EBRB systems.

5.3. Comparison with other rule-based systems

In order to further validate the effectiveness and efficiency of the proposed method, the classification accuracy of the proposed method is further compared with several state-of-the-art rule-based systems, including fuzzy rule-based classification system (FRBCS) [48], belief rule-based classification system (BRBCS) [7], fuzzy rule approach based on genetic cooperative-competitive learning (GCCLFR) proposed by Ishibuchi et al. [49], fuzzy rule-based system with rule weight specification (WFRBCS) proposed by Ishibuchi and Yamamoto [50], selection and reduction-based belief rule-based system based on greedy method (SR-BRB) [8], and EBRB with Euclidean distance measure (C-EBRB). The comparison results are shown in Table 6.

Table 6: Classification accuracy of the proposed method compared with other rule-based systems

Dataset	FRBCS	BRBCS	GCCLFR	WFRBCS	SR-BRB	C-EBRB	ESM-EBRB
Avila	35.52%	36.89%	41.38%	8.31%	25.89%	41.29%	**42.79%**
Banana	57.47%	61.73%	56.72%	60.32%	62.92%	61.57%	**62.84%**
Banknote	93.78%	95.68%	88.40%	95.12%	97.59%	98.44%	**99.05%**
Breast	52.57%	67.33%	57.53%	43.38%	68.18%	69.53%	**71.28%**
Cancer	93.16%	94.84%	86.64%	92.62%	94.91%	96.35%	**97.01%**
Car	83.14%	90.37%	70.02%	77.26%	91.03%	83.72%	**93.46%**
Diabetes	66.61%	70.32%	64.97%	72.90%	72.91%	74.74%	**77.24%**
Ecoli	75.70%	77.58%	52.08%	70.85%	78.26%	77.48%	**83.72%**
Glass	63.58%	68.76%	49.51%	60.72%	**69.05%**	61.79%	62.82%
Haberman	69.61%	68.74%	73.20%	73.21%	**75.33%**	73.14%	74.12%
Iris	93.33%	94.33%	96.67%	94.00%	**98.67%**	95.33%	98.00%
Knowledge	78.91%	86.13%	86.35%	66.74%	87.54%	85.65%	**89.03%**
Letter	89.09%	93.46%	32.26%	33.74%	**94.13%**	92.92%	93.53%
Liver	57.68%	59.91%	59.42%	57.97%	60.87%	61.18%	**64.08%**
Mammographic	71.80%	78.42%	77.71%	79.64%	**82.17%**	79.08%	80.48%
Pageblocks	91.63%	**94.14%**	89.78%	91.92%	96.27%	93.51%	93.53%
Red wine	47.60%	58.29%	47.40%	51.47%	58.72%	59.21%	**62.15%**
Seeds	81.57%	91.90%	90.48%	91.43%	94.29%	92.00%	**94.76%**
Transfusion	72.61%	76.20%	76.20%	76.60%	76.87%	76.58%	**78.67%**
Vehicle	58.49%	69.85%	52.83%	60.17%	71.28%	70.99%	**74.71%**
Vertebral	74.58%	80.03%	60.00%	60.00%	**83.97%**	69.61%	79.96%
Vowel	89.34%	93.74%	39.29%	48.08%	94.56%	94.32%	**98.63%**
Wine	90.95%	94.31%	93.79%	92.14%	94.90%	95.53%	**97.84%**
Wisconsin	91.06%	94.44%	95.61%	87.21%	95.32%	94.16%	**97.81%**
Yeast	46.23%	55.76%	34.02%	25.74%	52.68%	49.81%	**57.48%**

As shown in Table 6, the proposed method could achieve generally favorable results compared with other rule-based systems. For example, for Banknote dataset, the accuracy of ESM-EBRB is 99.05%, which is significantly better than FRBCS (93.78%), BRBCS (95.68%), GCCLFR (88.40%), WFRBCS (95.12%) , SR-BRB (97.59%), and C-EBRB (98.44%). However, it also should be noted that the classification results of the proposed method for multi-class datasets are less satisfying, for example, for Glass dataset, the proposed method only reaches better results than GCCLFR, WFRBCS and C-EBRB. That is because when the number of classes increases, simply adjusting the similarity measure may not be efficient enough, and some inconsistent rules could still be activated, thus affecting the classification accuracy. For two-class or three-class datasets, the inconsistency is relatively low, thus better accuracy could be obtained. Therefore, it can be concluded that the proposed method could achieve better accuracy than other rule-based systems for small-scale datasets.

5.4. Comparison with conventional machine learning methods

To further verify the validity of the proposed method, the classification results of the proposed method are also compared with several conventional machine learning methods. The machine learning methods are conducted using the Waikato Environment for Knowledge Analysis (WEKA) software [51], where k-nearest neighbor (KNN), naive Bayes (NB), C4.5, support vector machine (SVM), and artificial neural network (ANN) are utilized. For these methods, the settings are as follows: k for kNN is set to be 5% of the samples in the training set, i.e., the top 5% nearest neighbors in the training set are used (decimals are omitted), the minimum number of instances per leaf is set as 2 for C4.5, the polynomial kernel is adopted in SVM and 50% of the sum of the number of attributes and classes is the number of hidden layers for ANN. The results of the aforementioned machine learning methods, as well as ESM-EBRB, are shown in Table 7.

Based on the comparison of classification results of the proposed method with conventional machine learning methods, it can be concluded that the proposed method can achieve relatively satisfying results. For example, for Cancer and Iris datasets, the proposed method obtains 97.01% and 98.00% accuracy, respectively, outperforming other methods. As for other datasets such as Banknote, Vehicle and Wine, though the proposed method fails to generate the highest accuracy, its results are still relatively high, better than most of the other methods. However, it should be noted that for some multi-class datasets, such as Glass and Ecoli, the result of the proposed method is less satisfying, which indicates that the proposed method is less ideal for multi-class datasets. Furthermore, from the comparison, it can be found that there is no universal method that can achieve the best accuracy for all datasets, as many factors such as internal structure and noise data would likely affect the classification results. It can be concluded that the proposed method provides an effective and efficient way for small-scale classification problems.

Table 8 shows the running time of these methods on the testing dataset. From Table 8, it can be observed that the running time of ESM-EBRB is clearly higher than the comparative methods for many datasets, which can be explained by the fact that ESM-EBRB combines the results of several EBRB systems when determining the classification. However, it is worth noting that the differences in the running time are not very significant, which can

Table 7: Classification accuracy of the proposed method compared with conventional machine learning methods

Dataset	kNN	NB	C4.5	SVM	ANN	ESM-EBRB
Banknote	98.98%	84.26%	98.54%	98.03%	**99.93%**	99.05%
Breast	70.67%	70.59%	70.63%	69.50%	63.54%	**71.28%**
Cancer	95.96%	92.97%	93.32%	62.74%	96.31%	**97.01%**
Car	93.11%	85.60%	92.34%	91.99%	**94.99%**	93.46%
Diabetes	74.09%	76.30%	73.82%	65.10%	75.39%	**77.24%**
Ecoli	85.71%	85.42%	84.23%	75.60%	**86.01%**	83.72%
Glass	66.36%	48.60%	66.82%	**68.69%**	67.76%	62.82%
Haberman	70.22%	69.60%	**75.53%**	71.11%	74.55%	74.12%
Iris	96.67%	96.00%	96.00%	96.67%	97.33%	**98.00%**
Knowledge	84.49%	79.53%	88.40%	85.56%	88.97%	**89.03%**
Liver	61.88%	54.80%	63.38%	59.49%	**64.46%**	64.08%
Mammographic	78.46%	78.36%	82.10%	79.29%	**80.96%**	80.48%
Red wine	59.59%	54.48%	60.04%	61.12%	**63.03%**	62.15%
Seeds	92.38%	91.43%	91.90%	90.48%	**95.24%**	94.76%
Transfusion	75.16%	73.54%	78.01%	74.44%	77.92%	**78.67%**
Vehicle	70.00%	42.42%	71.87%	59.95%	**76.47%**	74.71%
Vertebral	80.01%	76.66%	79.00%	78.89%	79.25%	**79.96%**
Vowel	97.07%	64.59%	80.04%	84.14%	92.81%	**98.63%**
Wine	93.29%	90.71%	92.84%	93.13%	97.82%	**97.84%**
Yeast	58.22%	57.61%	55.39%	43.26%	**59.03%**	57.48%

be tolerated in most cases. Nevertheless, the applicability of the ESM-EBRB method for large-scale datasets could be limited.

6. Conclusion

In this paper, in order to address the insufficiency in the similarity measure of the conventional EBRB system, a variety of similarity measures are analyzed, and their application in the EBRB system is studied. Based on that, an ensemble method for EBRB systems with different similarity measures is proposed. Twenty classification datasets are tested with 10-fold cross-validation to validate the effectiveness and efficiency of the proposed method, and the results are compared with some of the state-of-the-art rule-based systems and machine learning methods. The main contributions of this paper can be summarized into three aspects below:

(1) Previous studies on EBRB systems simply use Euclidean distance

Table 8: Running time (s) of the proposed method compared with conventional machine learning methods

Dataset	kNN	NB	C4.5	SVM	ANN	ESM-EBRB
Banknote	0.4391	0.2355	0.1874	0.1541	0.1832	0.4551
Breast	0.0018	0.0016	0.0011	0.0021	0.0011	0.0015
Cancer	0.2764	0.1829	0.1296	0.1545	0.1607	0.2563
Car	0.3546	0.3241	0.2864	0.3053	0.2958	0.3428
Diabetes	0.0574	0.0483	0.0387	0.0362	0.0390	0.0597
Ecoli	0.0132	0.0127	0.0113	0.0098	0.0118	0.0152
Glass	0.0118	0.0075	0.0068	0.0059	0.0089	0.0113
Haberman	0.0089	0.0067	0.0053	0.0049	0.0058	0.0085
Iris	0.0019	0.0016	0.0014	0.0014	0.0012	0.0018
Knowledge	0.0098	0.0079	0.0073	0.0068	0.0073	0.0121
Liver	0.0213	0.0189	0.0174	0.0167	0.0182	0.0193
Mammographic	0.0489	0.0346	0.0322	0.0310	0.0347	0.0488
Red wine	0.6315	0.5426	0.4858	0.4507	0.5029	0.7326
Seeds	0.0069	0.0064	0.0063	0.0059	0.0067	0.0068
Transfusion	0.0089	0.0064	0.0058	0.0053	0.0063	0.0076
Vehicle	0.2674	0.1583	0.1428	0.1271	0.1375	0.2400
Vertebral	0.0084	0.0070	0.0068	0.0053	0.0060	0.0076
Vowel	0.0846	0.0736	0.0633	0.0582	0.0610	0.0763
Wine	0.0143	0.0127	0.0119	0.0103	0.0117	0.0136
Yeast	0.4128	0.3726	0.2487	0.1878	0.2175	0.4038

to calculate the similarity between the input and the extended belief rule, which could lead to counterintuitive results in some cases, thus affecting the performance of the EBRB system. Therefore, a variety of similarity measures are analyzed, and their application in the EBRB system is studied to take full advantage of each measure.

(2) According to the analysis of different similarity measures, an adaptive weight learning method based on the DE algorithm for EBRB systems with different similarity measures is proposed to determine the weight of each EBRB system with different similarity measures. Using this method, the optimal weights for different EBRB systems can be obtained for combination.

(3) The detailed process of the ensemble method for EBRB systems with different similarity measures is introduced, which can generate better results by combining the inferential results of EBRB systems with different

similarity measures. Twenty classification datasets are used for validation, and it is validated that the proposed EBRB ensemble method can obtain better results than other EBRB systems with different similarity measures. More importantly, the proposed method could achieve better performance than other rule-based systems and conventional machine learning methods on some classification datasets.

For future research, as the proposed method is less suitable for multi-class datasets while adjusting rule activation method has been proven to be effective for some multi-class datasets, how to combine the proposed method and modified rule activation method should be studied to further promote the application of the EBRB system.

Acknowledgment

This work was supported by the Shandong Provincial Natural Science Foundation under Grant No. ZR2023QF148.

References

[1] R. Sun, Robust reasoning: integrating rule-based and similarity-based reasoning, Artificial Intelligence 75 (1995) 241–295.

[2] M. Negnevitsky, Artificial intelligence: a guide to intelligent systems, Pearson education, 2005.

[3] F. Hayes-Roth, Rule-based systems, Communications of the ACM 28 (1985) 921–932.

[4] A. Ligeza, Logical foundations for rule-based systems, volume 11, Springer, 2006.

[5] P.-C. Chang, C.-H. Liu, A tsk type fuzzy rule based system for stock price prediction, Expert Systems with applications 34 (2008) 135–144.

[6] P. Angelov, R. Yager, A new type of simplified fuzzy rule-based system, International Journal of General Systems 41 (2012) 163–185.

[7] L. M. Jiao, Q. Pan, T. Denoeux, Y. Liang, X. X. Feng, Belief rule-based classification system: Extension of frbcs in belief functions framework, Information Sciences 309 (2015) 26–49.

[8] F. Gao, A. Zhang, W. Bi, J. Ma, A greedy belief rule base generation and learning method for classification problem, Applied Soft Computing 98 (2021) 106856. doi:10.1016/j.asoc.2020.106856.

[9] F. Gao, Density-based approach for fuzzy rule interpolation, Applied Soft Computing 143 (2023) 110402.

[10] J.-B. Yang, J. Liu, J. Wang, H.-S. Sii, H.-W. Wang, Belief rule-base inference methodology using the evidential reasoning approach-rimer, IEEE Transactions on systems, Man, and Cybernetics-part A: Systems and Humans 36 (2006) 266–285.

[11] J. Liu, L. Martinez, A. Calzada, H. Wang, A novel belief rule base representation, generation and its inference methodology, Knowledge-based systems 53 (2013) 129–141.

[12] L.-H. Yang, J. Liu, Y.-M. Wang, L. Martínez, A micro-extended belief rule-based system for big data multiclass classification problems, IEEE Transactions on Systems, Man, and Cybernetics: Systems (2018).

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

[14] L.-H. Yang, J. Liu, F.-F. Ye, Y.-M. Wang, C. Nugent, H. Wang, L. Martínez, Highly explainable cumulative belief rule-based system with effective rule-base modeling and inference scheme, Knowledge-Based Systems 240 (2022) 107805.

[15] T. U. Ahmed, M. N. Jamil, M. S. Hossain, R. U. Islam, K. Andersson, An integrated deep learning and belief rule base intelligent system to predict survival of covid-19 patient under uncertainty, Cognitive Computation (2022) 1–17.

[16] S. Kabir, R. U. Islam, M. S. Hossain, K. Andersson, An integrated approach of belief rule base and convolutional neural network to monitor air quality in shanghai, Expert Systems with Applications 206 (2022) 117905.

[17] L.-H. Yang, F.-F. Ye, Y.-M. Wang, Y.-X. Lan, C. Li, Extended belief rule-based system using bi-level joint optimization for environmental investment forecasting, Applied Soft Computing 140 (2023) 110275.

[18] L.-H. Yang, J. Liu, Y.-M. Wang, L. Martínez, Comparative analysis on extended belief rule-based system for activity recognition, in: Data Science and Knowledge Engineering for Sensing Decision Support: Proceedings of the 13th International FLINS Conference (FLINS 2018), volume 11, World Scientific, 2018, p. 430.

[19] L.-H. Yang, J. Liu, Y.-M. Wang, C. Nugent, L. Martínez, Online updating extended belief rule-based system for sensor-based activity recognition, Expert Systems with Applications 186 (2021) 115737.

[20] F. Gao, W. Bi, A fast belief rule base generation and reduction method for classification problems, International Journal of Approximate Reasoning 160 (2023) 108964.

[21] F.-F. Ye, L.-H. Yang, Y.-M. Wang, L. Chen, An environmental pollution management method based on extended belief rule base and data envelopment analysis under interval uncertainty, Computers & Industrial Engineering (2020) 106454.

[22] A. Zhang, F. Gao, M. Yang, W. Bi, Belief rule-based dependence assessment method under interval uncertainty, Quality and Reliability Engineering International 36 (2020) 2459–2477.

[23] F. Gao, A. Zhang, W. Bi, Weapon system operational effectiveness evaluation based on the belief rule-based system with interval data, Journal of Intelligent & Fuzzy Systems 39 (2020) 6687–6701. doi:10.3233/jifs-190651.

[24] F.-F. Ye, L.-H. Yang, Y.-M. Wang, H. Lu, A data-driven rule-based system for china's traffic accident prediction by considering the improvement of safety efficiency, Computers & Industrial Engineering 176 (2023) 108924.

[25] L.-H. Yang, F.-F. Ye, Y.-M. Wang, Ensemble belief rule base modeling with diverse attribute selection and cautious conjunctive rule for classification problems, Expert Systems with Applications 146 (2020) 113161.

[26] Z. Zhou, Z. Feng, C. Hu, G. Hu, W. He, X. Han, Aeronautical relay health state assessment model based on belief rule base with attribute reliability, Knowledge-Based Systems 197 (2020) 105869.

[27] Y. Cao, Z. J. Zhou, C. H. Hu, S. W. Tang, J. Wang, A new approximate belief rule base expert system for complex system modelling, Decision Support Systems (2021) 113558.

[28] L.-H. Yang, J. Liu, Y.-M. Wang, H. Wang, L. Martínez, Enhancing extended belief rule-based systems for classification problems using decomposition strategy and overlap function, International Journal of Machine Learning and Cybernetics (2021) 1–27.

[29] A. Zhang, F. Gao, M. Yang, W. Bi, A new rule reduction and training method for extended belief rule base based on dbscan algorithm, International Journal of Approximate Reasoning (2020).

[30] L. H. Yang, J. Liu, Y. M. Wang, L. Martinez, Extended belief-rule-based system with new activation rule determination and weight calculation for classification problems, Applied Soft Computing 72 (2018) 261–272.

[31] L.-H. Yang, Y.-M. Wang, Y.-X. Lan, L. Chen, Y.-G. Fu, A data envelopment analysis (dea)-based method for rule reduction in extended belief-rule-based systems, Knowledge-Based Systems 123 (2017) 174–187.

[32] H. Zhu, M. Xiao, L. Yang, X. Tang, Y. Liang, J. Li, A minimum centre distance rule activation method for extended belief rule-based classification systems, Applied Soft Computing (2020) 106214.

[33] S.-H. Cha, Comprehensive survey on distance/similarity measures between probability density functions, International Journal of Mathematical models and Methods in Applied Sciences 1 (2007) 300–307.

[34] A. L. Jousselme, P. Maupin, Distances in evidence theory: Comprehensive survey and generalizations, International Journal of Approximate Reasoning 53 (2012) 118–145. doi:10.1016/j.ijar.2011.07.006.

[35] C. C. Aggarwal, A. Hinneburg, D. A. Keim, On the surprising behavior of distance metrics in high dimensional space, in: International conference on database theory, Springer, 2001, pp. 420–434.

[36] F. Van Der Heijden, R. P. Duin, D. De Ridder, D. M. Tax, Classification, parameter estimation and state estimation: an engineering approach using MATLAB, John Wiley & Sons, 2005.

[37] L. Jost, Entropy and diversity, Oikos 113 (2006) 363–375.

[38] M.-M. Deza, E. Deza, Dictionary of distances, Elsevier, 2006.

[39] S. E. Schaeffer, Graph clustering, Computer science review 1 (2007) 27–64.

[40] P. Legendre, L. F. Legendre, Numerical ecology, Elsevier, 2012.

[41] D. J. MacKay, D. J. Mac Kay, Information theory, inference and learning algorithms, Cambridge university press, 2003.

[42] D. G. Gavin, W. W. Oswald, E. R. Wahl, J. W. Williams, A statistical approach to evaluating distance metrics and analog assignments for pollen records, Quaternary Research 60 (2003) 356–367.

[43] A.-L. Jousselme, D. Grenier, E. Bossé, A new distance between two bodies of evidence, Information Fusion 2 (2001) 91–101.

[44] R. Storn, K. Price, Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces, Journal of global optimization 11 (1997) 341–359.

[45] L.-H. Yang, Y.-M. Wang, J. Liu, L. Martínez, A joint optimization method on parameter and structure for belief-rule-based systems, Knowledge-Based Systems 142 (2018) 220–240.

[46] D. Dua, C. Graff, UCI machine learning repository, 2017. URL: `http://archive.ics.uci.edu/ml`.

[47] I. Triguero, S. Gonzalez, J. M. Moyano, S. Garcia, J. Alcala-Fdez, J. Luengo, A. Fernandez, M. J. del Jesus, L. Sanchez, F. Herrera, Keel 3.0: An open source software for multi-stage analysis in data mining, International Journal of Computational Intelligence Systems 10 (2017) 1238.

[48] Z. Chi, H. Yan, T. Pham, Fuzzy algorithms: with applications to image processing and pattern recognition, volume 10, World Scientific, 1996.

[49] H. Ishibuchi, T. Nakashima, T. Murata, Performance evaluation of fuzzy classifier systems for multidimensional pattern classification problems, IEEE Transactions on Systems Man and Cybernetics Part B-Cybernetics 29 (1999) 601–618. doi:10.1109/3477.790443.

[50] H. Ishibuchi, T. Yamamoto, Rule weight specification in fuzzy rule-based classification systems, IEEE Transactions on Fuzzy Systems 13 (2005) 428–435. doi:10.1109/tfuzz.2004.841738.

[51] G. Holmes, A. Donkin, I. H. Witten, Weka: A machine learning workbench, in: Proceedings of ANZIIS'94-Australian New Zealnd Intelligent Information Systems Conference, IEEE, 1994, pp. 357–361.