IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 9, 2020, accepted July 22, 2020, date of publication July 28, 2020, date of current version August 7, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3012453

Construction and Reasoning Approach of Belief Rule-Base for Classification Base on Decision Tree

YANGGENG FU, ZEFENG YIN, MANNA SU, YINGJIE WU, AND GENGGENG LIU, (Member, IEEE)
College of Mathematics and Computer Science, Fuzhou University, Fuzhou 350116, China

Corresponding author: Genggeng Liu (liugenggeng@fzu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 71501047 and Grant 61773123, and in part by the Natural Science Foundation of Fujian Province, China, under Grant 2019J01647.

ABSTRACT The classical belief rule-based (BRB) systems are usually constructed by arranging and combining referential values of antecedent attributes or by setting special fixed values, which can lead to overly large size of BRB systems in complex problems. This paper combines the decision tree classification method to analyze the information of data and extract the rules. Based on this, a new rule representation method with referential interval is proposed and the rule base is constructed according to the support degree and belief degree of the data. In the newly proposed method, the introduction of decision tree ensures that the size of the rule base is reasonable. Moreover, the rule parameters trained by the differential evolution (DE) algorithm are optimized and adjusted to further improve the system performance. The experiments are conducted on several commonly used public classification datasets. And the proposed algorithm can achieve better accuracy results compared with classical classification methods and the existing classification methods of BRB systems on average. The experimental results validate the reasonableness and effectiveness of the BRB construction method proposed in this paper.

INDEX TERMS Intelligent decision, belief rule-based system, decision tree, parameter learning.

I. INTRODUCTION

With the deepening of informatization in various fields, a large amount of raw data has emerged. With the arrival of the era of big data, the scale and dimension of data have become two aspects that must be considered in the research process. How to obtain the related information from the massive data and analyze it has become the main research object of data mining. As an important field in data mining, data classification uses known classification label data to classify unknown classification label data by extracting data features for commonality and difference analysis [1]. At present, the existing classification algorithms mostly use classical data mining methods, such as Bayesian algorithm, support vector machine (SVM) classifier, rough set algorithm, and neural network to construct the classifiers, which cannot effectively handle uncertain information and fuzzy information in data. At the same time, based on the background of big data, it is difficult to explain the classification rules for the case of large-scale parameters in complex classification problems. Therefore, it is difficult to adjust the specific parameters through logic analysis to improve the performance of the classifier.

At present, D-S evidence theory [2], [3], decision theory [4], fuzzy theory [5], and classical IF-THEN rule base [6] have provided a good theoretical basis for solving multi-attribute decision-making problems. However, for the data with a large number of uncertain information problems, it still requires a decision-making method that can effectively deal with uncertain information. Therefore, based on the theories above, Yang *et al.* [7] proposed BRB system based on the BRB inference methodology using the evidential reasoning (RIMER) approach in 2006. Compared with the classical IF-THEN rule base, BRB is a new method proposed to deal with fuzzy information based on the classical processing of uncertain information. The method embeds belief distribution in the consequent term of each rule. The method also calculates the rule activation weights by matching the antecedent attributes in reasoning. The model trained by this method is more in line with the human way of thinking and has a strong explanatory. This method has been successfully applied in the engineering fields of graphite component detection [7],

The associate editor coordinating the review of this manuscript and approving it for publication was Jerry Chun-Wei Lin.

oil pipeline leak detection [8], fault diagnosis of flight control system [9], intension identification in air defence [10], and so on.

Although BRB has a strong ability to handle fuzzy and uncertain information, it is only suitable for classification problems with a small number of antecedent attributes. If there are a large number of antecedent attributes, to cover all possible cases, the number of rules in this system is bound to increase exponentially. Based on this problem, many scholars have researched and improved the method of selecting rules and building the rule base. For example, Jiao *et al.* [11] proposed BRB classification system (BRBCS) through fuzzy rule-based classification systems [12]–[14]. The BRBCS rule represents the antecedent term of BRB as a fuzzy array, so the number of rules in the BRBCS is limited by the number of fuzzy intervals and the amount of sample data. For the large number of parameters in BRBCS, Liu *et al.* [15] trained it by DE algorithm to improve system performance. However, for the classification problem with a large number of antecedent attributes and high data dimension, the number of rules constructed by dividing the fuzzy intervals will increase exponentially in BRBCS, and the massive parameters will be difficult to train and adjust. Therefore, Ye *et al.* [16] proposed a method of directly setting the number of rules in the rule base according to the number of classes of the classification result. The referential values of attributes of the rules are linearly combined to limit the size of the rule base to the range that can be adjusted for parameter training. However, if the value range of the antecedent attribute does not have a simple linear relationship with the result of classification, an obvious error will occur in this method.

To solve the problem described above effectively, we propose a new constructing and reasoning approach of BRB. The main contribution of this paper can be summarized as follows.

1) We combine C4.5 decision tree algorithm to determine the number of rules and the scope of antecedent attributes. This method can effectively reduce the number of rules in BRB. And it also provides a reasonable reference for the origin of the rule.
2) Based on C4.5 algorithm, a new rule representation method with referential interval of BRB system is proposed and the rule base is constructed according to the support degree and belief degree of data. We modify the representation of BRB so that the referential value of antecedent attriture is replaced by referential interval. Based on this, we propose a new individual matching method and a setting method of belief distribution. DE algorithm is combined to optimize the parameters of BRB system.
3) We compare the newly proposed method with some classical classification methods and existing classification methods of BRB system. We test several commonly used public classification datasets from the University of California at Irvine (UCI). The reasonableness, effectiveness, and superiority in the accuracy of the proposed method are all validated.

The remainder of the paper is organized as follows. Section II introduces the knowledge of BRB theory and the classification method of BRB system and expounds on the problems in the rule construction. Section III introduces the C4.5 decision tree algorithm and the classification BRB construction and reasoning method based on the rules extracting from the decision tree. To prove the good performance of the proposed method, several comparisons on common classification datasets are presented in Section IV. Section V concludes this paper.

II. INTRODUCTION OF BRB SYSTEM

The BRB system proposed by Yang *et al.* [7] contains two aspects: the BRB representation and the BRB inference methodology using the evidential reasoning approach. The section briefly introduces the relevant theoretical knowledge of BRB system.

A. REPRESENTATION OF BRB

To express the uncertainty information, Yang *et al.* [7] introduced the belief distribution of consequent term based on the classical IF-THEN rule, and added the antecedent attribute weight and the rule weight to reflect the importance of the attribute and the rule. In the rule base, the kth ($k = 1, \ldots, L$) rule can be expressed as follows:

$$\begin{aligned} R_k :\ & \textit{if } x_i \textit{ is } A_i^k \quad i = 1, 2, \ldots, T_k \\ & \textit{then } \{(D_j, \beta_j^k); \quad j = 1, 2, \ldots, N\} \\ & \textit{with rule weight } \theta_k \textit{ and attribute weight } \delta_i^k \end{aligned} \tag{1}$$

where L represents the number of rules in the system. A_i^k represents the referential value of the ith antecedent attribute in the kth rule. T_k represents the number of antecedent attributes in the antecedent term of the kth rule. D_j represents the jth referential value of the consequent attribute. β_j^k denotes the belief degree to D_j in the kth rule. N represents the number of referential values of consequent attribute. θ_k and δ_i^k denote the weight of the kth rule and the weight of the ith antecedent attribute in the kth rule, respectively.

B. BRB INFERENCE METHODOLOGY USING THE EVIDENTIAL REASONING APPROACH

The BRB inference methodology using the evidential reasoning approach is the core of BRB system. It mainly consists of two parts: activation of rules and synthesis of activated rules.

1) ACTIVATION WEIGHT CALCULATION AND BELIEF CORRECTION

For an input sample for the system $X(x_1, x_2, \ldots, x_{T_k})$, the activation weight of kth rule is as follows:

$$\omega_k = \frac{\theta_k \prod_{i=1}^{T_k} \left(\alpha_i^k\right)^{\overline{\delta_i^k}}}{\sum_{l=1}^{L} \left[\theta_l \times \prod_{i=1}^{T_k} \left(\alpha_i^l\right)^{\overline{\delta_i^l}}\right]} \tag{2}$$

Among them, the weight of the antecedent attributes is normalized as: $\overline{\delta_i^k} = \frac{\delta_i^k}{max_i=\{\delta_i^k|i=1,\ldots,T_k\}}$. α_i^k represents the individual matching degree of the input sample $X(x_1, x_2, \ldots, x_{T_k})$ for the ith antecedent attribute, and the calculation equation is as follows:

$$\alpha_i^k = \begin{cases} \frac{A_{i(c+1)} - x_i}{A_{i(c+1)} - A_{i(c)}} & A_i^k = A_{i(c)}, A_{i(c)} \le x_i \le A_{i(c+1)} \\ \frac{x_i - A_{i(c)}}{A_{i(c+1)} - A_{i(c)}} & A_i^k = A_{i(c+1)}, A_{i(c)} \le x_i \le A_{i(c+1)} \\ 0 & otherwise \end{cases} \tag{3}$$

where $A_{i(c)}$ represents the cth referential value of the ith antecedent attribute.

When $\sum_{j=1}^{N} \beta_j^k = 1$, the rule is called complete. Conversely, when $\sum_{j=1}^{N} \beta_j^k < 1$, the belief distribution of consequent term needs to be corrected as follows:

$$\overline{\beta_j^k} = \beta_j^k \frac{\sum_{t=1}^{T_k}\left(\tau(t,k)\sum_{i=1}^{|A_t|}\alpha_{t,i}\right)}{\sum_{t=1}^{T_k}\tau(t,k)}$$

Among them, $\tau(t,k) = \begin{cases} 1 & U_t \in R_k,\ t=1,\ldots,T_k \\ 0 & otherwise \end{cases}$ (4)

2) RULE SYNTHESIS

For the activated rules, the Evidential Reasoning (ER) algorithm [7] is used for synthesis.

First, convert the resulting belief distribution to the probability mass for ER synthesis:

$$m_j^k = \omega_k \beta_j^k \tag{5}$$

$$\widetilde{m_D^k} = \omega_k\left(1 - \sum_{j=1}^{N}\beta_j^k\right) \tag{6}$$

$$\overline{m_D^k} = 1 - \omega_k \tag{7}$$

where m_j^k represents the credibility of the jth referential value of consequent attritube of the kth rule. $\widetilde{m_D^k}$ denotes the credibility of the kth rule that is not assigned to any referential value of consequent attritube due to incomplete belief distribution. $\overline{m_D^k}$ denotes the credibility of the kth rule that is not assigned to any referential value of consequent attritube due to activation weights.

For probability mass, the ER analytical formulas [17] is used to synthesize the activation rules:

$$C_j = k\left[\prod_{l=1}^{L}\left(m_j^l + \widetilde{m_D^l} + \overline{m_D^l}\right) - \prod_{l=1}^{L}\left(\overline{m_D^l} + \widetilde{m_D^l}\right)\right] \tag{8}$$

$$\widetilde{C_D} = k\left[\prod_{l=1}^{L}\left(\widetilde{m_D^l} + \overline{m_D^l}\right) - \prod_{l=1}^{L}\overline{m_D^l}\right] \tag{9}$$

$$\overline{C_D} = k\prod_{l=1}^{L}\overline{m_D^l} \tag{10}$$

$$k^{-1} = \sum_{j=1}^{N}\prod_{l=1}^{L}\left(m_j^l + \widetilde{m_D^l} + \overline{m_D^l}\right) - (N-1)\prod_{l=1}^{L}\left(\widetilde{m_D^l} + \overline{m_D^l}\right) \tag{11}$$

According to the rule synthesis results, the belief distribution of each referential value of consequent attritube is obtained:

$$\beta_j = \frac{C_j}{1 - \overline{C_D}} \tag{12}$$

$$\beta_D = \frac{\widetilde{C_D}}{1 - \overline{C_D}} \tag{13}$$

where β_j presents the belief of the system inference for the jth referential value of consequent attritube D_j. β_D indicates the belief not assigned to any referential value of consequent attritube.

According to the belief distribution of consequent term and utility value $\mu = \{\mu_1, \mu_2, \ldots, \mu_N\}$, we can get the result utility:

$$\mu_{min} = (\beta_1 + \beta_D)\,\mu_1 + \sum_{j=2}^{N}\beta_j\mu_j \tag{14}$$

$$\mu_{max} = (\beta_N + \beta_D)\,\mu_N + \sum_{j=1}^{N-1}\beta_j\mu_j \tag{15}$$

$$\mu_{avg} = \frac{\mu_{min} + \mu_{max}}{2} \tag{16}$$

C. BRB PARAMETER TRAINING MODEL

When solving the complex decision problems, it is difficult for experts to give accurate system parameter values based on historical experience. Therefore, Yang *et al.* [18] proposed to use the prediction results of BRB system compared with the actual values to correct the system parameters, train the obtained belief degree β_j^k, rule weights θ_k, and antecedent attribute weights δ_i of BRB system. Chen *et al.* [19] proposed a global parameter optimization model by considering the referential values A_i of antecedent attritubes as parameters based on [18]. Based on the parameter optimization model in [19], the models of [20] and [21] also included the antecedent attribute weights, referential values of antecedent attributes, and referential values of consequent attributes into parameter training, and constructed a new parameter optimization model as shown in Figure 1.

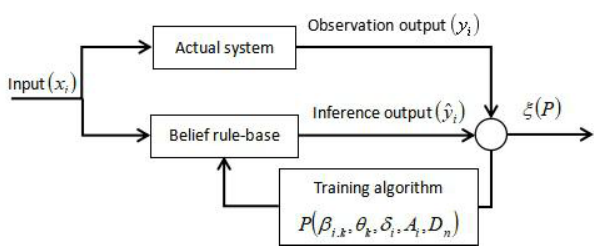

FIGURE 1. Global parameter optimization training model.

D. QUESTION PUTTING FORWARD

In the classical classification method of BRB, the combinatorial explosion problem of the state space is tricky. This question refers to the fact that when the size of the data is large, the combination of antecedent attributes will then increase exponentially. For example, when the dataset has 5 antecedent attributes and each antecedent attribute has 5 referential values, a total of 3125 rules are required to ensure that the BRB system covers all possible situations. If an antecedent attribute is added on this basis, and there are also 5 referential values, 15625 rules are required. The number of rules will increase extremely rapidly. This problem will become a time bottleneck in constructing BRB system, and will seriously slow down the efficiency of parameter training, and must be properly resolved.

In the current classification method of BRB, Jiao *et al.* [11] proposed that BRBCS, using fuzzy arrays to skip the selection step of the attribute referential set, and make the rule selection more objective, and eliminate the impact of the size of the attribute referential set on the rule base. However, this method only sets the relevant weight coefficients based on support degree and belief degree of data for the rule. It relies heavily on the quality of the dataset. For the complex parameter settings, it is difficult to fit the actual situation through intuitive weight calculation. The combinatorial explosion of the state space still exists. Ye *et al.* [16] proposed a linear combination of the referential values of antecedent attributes instead of the original Cartesian product combination. This method sets the number of rules according to the number of classification results that it is no longer limited by the number of attributes. However, when the method deals with complex classification problems, or when the value range of the antecedent attribute does not have a simple linear relationship with the type of the result, the method will cause obvious errors. And there are also certain subjective factors in the setting of the number of rules.

To this end, this paper proposes a new method for constructing BRB system with referential interval for classification on decision trees. The method extracts rule according to the classification result of the decision tree. Then using DE algorithm, the weights set by the data support degree and the belief degree are further trained during the construction of BRB system. While reducing the size of the system, the method also reasonably determines the scope of antecedent attributes. The new system can effectively avoid the influence of subjective factors, and thus improve the accuracy of the system.

III. THE PROPOSED METHOD FOR CONSTRUCTING CLASSIFICATION BRB BASED ON DECISION TREE

In this section, we will introduce the new proposed methods, including

1) Introduction of decision tree algorithms.
2) Introduction of improved BRB method based on decision tree algorithm. And the setting of related initial parameters.
3) Introduction of DE algorithm for parameter learning.

A. REPRESENTATION OF DECISION TREE

The decision tree model classifies the samples by using the probability information of the sample attributes and the branching strategy of the nodes. And the nodes in the tree represent partial samples, while the branches are represented as attribute partitions. The attribute value range of each leaf node corresponds to the interval of the attribute divided by the path from the root node to the leaf node. The whole tree reflects the intuitive relationship between attributes and classification results [22]. For each leaf node in the tree, the division of path from the root node to itself can extract a BRB rule. The range of referential value range of each antecedent attribute can be got on this division. The proposed method can also provide the relevant membership information for rule matching in inference.

Suppose an antecedent attribute A_i $(i = 1, \ldots T)$, the data sample value interval is $[a_i, b_i]$, and the attribute division point set after the discretization of the antecedent attribute is $Q_i = \left\{q_1^i, q_2^i, \ldots q_{N_i}^i\right\}$. The decision tree algorithm needs to select an antecedent attribute and a value to split the data. It's usually determined by information entropy in ID3 algorithm or information gain rate in C4.5 algorithm. The decision tree construction algorithm can be expressed in Figure 2.

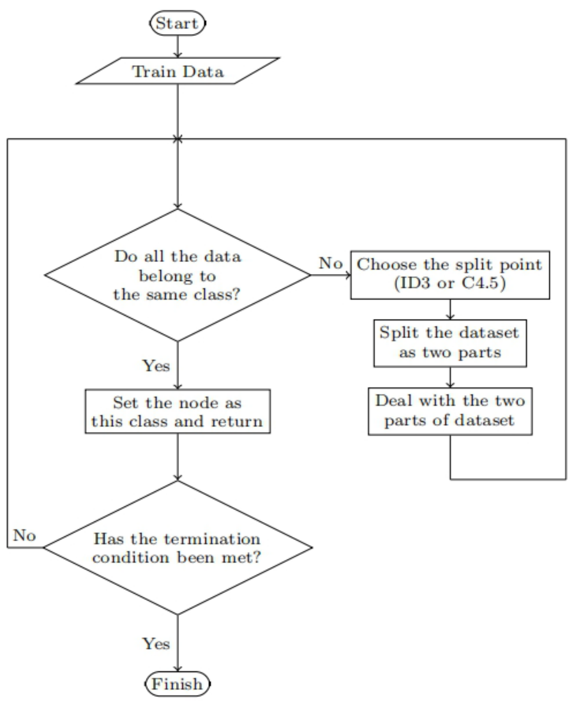

FIGURE 2. The algorithm flow chart of decision tree construction algorithm.

When constructing the decision tree from top to bottom, the C4.5 decision tree replaces the attribute with the attribute division point for the continuous antecedent attribute. It improves the split information of the ID3 decision tree using the information entropy as the branch strategy. The information gain rate to represent the split information of the attribute split point on the current node q_g^i is calculated as follow:

$$GainRatio\left(q_g^i\right) = \frac{Gain\left(q_g^i\right)}{SplitInfo\left(q_g^i\right)} \tag{17}$$

where $Gain\left(q_g^i\right)$ is the same as calculated in ID3 algorithm. It is represented by the difference between the category information entropy $Info\,(U)$ of the current node and the information entropy $Info_{q_g^i}\,(U)$ of the attribute division point q_g^i.

$$Gain\left(q_g^i\right) = Info\,(U) - Info_{q_g^i}\,(U) \tag{18}$$

The category information entropy is calculated by

$$Info\,(U) = -\sum_{i=1}^{m} p_i \log_2 (p_i) \tag{19}$$

where m is the current number of node categories, and p_i is the probability that the ith category is in the node.

Information entropy of attribute division point q_g^i is calculated by

$$Info_{q_g^i}\,(U) = \sum_{d=1}^{2} \left(\frac{|U_d|}{|U|} \times Info\,(U_d) \right) \tag{20}$$

The split information $SplitInfo\left(q_g^i\right)$ represents the breadth and uniformity of the split sample set according to the attribute q_g^i. It is defined as follows:

$$SplitInfo\left(q_g^i\right) = -\sum_{d=1}^{2} \left(\frac{|U_d|}{|U|} \times \log \frac{|U_d|}{|U|} \right) \tag{21}$$

The naive ID3 algorithm can only handle discrete attributes and tends to select the attribute with more referential values as the splitting attribute. Compared with the ID3 decision tree, the C4.5 decision tree can handle non-discrete feature data and effectively avoid the disadvantage in ID3 through the information gain rate. Decision tree algorithm provides a theoretical basis for the generation of BRB rules in the newly proposed method.

B. CONSTRUCTION METHOD OF BRB WITH REFERENTIAL INTERVAL

Based on the classification results and the characteristics of the decision tree, a new rule representation and the setting methods of parameters such as rule weight and belief distribution of consequent term are proposed in this paper. And then, a method for constructing classification BRB based on decision tree is proposed.

1) NEW PROPOSED RULE REPRESENTATION METHOD

According to the C4.5 decision tree constructed by training set U, each branch from the root node to a leaf node is taken to construct a rule in BRB. The single referential value of antecedent attribute A_i is replaced by a interval, which corresponds to the value range of leaf node attribute A_i in the decision tree branch $[a_i^k, b_i^k)$. The rule representation method of BRB system is modified as follows:

$$\begin{aligned} R_k:\ & if\ x_i \in [a_i^k, b_i^k); \quad i = 1, 2, \ldots, T_k \\ & then\ \{(D_j, \overline{\beta_j^k}); \quad j = 1, 2, \ldots, N\} \\ & with\ rule\ weight\ \theta_k\ and\ attribute\ weight\ \delta_i^k \end{aligned} \tag{22}$$

where a_i^k represents the lower bound of the ith attribute referential value interval of the kth rule, b_i^k represents the upper bound of the ith attribute referential value interval of the kth rule, and the definition of other parameters is the same as those in (1).

2) NEW PROPOSED RULE MATCHING METHOD

According to the newly proposed rule representation method, the similarity between the sample and the rule is measured by calculating the membership degree of the sample data and the range of the rule antecedent attribute value range. And then a new individual matching method is proposed. Suppose sample $X\left(x_1, x_2, \ldots, x_{T_k}\right)$, the equation for calculating the matching degree of the ith antecedent attribute corresponding to the kth rule is expressed by the usual symmetric membership function as follows:

$$\mu_i^k\,(x_i) = \frac{1}{1 + e^{-dis_i^k(x_i)}} \tag{23}$$

where $dis_i^k\,(x_i)$ represents the distance from x_i to ith antecedent attribute in the kth rule.

This function is similar to the sigmoid function, whose function expression is $f(x) = \frac{1}{1+e^{-x}}$. The graph of the sigmoid function is shown in Figure 3.

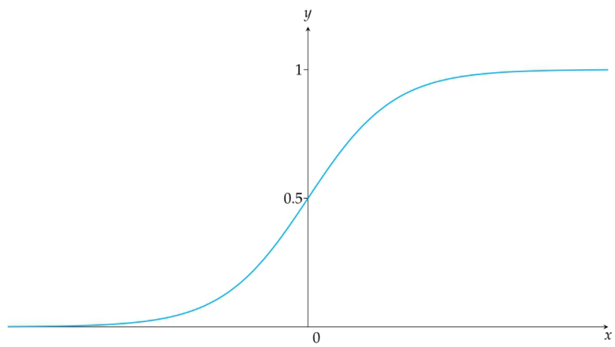

FIGURE 3. The graph of the sigmoid function.

As shown in the graph of the sigmoid function, we can know that this function is symmetrical about the point (0,0.5). In the process of x changes from 0 to positive infinity, the slope of the function becomes smaller and smaller.

The advantage of this property is related to the rules generated by the decision tree. In the decision tree algorithm, the influence of boundaries is considerable. As long as an item of data crosses the boundary, the data may be classified into other classes. In BRB, we reduce this effect by calculating the individual matching degree. But this effect cannot be completely eliminated. So this function was introduced in the newly proposed method.

The range of $dis_i^k(x_i)$ described below is [-1,1], which cannot cover the domain of the sigmoid function. So, it is necessary to expand $dis_i^k(x_i)$ proportionally so that the range of values for this function can be more close to (0,1).

The distance from x_i to the kth rule in the ith antecedent attribute can be calculated as the following equation:

$$dis_i^k(x_i) = \begin{cases} -\lambda\left|x_i - a_i^k\right| & lowMid_i^k \le x_i < a_i^k \\ \lambda\left|x_i - a_i^k\right| & a_i^k \le x_i < mid_i^k \\ \lambda\left|x_i - b_i^k\right| & mid_i^k \le x_i < b_i^k \\ -\lambda\left|x_i - b_i^k\right| & b_i^k \le x_i < upMid_i^k \\ -\infty & otherwise \end{cases} \tag{24}$$

where $i = 1, \ldots T_k$, and $k = 1, \ldots, L$. The membership function parameter λ is calculated according to the antecedent attribute division interval:

$$\lambda = \begin{cases} \dfrac{1}{a_i^k - lowMid_i^k} & lowMid_i^k \le x_i < a_i^k \\ \dfrac{1}{mid_i^k - a_i^k} & a_i^k \le x_i < b_i^k \\ \dfrac{1}{upMid_i^k - b_i^k} & b_i^k \le x_i < upMid_i^k \end{cases} \tag{25}$$

where mid_i^k is the intermediate value of the ith antecedent attribute referential value interval of the kth rule. After the referential value interval is divided according to the decision tree classification, we can get several adjacent continuous intervals on each antecedent attribute. $lowMid_i^k$ represents the intermediate values of the last interval adjacent to the ith antecedent attribute referential value interval of the kth rule. $upMid_i^k$ represents the intermediate values of the next interval adjacent to the ith antecedent attribute referential value interval of the kth rule.

Therefore, the total matching degree $\mu^k(X)$ of the sample $X\left(x_1, x_2, \ldots, x_{T_k}\right)$ for the kth rule is

$$\mu^k(X) = \prod_{i=1}^{T_k} \mu_i^k(x_i) \tag{26}$$

The activation weight of the sample $X\left(x_1, x_2, \ldots, x_{T_k}\right)$ corresponding to the kth rule is calculated according to the individual matching degree of the antecedent attribute:

$$\omega_k = \mu^k(X)\,\theta_k \tag{27}$$

where θ_k is the rule weight of the kth rule.

3) NEW PROPOSED BELIEF DISTRIBUTION OF CONSEQUENT TERM SETTING METHOD

The belief distribution of each consequent term in a rule is determined by the data samples supporting the rule. To utilize the ambiguity and uncertainty of the information in BRB, the support degree is based on the consistency and inconsistency of related data samples. According to [11], the belief distribution of the kth rule is set, and the number of consequent terms is equal to the number of classification classes:

$$\beta_m^k = \frac{1}{1-K^k}\left(1 - \prod_{X_i \in R_m^k}\left(1-\mu^k(X_i)\right)\right) \times \prod_{r \ne m}\prod_{X_i \in R_r^q}\left(1-\mu^k(X_i)\right), \quad m = 1, \ldots, M \tag{28}$$

$$\beta_\Omega^k = \frac{1}{1-K^k}\prod_{r=1}^{M}\prod_{X_i \in R_r^k}\left(1-\mu^k(X_i)\right) \tag{29}$$

R_m^k represents the subset of data supporting the kth rule in the sample dataset R^k. Moreover, the classification result of this subset of data is m. $\mu^k(X_i)$ denotes the matching degree of the sample X_i data calculated according to (26) for the kth rule, and M is the number of classification classes.

K^k represents the mass of the conflict probability about the sample dataset supporting the rule k relative to rule k. It is calculated as follows:

$$K^k = 1 + (M-1)\prod_{r=1}^{M}\prod_{X_i \in R_r^k}\left(1-\mu^k(X_i)\right) - \sum_{m=1}^{M}\prod_{r \ne m}\prod_{X_i \in R_r^k}\left(1-\mu^k(X_i)\right) \tag{30}$$

4) NEW RULE WEIGHT SETTING METHOD

The rule is derived from the corresponding branch of the decision tree. Therefore, when it is used to balance the importance degree of the rule in the classification, it should be determined by the support degree and the belief degree of the corresponding data belonging to the branch of the rule [11], which is calculated as follows:

$$\theta^k \propto c\left(R^k\right) \circ s\left(R^k\right), \quad k = 1, \ldots, L; \tag{31}$$

where $c\left(R^k\right)$ represents the belief degree of the dataset R^k supporting the kth rule on the rule: $c\left(R^k\right) = 1 - \overline{K^k}$. $\overline{K^k}$ represents the average conflict factor, which is calculated as follows:

$$\overline{K^k} \begin{cases} 0 & \left|R^k\right| = 1 \\ \dfrac{1}{\left|R^k\right|\left(\left|R^k\right|-1\right)} \times \sum_{\substack{X_i, X_j \in R^k \\ c(X_i) \ne c(X_j)}} \mu_i^k(X_i)\,\mu_j^k(X_j) & otherwise \end{cases} \tag{32}$$

where R^k represents the size of the sample dataset supporting the rule k. $s\left(R^k\right)$ represents the support degree of the dataset R^k supporting the kth rule: $s\left(R^k\right) = \frac{|R^k|}{N}$, where N is the total number of data samples. $c(X_i)$ represents the class label of ith data.

When the rule has a higher support degree, it can represent more data features and should have a higher weight. At the same time, to avoid the influence of the classification error of the decision tree on the system performance of BRB, the weight of the rule is positively correlated with the corresponding belief degree of the data. The equation for calculating the weight of the rule according to [11] is as follows:

$$\theta^k = \frac{c\left(R^k\right) \times s\left(R^k\right)}{max\left\{c\left(R^k\right) \times s\left(R^k\right)\right\}} \tag{33}$$

C. DE ALGORITHM FOR PARAMETER TRAINING METHOD

After determining the BRB model and initial parameters, we need to train the parameters so that it can better fit the distribution of the dataset.

DE algorithm is a heuristic global optimization technology based on population. This algorithm is mainly used to solve real number optimization problems. As a genetic algorithm, the DE algorithm is also an optimization algorithm based on modern intelligence theory, which guides the direction of optimization search through the swarm intelligence generated by cooperation and competition among individuals within the swarm. We explain the algorithm flow of the DE algorithm through the parameter optimization problem in this paper. According to the definition of (22), a parameter model for new proposed BRB can be formally represented as follows:

$$Q =< \theta, \delta, \beta > \tag{34}$$

where Q represents the parameter vector, θ is the vector of rule weights, δ is the vector of attributes weights, and β is the vector of given belief degrees. Other intermediate parameters are directly calculated by Q.

Then the parameter training will become the following optimization problems that need to be solved:

$$\begin{aligned}
&min\, f(Q, A, U)\\
&s.t.\ 0 \le \beta_j^k \le 1; \quad j = 1, 2, \ldots, N;\ k = 1, 2, \ldots, L;\\
&\quad \sum_{j=1}^{N} \beta_j^k = 1\\
&\quad 0 \le \theta_k \le 1; \quad k = 1, 2, \ldots, L;\\
&\quad 0 \le \delta_i^k \le 1; \quad i = 1, 2, \ldots, T_k;\ k = 1, 2, \ldots, L;
\end{aligned} \tag{35}$$

where f represents the objective function for parameter optimization, A is the vector of referential values of antecedent attributes, and U is the vector of referential values of consequent attributes. For the definition of other variables, see (22).

The flow of the DE algorithm is as follows:

1) Initialize population. In the DE algorithm, each individual in the population is a feasible solution for parameters. We can treat it as a D-dimensional solution vector. D represents the number of parameters. For each element in the vector, we randomly assign a value to it:

$$x_j = x_j^L + rand(0, 1) \times (x_j^U - x_j^L) \tag{36}$$

where x_j represents the j-th element in each individual. x_j^U represents the upper bound of the jth element. x_j^L represents the lower bound of the j-th element. $rand(0, 1)$ represents a random number between 0 and 1.

2) Mutation. DE algorithm realizes individual mutation through difference strategy. A common difference strategy is to select two different individuals and scale their vector difference with the individual to be mutated for vector synthesis:

$$v_i^{g+1} = x_a^g + F \times (x_b^g - x_c^g), \quad i \ne a \ne b \ne c \tag{37}$$

where v_i^{g+1} represents the i-th individual of the $(g+1)$-th population. x_a^g represents the a-th individual of the g-th population. x_b^g and x_c^g are the same. F is the scaling factor.

When an element of the solution vector in the mutation process is out of range. This element will be randomly generated again by the initialization step.

3) Cross operation. Perform individual cross operations on the gth generation population and the $(g+1)$th generation population resulting from mutation:

$$u_i^{g+1} = \begin{cases} v_i^{g+1} & if\ rand(0, 1) \le CR\ or\ j = j_{rand} \\ x_i^g & otherwise \end{cases} \tag{38}$$

where CR is the cross probability. j_{rand} is a random integer between 1 and D. The meaning of j_{rand} is to ensure that at least one of the mutated elements can be saved to the next generation population.

4) Selection. We use greedy strategies to select individuals of the next generation:

$$x_i^{g+1} = \begin{cases} u_i^{g+1} & if\ f(u_i^{g+1}, A, U) < f(x_i^g, A, U) \\ x_i^g & otherwise \end{cases} \tag{39}$$

Through the iteration of the above method, we get the best solution vector as our final BRB system parameters.

D. ALGORITHM FLOW

The rules extracted from the decision tree are used to construct the initial rules according to the above parameter setting method. Combined with the new rule matching method, the parameters are optimized by using the DE algorithm. Then the final classification BRB is constructed by the method described above.

The algorithm flow chart of the rule building method of classification BRB based on decision tree is shown in Figure 4.

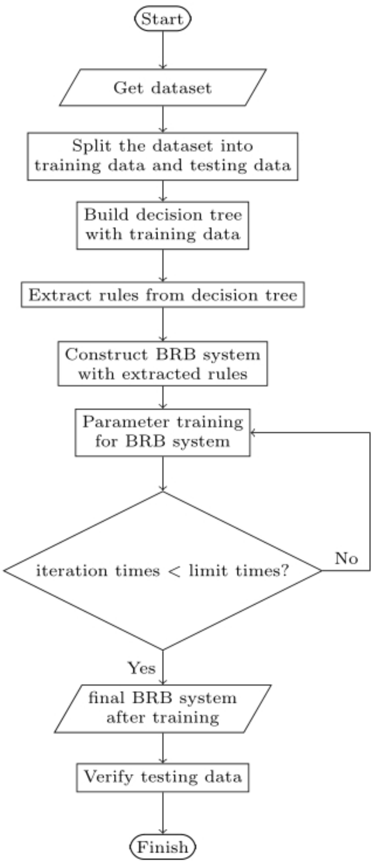

FIGURE 4. Algorithm flow chart of classification BRB based on C4.5 decision tree.

IV. EXPERIMENTAL RESULTS

In this section, we introduce the experimental part with three subsections. The first subsection introduces our experimental environment and the datasets we used. The second subsection introduces the effect of rule reduction in the new proposed BRB construction method. The third subsection introduces the accuracy comparison effect of the newly proposed method and other classification methods and summarizes the experiments.

A. EXPERIMENTAL ENVIRONMENT

The experiment runs on Intel(R) Core(TM) i7-6700 @ 3.40GHz CPU, 16GB RAM, Windows10 operating system environment. We write the algorithm in R and C++ language. In the parameter training, the DE algorithm is selected to train the parameters. The population of DE algorithm is 100 and the iteration number is 300.

The experimental dataset uses ten common classification datasets from UCI [23] (University of California at Irvin) website. Table 1 lists the number of antecedent attributes in the ten datasets, the number of classes, and the size of the dataset.

TABLE 1. Information of experimental datasets.

Dataset name	Attribute quantity	Class quantity	Data quantity
Pima	8	2	768
Mammographic	5	2	830
Bupa	6	2	345
Wine	13	3	178
Iris	4	3	150
Seeds	7	3	210
Contraceptive	9	3	1473
Glass	9	7	214
Ecoli	7	8	336
Yeast	8	10	1484

In the experiment, the correctness and robustness of the algorithm are verified by ten-fold cross-validation, which runs by randomly selecting 90% of the data set as the training data, and the remaining 10% as the testing data.

B. RULE REDUCTION EFFECT COMPARED WITH OTHER RULE-BASED SYSTEMS

The size of the BRB constructed by BRBCS and divided fuzzy sets proposed in [11] is affected by the number of divisions and the number of features in the dataset. Even if the number of divisions of the original text is small, when the number of features is large, the number of rules in BRBCS will still produce a combinatorial explosion of the state space.

Since the training data is different in the ten-fold cross-validation, the training trees constructed by the ten-fold cross-validation are different. The node division attributes will change, and the tree shape and the number of leaf nodes will also be slightly different. The average values for the number of rules are reflected in Table 2.

TABLE 2. Comparison of average rule quantities of BRBCS.

Dataset name	BRBCS	This paper
Pima	6561	23.9
Mammographic	243	22.4
Bupa	729	28.6
Wine	1594323	5.6
Iris	81	5.0
Seeds	2187	7.8
Contraceptive	19683	29.2
Glass	19683	34.5
Ecoli	2187	29.6
Yeast	6561	51.1

In extreme cases, some classes with few samples will be removed from the model due to the pruning process in the decision tree algorithm. For example, in some datasets, there is only one sample of a class, and then this data will be regarded as noise by the decision tree algorithm. The corresponding decision tree node will be deleted due to pruning. So the class will disappear in the model.

To prevent this issue, We oversampled the data of each class. It can ensure that the number of samples in each class is consistent. After that, we can use C4.5 algorithm to build the decision tree.

As can be seen from Table 2, in dataset Wine, when the number of attributes is 13, and the number of rules in BRBCS has exploded to 1594323. Under the condition of no attribute reduction, it is very difficult to train massive parameters. The size of the BRB built by the method mentioned above does not have satisfactory efficiency in both parameter training and testing. Besides, the excess rules produced by combinatorial states explosion can not be trained by swarm intelligence to improve the performance of the system. And most of the excessive rules do not have the support of data samples and are rarely activated in the final reasoning synthesis. However, this still costs matching operations when data matching rules, which wastes a lot of running time. Few of them are supported by very small samples, which may result in the conflict between rules and others because of very small noise data. And then, it affects the accuracy of the rules that are supported by a large number of data samples.

In this paper, the value range of rule number and antecedent attributes are divided by the decision tree. Under the premise of a large reduction of the rule base, the influence of rules conflict caused by individual noise or edge data on system performance can be avoided. And the efficiency of parameter training and the overall decision accuracy of the system are improved.

To verify the rule reduction effect of the new method, we compare it with the rule generation methods of some existing rule-based systems, including extended belief rule-based (EBRB) system [24], and fuzzy rule-based classification system (FRBCS) [25].

According to the results of Table 3, we can know that in the above data set, the number of rules by the method described in this paper is less than the other two methods in general. This method greatly reduces the number of rules in BRB system. The another advantage is that, in the decision tree algorithm, we can flexibly adjust the number of rules in the rule base by modifying the pruning threshold to meet our various needs for BRB system.

TABLE 3. Comparison of rules quantities of BRB.

Dataset name	FRBCS	EBRB	This paper
Ecoli	105.8	269	**29.6**
Glass	80.4	172	**34.5**
Iris	42.6	120	**5.0**
Pima	190.8	615	**23.9**
Seeds	96.8	168	**7.8**
Yeast	208	1188	**51.1**

C. ANALYSIS OF EXPERIMENTAL RESULTS

1) SELECTION OF PARAMETER TRAINING ALGORITHM

To ensure the smooth progress of parameter training, we need to define a loss function to evaluate the fitness of the model. In classification problems, the cross-entropy loss function is a commonly used loss function, which is defined as follows:

$$L = \frac{1}{N}\sum_{i=1}^{N} - \sum_{c=1}^{M} y_{ic} log(p_{ic}) \tag{40}$$

where N represents the number of samples. L represents the values of the function. M represents the class quantity of samples. y_{ic} is the indicator variable. When c equal to class of the ith sample, y_{ic} is 1 and 0 otherwise. p_{ic} represents the referential value that the model judges that ith sample belongs to class c.

In the choice of parameter training algorithm, we compared the three algorithms: DE algorithm, artificial bee colony (ABC) algorithm, and particle swarm optimization (PSO) algorithm. The classification accuracy is listed in Table 6. And an example of the loss function changes with the number of iterations on dataset Iris was given in Figure 5.

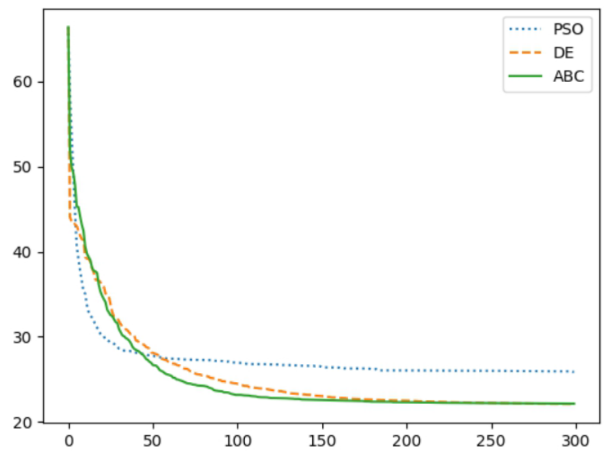

FIGURE 5. The loss function changes with the number of iterations on Iris dataset.

According to Figure 5, we find that when the accuracy is same, the convergence efficiency of PSO algorithm is worse than that of the other two algorithms. According to Table 6, we find that the accuracy of DE algorithm is better than that of ABC algorithm. So in the subsequent experiments, we choose DE algorithm for parameter training.

2) COMPARISON WITH CLASSICAL METHODS

To further verify the performance of the proposed BRB learning method, the performance of this paper is compared with the performance of the classical algorithms. K-nearest neighbour (KNN), naive Bayes (NB), C4.5, and support vector machine (SVM) are used to be compared in this paper.

TABLE 4. Comparison of the average accuracy for different classical classification methods.

Dataset	Classification Method				
	KNN	NB	C4.5	SVM	This paper
Mammographic(%)	78.46(4)	78.36(5)	81.20(2)	79.29(3)	**81.93(1)**
Iris(%)	**96.67(1)**	96.00(3)	94.67(4)	**96.67(1)**	96.00(3)
Seeds(%)	92.38(2)	91.43(3)	90.48(4)	90.48(4)	**94.29(1)**
Glass(%)	66.36(4)	48.60(5)	67.36(3)	68.69(2)	**72.92(1)**
Ecoli(%)	**85.71(1)**	85.42(2)	84.23(3)	75.60(5)	82.44(4)
Yeast(%)	**58.22(1)**	57.61(3)	55.39(4)	43.26(5)	57.82(2)
Average rank	2.2(2)	3.5(5)	3.3(3)	3.3(3)	**2.0(1)**

TABLE 5. Comparison of the average accuracy for novel BRB methods.

Dataset	Classification Method				
	EBRB	SRA-EBRB	VP-EBRB	BA-EBRB	This paper
Pima(%)	70.87(5)	71.71(4)	74.61(2)	72.80(3)	**75.01(1)**
Mammographic(%)	79.67(5)	**82.53(1)**	80.95(3)	79.81(4)	81.93(2)
Bupa(%)	65.02(5)	**70.46(1)**	66.67(4)	69.20(2)	68.42(3)
Wine(%)	96.32(4)	96.85(2)	93.01(5)	**97.02(1)**	96.60(3)
Iris(%)	95.26(4)	94.80(5)	**96.27(1)**	95.26(3)	96.00(2)
Seeds(%)	91.33(4)	91.24(5)	92.48(3)	93.95(2)	**94.29(1)**
Glass(%)	67.85(5)	**73.08(1)**	71.03(4)	72.32(3)	72.92(2)
Yeast(%)	45.61(5)	56.85(4)	**58.83(1)**	58.63(2)	57.82(3)
Average rank	3.7(5)	2.3(3)	2.3(3)	2.0(2)	**1.9(1)**

TABLE 6. Comparison of the average accuracy for different parameter training algorithms.

Dataset name	DE	PSO	ABC
Pima(%)	**75.01**	74.75	74.88
Mammographic(%)	**81.93**	81.56	81.69
Bupa(%)	**68.42**	64.38	68.12
Wine(%)	**96.60**	93.20	**96.60**
Iris(%)	**96.00**	**96.00**	**96.00**
Seeds(%)	**94.29**	90.95	93.81
Contraceptive(%)	**54.97**	53.81	54.83
Glass(%)	**72.92**	67.34	71.54
Ecoli(%)	**82.44**	81.57	81.83
Yeast(%)	**57.82**	55.86	57.14

The comparison results are listed in Table 4. Among them, the experimental results of C4.5 are obtained through our experiments. Other experimental results are cited from [26].

According to Table 4, the proposed algorithm has shown the highest accuracy on the Mammograohic, Seeds, and Glass. In the dataset with dense data distribution and little difference in antecedent attribute values, since the dataset is more complicated, there is no simple linear correlation between the antecedent attribute and the classification result. Therefore, there is a certain difference in accuracy by the linear combination algorithm [16]. On the other side, the proposed algorithm extracts the antecedent attribute interval from the decision tree to construct the rules, which effectively improves the accuracy of the BRB system in the classification. In the dataset with a large number of antecedent attribute, both the algorithm in [11] and the algorithm in [15] need to divide the fuzzy interval construction rules due to the expansion of the attribute scale. It is difficult to construct effective parameter training for the massive rules. In this paper, the rule constructed by the decision tree reduces most of the useless rules under the data clustering of the decision tree, and then improves the efficiency of parameter training and the performance of the classification BRB effectively.

3) COMPARISON WITH NOVEL BRB SYSTEMS

To further prove the effectiveness of the proposed method, the results of the proposed method are also compared with several novel BRB systems. Those approaches for comparison are EBRB [24], SRA-EBRB [27], VP-EBRB [28] and BA-EBRB [29]. The comparison results were listed in Table 5.

Although the proposed method cannot achieve the best accuracy in most datasets, its ranking in comparison is relatively stable, and the average ranking is also the best among them. But compared with the number of rules of EBRB, the number of rules generated by the proposed method is greatly reduced. And the proposed method still does not fail in accuracy. The results of the experiment can show that the proposed method is a powerful method that can effectively optimize the size of BRB system. However, the accuracy of BRB system generated based on the decision tree is more dependent on the distribution of the datasets.

V. CONCLUSION

This paper proposes a method for building a BRB system based on the decision tree and modifies the classical BRB so that the antecedent attritubes of rule can be expressed by intervals instead of a single value. On this basis, the regular individual matching method is improved, and the sigmoid function is introduced to the method for calculating individual matching degree, making it closer to the meaning of the model. After that, we appropriately select the other parameters of the rule base. By comparing the efficiency of the three swarm intelligent algorithms in the proposed BRB system, the DE algorithm is used to optimize the initial set of rule parameters most effectively. The performance of the proposed algorithm is verified on the several UCI public classification datasets, and the superiority of the algorithm is verified by comparing with other existing classification methods.

BRB has advantages in dealing with fuzzy and uncertain information. It can effectively use fuzzy information and uncertain information that cannot be used by other classical classification algorithms. Therefore, the proposed method can improve the performance of the classifier when solving classification problems. But it is also directly affected by the data, a small number of noise data and edge data may have a significant impact on the overall accuracy of the system. Therefore, the important direction of future research includes:

1) How to deal with the influence of the noise data and the edge data on the system in the classification problem?
2) Is there a better parameter learning algorithm?

REFERENCES

[1] C.-X. Cui, *Intelligent Classification Method*. Beijing, China: Meteorology Press, 2009.
[2] A. P. Dempster, "A generalization of Bayesian inference," *J. Roy. Stat. Soc. B, Methodol.*, vol. 30, no. 2, pp. 205–232, 1968.
[3] G. Shafer, *A Mathematical Theory of Evidence*. Princeton, NJ, USA: Princeton Univ. Press, 1976.
[4] C.-L. Huang and K. Yong, *Multiple Attribute Decision Making Methods and Applications. A State-of Art Survey*. Berlin, Germany: Springer, 1981.
[5] L. A. Zadeh, "Fuzzy sets," *Inf. Control*, vol. 8, no. 3, pp. 338–353, Jun. 1965.
[6] R. Sun, "Robust reasoning: Integrating rule-based and similarity-based reasoning," *Artif. Intell.*, vol. 75, no. 2, pp. 241–295, Jun. 1995.
[7] J.-B. Yang, J. Liu, J. Wang, H.-S. Sii, and H.-W. Wang, "Belief rule-base inference methodology using the evidential reasoning approach-RIMER," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 36, no. 2, pp. 266–285, Mar. 2006.
[8] D.-L. Xu, J. Liu, J.-B. Yang, G.-P. Liu, J. Wang, I. Jenkinson, and J. Ren, "Inference and learning methodology of belief-rule-based expert system for pipeline leak detection," *Expert Syst. Appl.*, vol. 32, no. 1, pp. 103–113, Jan. 2007.
[9] Y.-Z. Lv, M.-Q. Xiao, Q. Liu, X.-L. Tang, and J.-F. Li, "Fault diagnosis of flight control system based on belief rule base reasoning," *Comput. Meas. Control*, vol. 27, no. 6, pp. 13–17, 2019.
[10] Y. He, L.-L. Chang, J. Jiang, and Y.-J. Tan, "Intension identification in air defense based on belief rule base expert system under expert guidance," *Fire Control Command Control*, vol. 42, no. 9, pp. 7–12, 2017.
[11] L. Jiao, T. Denoeux, and Q. Pan, "A hybrid belief rule-based classification system based on uncertain training data and expert knowledge," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 46, no. 12, pp. 1711–1723, Dec. 2016.
[12] Z. Chi, H. Yan, and T. Pham, *Fuzzy Algorithms: With Applications to Image Processing and Pattern Recognition*. Singapore: World Scientific, 1996.
[13] H. Ishibuchi, K. Nozaki, and H. Tanaka, "Distributed representation of fuzzy rules and its application to pattern classification," *Fuzzy Sets Syst.*, vol. 52, no. 1, pp. 21–32, Nov. 1992.
[14] C.-T. Sun, *Rule-Base Structure Identification in an Adaptive-Network-Based Fuzzy Inference System*. Piscataway, NJ, USA: IEEE Press, 1994.
[15] W.-L. Liu, H.-J. Wang, Y.-G. Fu, L.-H. Yang, and Y.-J. Wu, "Belief rule based inference methodology for classification based on differential evolution algorithm," *J. China Univ. Sci. Technol.*, vol. 46, no. 9, pp. 764–773, 2016.
[16] Q.-Q. Ye, L. H. Yang, Y. G. Fu, and X. C. Chen, "Classification approach based on improved belief rule-base reasoning," *Comput. Sci. Explor.*, vol. 10, no. 5, pp. 709–721, 2016.
[17] Y.-M. Wang, J.-B. Yang, D.-L. Xu, and K.-S. Chin, "The evidential reasoning approach for multiple attribute decision analysis using interval belief degrees," *Eur. J. Oper. Res.*, vol. 175, no. 1, pp. 35–66, Nov. 2006.
[18] J.-B. Yang, J. Liu, D.-L. Xu, J. Wang, and H. Wang, "Optimization models for training belief-rule-based systems," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 37, no. 4, pp. 569–585, Jul. 2007.
[19] Y.-W. Chen, J.-B. Yang, D.-L. Xu, Z.-J. Zhou, and D.-W. Tang, "Inference analysis and adaptive training for belief rule based systems," *Expert Syst. Appl.*, vol. 38, no. 10, pp. 12845–12860, Sep. 2011.
[20] W.-K. Wu, L.-H. Yang, Y.-G. Fu, L.-Q. Zhang, and X.-T. Gong, "Parameter training approach for belief rule base using the accelerating of Gra-dient algorithm," *Comput. Sci. Explor.*, vol. 8, no. 8, pp. 989–1001, 2014.
[21] Q. Su, Y. Longhao, and F. Yanggeng, "Parameter training approach based on variable particle swarm optimization for belief rule base," *Comput. Appl.*, vol. 34, no. 8, pp. 2161–2174, 2014.
[22] Q. J. Ross, *C4.5: Programs for Machine Learning*. San Diego, CA, USA: Morgan Kaufmann, 1993.
[23] D. Dua and C. Graff. *UCI Machine Learning Repository*. Accessed: 2019. [Online]. Available: http://archive.ics.uci.edu/ml
[24] J. Liu, L. Martinez, A. Calzada, and H. Wang, "A novel belief rule base representation, generation and its inference methodology," *Knowl.-Based Syst.*, vol. 53, pp. 129–141, Nov. 2013.
[25] L. Jiao, Q. Pan, T. Denœux, Y. Liang, and X. Feng, "Belief rule-based classification system: Extension of FRBCS in belief functions framework," *Inf. Sci.*, vol. 309, pp. 26–49, Jul. 2015.
[26] H. Zhu, M. Xiao, L. Yang, X. Tang, Y. Liang, and J. Li, "A minimum centre distance rule activation method for extended belief rule-based classification systems," *Appl. Soft Comput.*, vol. 91, Jun. 2020, Art. no. 106214.
[27] F. Wu, W. Wang, Y. Yang, Y. Zhuang, and F. Nie, "Classification by semi-supervised discriminative regularization," *Neurocomputing*, vol. 73, nos. 10–12, pp. 1641–1651, Jun. 2010.
[28] Y.-Q. Lin, Y.-G. Fu, Q. Su, Y.-M. Wang, and X.-T. Gong, "A rule activation method for extended belief rule base with VP-tree and MVP-tree," *J. Intell. Fuzzy Syst.*, vol. 33, no. 6, pp. 3695–3705, Nov. 2017.
[29] W. Fang, X. Gong, G. Liu, Y. Wu, and Y. Fu, "A balance adjusting approach of extended Belief-Rule-Based system for imbalanced classification problem," *IEEE Access*, vol. 8, pp. 41201–41212, 2020.
[30] L. Chang, Z. Zhou, Y. You, L. Yang, and Z. Zhou, "Belief rule based expert system for classification problems with new rule activation and weight calculation procedures," *Inf. Sci.*, vol. 336, pp. 75–91, Apr. 2016.
[31] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Ann. Eugenics*, vol. 7, no. 2, pp. 179–188, Sep. 1936.
[32] M. Aci and M. Avci, "K nearest neighbor reinforced expectation maximization method," *Expert Syst. Appl.*, vol. 38, no. 10, pp. 12585–12591, Sep. 2011.
[33] V. Athitsos and S. Sclaroff, "Boosting nearest neighbor classifiers for multiclass recognition," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR) Workshops*, Sep. 2005, p. 45.
[34] M. Fallahnezhad, M. H. Moradi, and S. Zaferanlouei, "A hybrid higher order neural classifier for handling classification problems," *Expert Syst. Appl.*, vol. 38, no. 1, pp. 386–393, Jan. 2011.

[35] T. Denœux, "Decision-making with belief functions: A review," *Int. J. Approx. Reasoning*, vol. 109, pp. 87–110, Jun. 2019.

[36] A. Calzada, J. Liu, H. Wang, and A. Kashyap, "A new dynamic rule activation method for extended belief rule-based systems," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 4, pp. 880–894, Apr. 2015.

[37] Y.-N. Wang, "Parameter learning for an intuitionistic fuzzy belief rule-based systems based on weight and reliability," *J. Adv. Comput. Intell. Intell. Informat.*, vol. 23, no. 2, pp. 219–228, Mar. 2019.

[38] I. Douven, "Optimizing group learning: An evolutionary computing approach," *Artif. Intell.*, vol. 275, pp. 235–251, Oct. 2019.

[39] L.-H. Yang, J. Liu, Y.-M. Wang, and L. Martínez, "New activation weight calculation and parameter optimization for extended belief rule-based system based on sensitivity analysis," *Knowl. Inf. Syst.*, vol. 60, no. 2, pp. 837–878, Aug. 2019.

[40] Y. Lin and Y. Fu, "A rule activation method for extended belief rule base based on improved similarity measures," *J. Univ. Sci. Technol. China*, vol. 48, no. 1, pp. 20–27, 2018.

YANGGENG FU received the Ph.D. degree from Fuzhou University, Fuzhou, China, in 2013. He is currently an Associate Professor of computer science with Fuzhou University. His research interests include data mining, machine learning, and intelligent decision support systems.

ZEFENG YIN received the B.S. degree from Fuzhou University, in 2018, where he is currently pursuing the master's degree. His research interests include intelligent decision technology, rule-based inference, big data analysis, and machine learning.

MANNA SU received the B.S. degree from Fuzhou University, in 2016, where she is currently pursuing the master's degree. Her research interests include intelligent decision technology, rule-based inference, and data mining.

YINGJIE WU received the Ph.D. degree from Southeast University, China, in 2012. He is currently a Professor of computer science with Fuzhou University, Fuzhou, China. His research interests include data mining, data security, privacy protection, and so on.

GENGGENG LIU (Member, IEEE) received the B.S. degree in computer science and the Ph.D. degree in applied mathematics from Fuzhou University, Fuzhou, China, in 2009 and 2015, respectively. He is currently an Associate Professor with the College of Mathematics and Computer Science, Fuzhou University. His research interest includes computational intelligence and its application.