Received January 27, 2020, accepted February 21, 2020, date of publication February 27, 2020, date of current version March 10, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2976708

A Balance Adjusting Approach of Extended Belief-Rule-Based System for Imbalanced Classification Problem

WEIJIE FANG [1], XIAOTING GONG [2], GENGGENG LIU [1], YINGJIE WU [1], AND YANGGENG FU [1]

[1]College of Mathematics and Computer Science, Fuzhou University, Fuzhou 350116, China
[2]College of Economics and Management, Fuzhou University, Fuzhou 350116, China

Corresponding author: Yanggeng Fu (ygfu@qq.com)

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

ABSTRACT The extended belief-rule-based (EBRB) system has become a widely recognized and effective rule-based system in decision-making. The system uses a data-driven method to generate the rule base by transforming each training sample into a rule. Hence, when an EBRB system is applied in an imbalanced classification dataset, the imbalance of training dataset will retain in the generated rule base. More specifically, the number of rules transformed from majority classes will be far greater than the rules transformed from minority classes. This issue usually leads to a sharp decrease in the accuracies of minority classes. This study analyses how the imbalance of training dataset exists in the generated EBRB and then proposes a Balance Adjusting (BA) approach to eliminate the influence of imbalance in the rule base. The BA approach adjusts rule activation weights of all activated rules, and further enhances the competitiveness of rules with higher activation weight during the rule aggregation process of the EBRB system. Several case studies in imbalanced benchmark classification datasets from UCI demonstrate how the use of the BA approach improves the performance of the EBRB system. This study also conducts a series of experiments to validate the improvement of the proposed approach compared with some conventional and recent existing works. The comparison results illustrate that the BA approach is feasible, effective and robust, and it performs well especially in large scale datasets. Moreover, the BA approach can also combine with various rule activation weight calculation methods, which means it might worth to be applied as a generic process before the rule aggregation process of the EBRB system.

INDEX TERMS Extended belief-rule-based system, rule activation weight, imbalanced classification problem.

I. INTRODUCTION

Data imbalance is a common problem occurs in classification. An imbalanced dataset consists of majority classes and minority classes. The former has more samples while the latter has less. Almost each of the existing decision models has a bias towards majority classes when applied in imbalanced classification datasets, unless the model has a special mechanism to address them. However, the accuracies of majority classes are usually less important than the accuracies of minority classes in data-imbalanced fields. For example, in disease surveillance, assume that 95% potential patients are healthy and the other 5% are ill. If a decision model infers that all the potential patients are healthy, it will obtain a high average accuracy of 95% without finding any true patient. Such a result is obviously meaningless and shall be improved.

The rule-based system is a kind of rapid developing and powerful decision support system, and it has become an important branch of artificial intelligence [1]. Generally, a rule-based system uses rules as its knowledge representation scheme, and uses a reasoning approach to infer the result of queries by activating and aggregating rules. The belief rule-based (BRB) system is a kind of rule-based system that embeds belief degrees in the consequent term of each rule,

The associate editor coordinating the review of this manuscript and approving it for publication was Valentina E. Balas.

and it has been widely applied in many fields [2]–[4]. It can handle both quantitative and qualitative information, and is considered to be more interpretable than deep-learning-based tools. The inference methodology of BRB system is the belief rule-base inference methodology using the evidential reasoning (RIMER) approach [5]. Recent years, a novel BRB system was proposed and studied [6]–[8], called extended belief-rule-based (EBRB) system. This system extends the conventional BRB system with belief degrees embedded in all the antecedent terms of each rule. Generally, the EBRB system uses the data-driven method to automatically generate rules from numerical data, without involving complex learning procedure to set parameters of the system. So the construction of the EBRB system is more steady and less time-consuming compared with other forms of BRB systems.

Although the data-driven method has many advantages, it brings an issue when applied in imbalanced datasets. Since each training sample will be transformed into a rule by the data-driven method, the number of rules belong to each class will also be the same as the number of training samples belong to each class, which means that the number of rules which belong to majority classes will be far greater than the number of rules which belong to majority classes. Such kind of rule base is called **imbalanced rule base** in this study. Besides, the performance of the RIMER methodology is similar to the weighted averages methodology when handling a large number of rules. Hence, even having a higher activation weight, the rules belong to minority classes in an imbalanced rule base will always be at a disadvantage during the rule aggregation process. So the reasoning result of an imbalanced EBRB system always has a bias towards majority classes. In order to fix this issue, this study proposes a Balance Adjusting (BA) approach. The approach increases the ratio of higher activation weights to lower activation weights, balances the sums of rule activation weights belong to the majority and the minority classes during the rule aggregation process, and thus further enhances the competitiveness of rules with higher activation weight. The BA approach improves the performance of EBRB system applied in imbalanced classification problems, and also proves that the system itself is a powerful tool for classification but limited by the imbalance in its rule base.

The remainder of this paper is organized as follows: Section II contains an overview about the BRB system and the EBRB system. Section III analyses the issue of the EBRB system applied in imbalanced classification datasets and then proposes the BA approach. Section IV demonstrates how the effect of the BA approach by case studies, and how to apply it to an EBRB system. Section V validates the effectiveness of the BA approach on various benchmark datasets compared with some recent existing works. Section VI is the conclusion of this paper.

II. OVERVIEW OF BRB SYSTEM AND EBRB SYSTEM

This section is a brief overview of the BRB system and the EBRB system. The overview of the EBRB system will be explained in three parts, i.e., how the systems represent knowledge, generate rules and obtain reasoning results. Some examples are provided in this section to help understand the overview.

A. OVERVIEW OF BRB SYSTEM

The BRB system is a kind of rule-based system, developed from the conventional IF-THEN rule-based system [9]. It can handle quantitative and qualitative knowledge, and deal with fuzzy uncertainty and probability uncertainty. The reasoning methodology of BRB system is the RIMER approach, which is based on D-S evidence theory, decision theory [10] and fuzzy theory [11].

Assume a BRB has L rules, T antecedent attributes and N referential values of its consequent attribute. The kth ($k = 1, \ldots, L$) rule in this BRB can be written as:

$$\begin{aligned} R_k: &\ \textit{if } U_i \textit{ is } A_{ij};\ i = 1, 2, \ldots, T \\ &\ \textit{then } D \textit{ is } \{(D_n, \beta_n^k);\ n = 1, 2, \ldots, N\} \end{aligned} \tag{1}$$

In (1), U_i is the ith antecedent attribute of this BRB, and A_{ij} is the jth ($j = 1, \ldots, J_i$) referential value of U_i. D is the consequent attribute of this BRB, D_n is the nth ($n = 1, \ldots, N$) referential value of D, and β_n^k is the belief degree to which D is evaluated to be D_n in the kth rule. For example, a belief rule for the Iris [12] classification may be written as a linguistic form (2a) or a numerical form (2b):

$$\begin{aligned} R_k: &\ \textit{if } U_1 \textit{ is Low} \wedge U_2 \textit{ is Medium} \\ &\ \textit{then } D \textit{ is } \{(\textit{Setosa}, 0.5), (\textit{Versicolor}, 0.3), \\ &\ (\textit{Virgica}, 0.2)\} \end{aligned} \tag{2a}$$

$$\begin{aligned} R_k: &\ \textit{if } U_1 \textit{ is } 4.3 \wedge U_2 \textit{ is } 3.2 \\ &\ \textit{then } D \textit{ is } \{(\textit{Setosa}, 0.5), (\textit{Versicolor}, 0.3), \\ &\ (\textit{Virgica}, 0.2)\} \end{aligned} \tag{2b}$$

In all existing kinds of BRB system, if $\sum_{n=1}^{N} \beta_n^k = 1$, then the kth rule is called complete, and if $\sum_{n=1}^{N} \beta_n^k < 1$, then the kth rule is called incomplete. For any rule in a BRB, $\sum_{n=1}^{N} \beta_n \leq 1$ and $\forall n \in N$, $\beta_n \geq 0$.

To distinguish the importance of different antecedent attributes and different rules, the BRB system uses δ_i to represent the weight of the ith antecedent attribute and θ_k to represent the weight of the kth rule. These parameters are called the system parameter of the BRB system.

The researches of the BRB system can be divided into three part: (1) parameter training optimization: mainly about seeking the optimal value of the system parameters [13]–[15]; (2) structure optimization: mainly about improving the rule retrieval efficiency and reasoning efficiency of the BRB system [16], [17]; (3) reasoning approach optimization: mainly about improving the rule activation method and reasoning approach of the BRB system [18], [19]. This study is a reasoning approach optimization that aims to improve the effectiveness of the EBRB system applied in imbalanced classification datasets.

B. EXTENDED BELIEF RULE

The extended belief rule is an extension of the conventional belief rule [6], and each antecedent attribute in the rule is a belief distribution but no referential value. Assume an EBRB has L rules, T antecedent attributes and N referential values of its consequent attribute. The kth ($k = 1, \ldots, L$) extended belief rule in this EBRB can be written as:

$$\begin{aligned} R_k : \ & \textit{if}\ U_i\ \textit{is}\ \{(A_{ij}, \alpha_{ij}^k);\ j = 1, 2, \ldots, J_i\}; \\ & i = 1, 2, \ldots, T \\ & \textit{then}\ D\ \textit{is}\ \{(D_n, \beta_n^k);\ n = 1, 2, \ldots, N\} \end{aligned} \tag{3}$$

In (3), U_i is the ith antecedent attribute of this EBRB, A_{ij} is the jth referential value of U_i, and α_{ij}^k is the belief degree to which U_i is evaluated to be A_{ij} in the kth rule. D is the consequent attribute of this EBRB, D_n is the nth ($n = 1, \ldots, N$) referential value of D, and β_n^k is the belief degree to which D is evaluated to be D_n in the kth rule. For example, an extended belief rule for the Iris [12] classification may be written as (4):

$$\begin{aligned} R_k : \ & \textit{if}\ U_1\ \textit{is}\ \{(Low, 0.8), (Medium, 0.2), \\ & (High, 0.0)\} \\ & \wedge\ U_2\ \textit{is}\ \{(Low, 0.1), (Medium, 0.9), \\ & (High, 0.0)\} \\ & \textit{then}\ D\ \textit{is}\ \{(Setosa, 1), (Versicolor, 0), \\ & (Virgica, 0)\} \end{aligned} \tag{4}$$

Due to the data-driven construction method, when an EBRB system is applied in classification problems, the belief degrees of consequent attribute's referential values in its extended belief rule will be either 1 or 0.

C. THE DATA-DRIVEN CONSTRUCTION METHOD OF EBRB SYSTEM

The most frequently used data-driven construction method of EBRB system is the utility-based transformation method. By this method, each numerical training sample will be transformed into a rule. The following example illustrates how the method works.

The first step of the method is to determine the referential values of each antecedent and consequent attribute. Generally, the referential values of each antecedent attribute are arranged incrementally, i.e., for the ith antecedent attribute, all the J_i referential values are arranged as $A_{i1} < A_{i2} < \ldots < A_{iJ_i}$. Assume the value of the ith antecedent attribute U_i in a training sample is x_i, then let α_{ij} be the belief degree to which x_i is evaluated to U_i's jth referential value A_{ij}, and then x_i can be represented using the following equivalent expectation [6]:

$$E(x_i) = \{(A_{ij}, \alpha_{ij});\ j = 1, 2, \ldots, J_i)\} \tag{5}$$

Suppose $t \in [1, J_i]$ and $A_{it} \leq x_i \leq A_{i(t+1)}$, then using utility-based equivalence transformation techniques [20] the belief degrees are generated as follow:

$$\alpha_{ij} = \begin{cases} \dfrac{A_{i(t+1)} - x_i}{A_{i(t+1)} - A_{it}} & j = t \\ 1 - \alpha_{it} & j = t + 1 \\ 0 & j \neq t, t + 1 \end{cases} \tag{6}$$

All these belief degrees form the belief distribution of U_i. The belief distribution of the consequent attribute can be generated in the same way according to the output of the training sample. In classification problems, the consequent attribute's referential values of EBRB system are arranged as the possible results of classification. As the output of each sample is one of those results, the belief degree to the output will be 1 and the others will be 0.

For example, suppose an EBRB for the Iris classification is arranged as:

$$\begin{aligned} U_1 &= \{Low(4.3), Medium(6.1), High(7.9)\} \\ U_2 &= \{Low(2.0), Medium(3.2), High(4.4)\} \\ D &= \{Setosa, Versicolor, Virgica\} \end{aligned} \tag{7}$$

Then a training sample $x = (6.1, 2.6, Virgica)$ will be transformed into the follow rule:

$$\begin{aligned} R_k : \ & \textit{if}\ U_1\ \textit{is}\ \{(Low, 0.0), (Medium, 1.0), \\ & (High, 0.0)\} \\ & \wedge\ U_2\ \textit{is}\ \{(Low, 0.5), (Medium, 0.5), \\ & (High, 0.0)\} \\ & \textit{then}\ D\ \textit{is}\ \{(Setosa, 0), (Versicolor, 0), \\ & (Virgica, 1)\} \end{aligned} \tag{8}$$

D. THE EVIDENTIAL REASONING APPROACH OF EBRB SYSTEM

The reasoning approach of EBRB system is similar to the one of the conventional BRB system. First, the antecedent attributes of the testing sample should also be transformed into belief distribution using the data-driven method, and then calculate the distance between the testing sample and each rule in EBRB. The Euclidean distance between input x_i's belief distribution $\{(A_{ij}, \alpha_{ij})\}$ and antecedent attribute U_i's belief distribution in the kth rule $\{(A_{ij}, \alpha_{ij}^k)\}$ is calculated as follow:

$$d(x_i, U_i)^k = \sqrt{\frac{\sum_{j=1}^{J_i} (\alpha_{ij} - \alpha_{ij}^k)^2}{2}} \tag{9}$$

Base on (9), the individual matching degree of x_i and U_i in the kth rule is:

$$S_i^k = 1 - d(x_i, U_i)^k \tag{10}$$

Base on the individual matching degrees, the conventional method to calculate activation weight ω^k for the kth rule is:

$$\omega^k = \frac{\theta_k * \prod_{i=1}^{T} (S_i^k)^{\bar{\delta}_i}}{\sum_{l=1}^{L} \left[\theta_l * \prod_{j=1}^{T} (S_j^l)^{\bar{\delta}_j}\right]} \tag{11}$$

where

$$\bar{\delta_i} = \frac{\delta_i}{\max\limits_{i=1,2,\ldots,T}\{\delta_i\}} \tag{12}$$

It is apparent that $0 \leq \omega_k \leq 1$, and $\sum_{i=1}^{L} \omega_i = 1$. After calculating the activation weight of each rule, all activated rules are aggregated using the RIMER analytical algorithm [21]. The reasoning conclusion of EBRB system is also a belief distribution represented as:

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

The belief degree $\beta_n(x)$ in $D(x)$ is generated as follows:

$$\beta_n(x) = \frac{\mu * \left[\prod_{k=1}^{L}(\omega_k \beta_i^k + \eta) - \prod_{k=1}^{L} \eta\right.}{1 - \mu * \left[\prod_{k=1}^{L}(1-\omega_k)\right]} \tag{14}$$

where

$$\eta = 1 - \omega_k \sum_{j=1}^{N} \beta_j^k \tag{15a}$$

$$\mu = \left[\sum_{j=1}^{N}\prod_{k=1}^{L}(\omega_k \beta_i^k + \eta) - (N-1)\prod_{k=1}^{L}\eta\right]^{-1} \tag{15b}$$

For classification problems, the referential value with the highest belief degree is generally regarded as the expected output of EBRB system, i.e.:

$$f_C(x) = D_c,\ c = \underset{n=1,2,\ldots,N}{\arg\max}\{\beta_n(x)\} \tag{16}$$

III. THE BALANCE ADJUSTING APPROACH

This section analyses the issue of a concrete imbalanced rule base and then proposes the BA approach to address the issue.

A. SIMILARITY BETWEEN RIMER METHODOLOGY AND WEIGHTED AVERAGES METHODOLOGY

Evidential reasoning (ER) approach was proposed by Yang and Singh [22] based on D-S evidence theory [23], [24]. In the D-S evidence theory, a proposition is represented by a frame of discernment Θ, which is a finite set of mutually exclusive elements. Each subset of the frame corresponds to an event that may be the solution to the proposition. Basic belief assignment function M is a mapping from each subset to [0, 1], and $M(A)$ is the belief degree to which event A is the solution. For function M, there is:

$$\begin{cases} M(\emptyset) = 0 \\ \sum_{A\in\Theta} M(A) = 1 \end{cases} \tag{17}$$

In BRB, the referential set of consequent attribute can be regarded as a special frame of discernment Θ' where $\forall A_i, A_j \in \{A | A \in 2^{\Theta'}, M(A) > 0\}$, and if $A_i \neq A_j$, there must be $A_i \cap A_j = \emptyset$.

For a certain frame of discernment, several sources of information may provide different value of function M, e.g., $M_1 \neq M_2$. M_1 and M_2 are called different evidences. In BRB, the rules can be regarded as the evidences, and the consequent belief distribution can be regarded as the value of function M. And the aggregation of rules is based on the D-S evidence combination approach as follow:

$$M(A_i \cap A_j) = \begin{cases} \dfrac{\sum M_1(A_i) * M_2(A_j)}{1-K} & A_i \cap A_j \neq \emptyset \\ 0 & A_i \cap A_j = \emptyset \end{cases} \tag{18}$$

where

$$K = \sum M_1(A_i) * M_2(A_j), \quad A_i \cap A_j = \emptyset \tag{19}$$

The RIMER recursive algorithm [25] can be regarded as aggregating rules one by one using (18), and this recursive algorithm has been equivalently transformed into the analytical algorithm as (14) in [21].

The RIMER methodology has something in common with the weighted averages methodology. As the weighted averages methodology is quite well-known, its introduction is omitted.

Assume two belief distributions are: $D_1 = (0.7, 0.1, 0.2)$ and $D_2 = (0.2, 0.5, 0.3)$. The aggregation result of D_1 and D_2 will be $D_{er} = (0.462, 0.289, 0.249)$ using the RIMER methodology and $D_{wa} = (0.45, 0.3, 0.25)$ using the weighted averages methodology. Moreover, the aggregation result of 100 D_1 and 100 D_2 using the RIMER methodology will be $D_{er} = (0.473, 0.291, 0.236)$. These examples indicate that the effect of the RIMER methodology is similar with the one of the weighted averages methodology. However, the former has a bias towards some of the belief degrees. Such a bias only depends on the aggregated belief distributions. Besides, the aggregation result of one D_1 and 100 D_2 using the RIMER methodology will be $D_{er} = (0.187, 0.527, 0.286)$, which indicates that the result of RIMER methodology will also be changed by the quantity difference of rules. Fig. 1 shows the changing of the aggregation results when the number of D_2 is increasing.

The increased number of rules is the equivalent of the increased weight of a rule in the weighted averages methodology, i.e., one rule whose weights are 100 is the equivalent

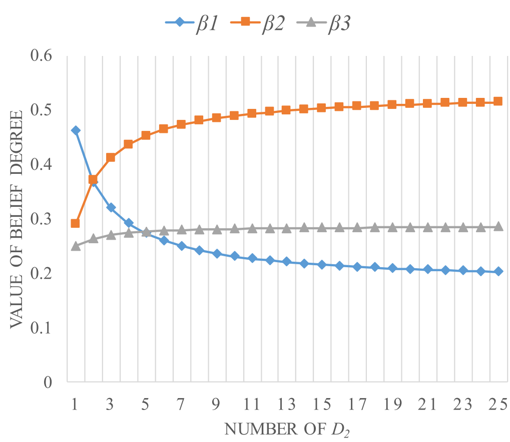


FIGURE 1. Changing of the aggregation result when the number of D_2 is increasing.

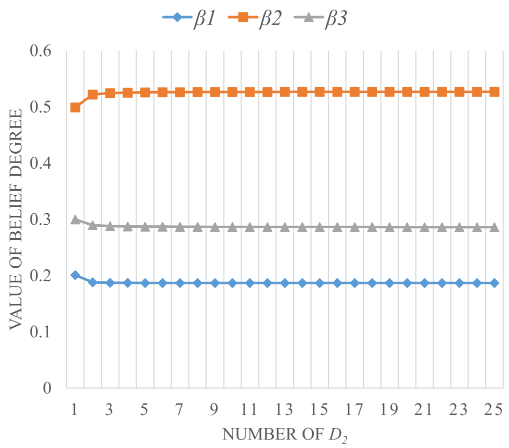


FIGURE 2. Aggregation result of one D_1 whose weight is 1 and n D_2 whose weights are $100/n$.

of 100 rules whose weights are 1. This property also applies to the RIMER methodology. Fig. 2 shows the aggregation results of one D_1 whose weight is 1 and n D_2 whose weights are $100/n$, and it can be noticed that those aggregation results merely have a little difference.

B. ISSUE OF EBRB SYSTEM WITH AN IMBALANCED RULE BASE

Since the data-driven method transforms each training sample into an extended belief rule, the number of generated rules belong to each class will equal to the number of training samples belong to each class. A rule base generated from an imbalanced dataset is an imbalanced rule base.

Take the classification dataset Thyroid [12] as an example. Thyroid is a series of datasets about thyroid disease research provided by the Garvan Institute in Sydney, Australia. Thyroid has 10 sub-datasets. This section selects one of its classification datasets. The quantity of samples in this dataset is 215, and each sample has five numerical antecedent attributes and a consequent attribute. The three classes in this dataset are Normal, Hyper, and Hypo, and the number of samples belong to each class are 150, 35 and 30, respectively. Thyroid is an imbalanced dataset, where Normal is the majority class and Hyper, Hypo are minority classes.

Uniformly and incrementally arrange 5 referential values for each antecedent attribute of the data, and then use stratified sampling to divide the dataset into 10 folds, where 9 folds are used for training and the other one fold are used for testing. For a testing sample of Hypo, the activation weights and consequent belief distribution of rules in this rule base are listed in Table 1, Table 2 and Table 3, respectively.

These rules have been sorted by their activation weights. Cause Thyroid is an imbalanced dataset, the training datasets obtained by stratified sampling are also imbalanced, and the rule base generated by the data-driven method will certainly be imbalanced. In such an imbalanced rule base, although the rules of Hypo exactly have higher activation weights, the quantity of them is much less than those of Normal.

TABLE 1. Rules of Normal in the imbalanced rule base.

rule No.	belief distribution	activation weight
1	(1.0, 0.0, 0.0)	0.0142
2	(1.0, 0.0, 0.0)	0.0135
3	(1.0, 0.0, 0.0)	0.0114
4	(1.0, 0.0, 0.0)	0.0105
5	(1.0, 0.0, 0.0)	0.0098
...	...	...
135	(1.0, 0.0, 0.0)	0.0005

TABLE 2. Rules of Hyper in the imbalanced rule base.

rule No.	belief distribution	activation weight
1	(0.0, 1.0, 0.0)	0.0060
2	(0.0, 1.0, 0.0)	0.0024
3	(0.0, 1.0, 0.0)	0.0022
4	(0.0, 1.0, 0.0)	0.0016
5	(0.0, 1.0, 0.0)	0.0016
...	...	...
31	(0.0, 1.0, 0.0)	0.0001

TABLE 3. Rules of Hypo in the imbalanced rule base.

rule No.	belief distribution	activation weight
1	(0.0, 0.0, 1.0)	0.0796
2	(0.0, 0.0, 1.0)	0.0682
3	(0.0, 0.0, 1.0)	0.0483
4	(0.0, 0.0, 1.0)	0.0403
5	(0.0, 0.0, 1.0)	0.0225
...	...	...
27	(0.0, 0.0, 1.0)	0.0006

For this sample, the sums of rule activation weights belong to the three classes are 0.570, 0.031 and 0.399, respectively, and the final belief distribution obtained is (0.590, 0.024, 0.386), which means the testing sample is misclassified into Normal. In the entire case study, the average accuracies of all the three classes are 100.00, 31.91 and 43.67, respectively. The imbalanced rule base brings high accuracies in majority classes but low accuracies in minority classes.

As discussed above, the RIMER methodology can be influenced by both quantity and weights of rules like the weighted averages methodology, so the reasoning result of an EBRB system will also be misled when the rule base is imbalanced.

C. ADJUSTING THE RULE BASE INTO BALANCED

Considering the rule base of a binary imbalanced classification problem, suppose the activation weight of the rule belonging to the majority class is $\omega_{majority}$ and the activation weight of the rule belonging to the minority class is

$\omega_{minority}$. Ideally, when $\omega_{majority} \geq \omega_{minority}$ there shall be $\sum \omega_{majority} \geq \sum \omega_{minority}$, and when $\omega_{majority} \leq \omega_{minority}$ there shall be $\sum \omega_{majority} \leq \sum \omega_{minority}$. But if the rule base is imbalanced, when $\omega_{majority} \leq \omega_{minority}$ there may still be $\sum \omega_{majority} \geq \sum \omega_{minority}$.

The condition $\omega_{majority} \leq \omega_{minority}$ is likely to occur when the testing sample belongs to the minority class. For easy understanding, simplify the RIMER methodology to the weighted averages methodology, and then the $\omega_{minority}$ is equal to $\beta_{minority}$ (because the values in consequent belief distribution are either 0 or 1 in classification problem) and vice versa, thus there may be $\sum \beta_{majority} \geq \sum \beta_{minority}$ which indicates a misclassification. In order to change the rule base to be balanced, a reasonable way is to simultaneously adjust the activation weights of all rules to make $\sum \omega_{majority} \leq \sum \omega_{minority}$ when the testing sample belongs to the minority class.

It can be noticed from (11) that there is $0 \leq \omega_k \leq 1$ for any rule. Therefore, for any two different rules with unequal activation weights ω_a, ω_b and a real number p, if $\omega_a > \omega_b$ and $p > 1$, there must be

$$\frac{\omega_a}{\omega_b} < \left(\frac{\omega_a}{\omega_b}\right)^p \tag{20}$$

Let $\hat{\omega}$ be the rule activation weight before normalization. To address the issue above, operation of the BA approach based on (20) can be represented as follow:

$$\begin{aligned} &\textit{adjust} && \bar{\omega}_k = \frac{\hat{\omega}_k^p}{\sum_{l=1}^{L} \hat{\omega}_l^p}, \quad k = 1, 2, \ldots, L \\ &\textit{to achieve} && \sum \bar{\omega}_{majority} \leq \sum \bar{\omega}_{minority} \end{aligned} \tag{21}$$

Real number p is called the balance parameter of the BA approach. Increasing the value of p can reduce the remainder of $\sum \bar{\omega}_{majority} - \sum \bar{\omega}_{minority}$. Fig. 3 shows the effect of p applied in the example of Section III-B. For that testing sample, set $p = 1.25$ is enough.

If the testing sample belongs to the majority class, the condition $\omega_{majority} \geq \omega_{minority}$ is likely to occur, and thus

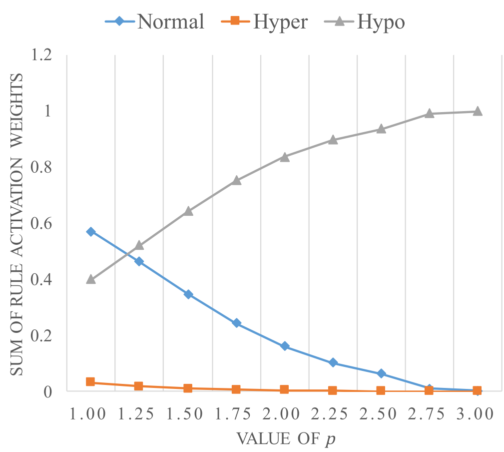


FIGURE 3. Sums of rule activation weights in the example with different values of p. The testing sample belongs to Hypo.

TABLE 4. Reasoning results of BA-EBRB with different value of p in Thyroid dataset.

p	Class 1 (150)	Class 2 (35)	Class 3 (30)	Ave. ACC
1.0 (Liu_EBRB)	100.00	31.91	43.67	81.47
2.0	100.00	62.17	70.00	89.46
3.0	100.00	77.08	80.00	93.39
4.0	99.73	84.83	82.00	94.61
5.0	99.40	94.17	83.00	95.90
6.0	99.40	96.25	87.67	97.07
7.0	99.40	95.83	90.33	97.35
8.0	99.40	98.00	93.33	97.91
9.0	99.53	96.83	91.33	97.77
10.0	99.40	96.83	92.67	97.90
11.0	99.33	96.50	92.00	97.95
12.0	99.33	96.92	91.67	98.00
13.0	99.27	96.58	91.67	97.59
14.0	99.33	97.08	93.33	97.63
15.0	99.20	97.25	93.33	97.59

$\sum \beta_{majority} \gg \sum \beta_{minority}$ is nearly for sure. Since the BA approach always increases the value of $\sum \omega_{majority}$, it will certainly bring a possibility of misclassification for majority classes. But, as demonstrated in the following section, such a possibility is quite little and acceptable.

IV. CASE STUDIES

The case studies in this section demonstrate the performance of EBRB system optimized by the BA approach (BA-EBRB) with 10-fold cross-validation. After that, an algorithm is proposed to set the value of the balance parameter p.

For each antecedent attribute in the following case studies, 5 referential values of antecedent attribute are arranged uniformly and incrementally. Number of samples, accuracy of each class and the average accuracy are shown in tables of reasoning result. Note that the average accuracy of the whole dataset does not equal to the weighted average of the accuracy of classes in 10-fold cross-validation because the distribution proportion of samples in different folds may not be precisely equal.

A. PERFORMANCE OF EBRB SYSTEM OPTIMIZED BY THE BA APPROACH

Take Thyroid dataset as the first case study. Table 4 lists the reasoning results of EBRB system when the value of p is gradually increased and Fig. 4 show the changing trend more clearly. For each class, its quantity is attached following the label. Note that $p = 1$ means the EBRB system is the conventional one, namely Liu-EBRB.

It can be concluded from the results that with the value of p increasing, the accuracies of the minority classes (Class 2 and

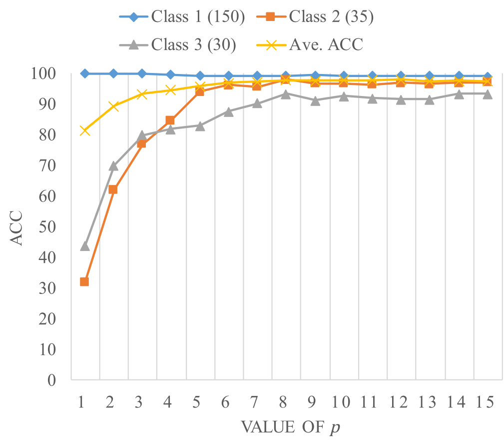


FIGURE 4. Reasoning results of BA-EBRB with different value of *p* in Thyroid dataset.

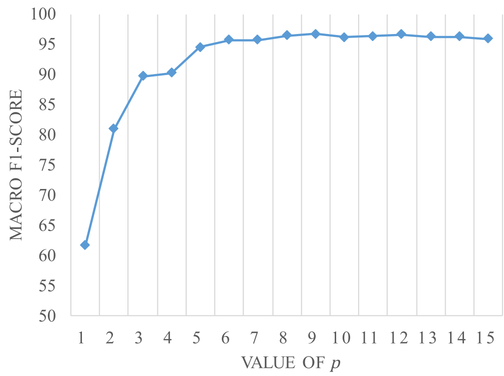


FIGURE 5. Macro F1-score of BA-EBRB with different value of *p* in Thyroid dataset.

Class 3) are significantly increasing while the accuracy of the majority class (Class 1) is slightly decreasing. Both of them finally tend to be stable.

The accuracy of majority class Normal seems to be excellent in Liu-EBRB but at a cost. Many samples are misclassified into the majority class, although it does not ruin the average accuracy because those minority classes are at a disadvantage in quantity. Is the result of Liu-EBRB acceptable? As the average accuracy is not the most reasonable evaluation metrics of imbalanced classification problems, this study uses Macro F1-score to evaluate the reasoning results of multi-classification. Fig. 5 shows the Macro F1-scores of this case study.

The Macro F1-score of Liu-EBRB is only 61.70, which is too low to be acceptable, but that of BA-EBRB is 96.76 at $p = 9$. The BA approach successfully increases the average accuracy by 20% and the Macro F1-score by 56% in Thyroid dataset. To further illustrate the effectiveness of the BA approach and seek the optimal value of p, the other two case studies are conducted in the Glass dataset and Bupa dataset. The detailed results are listed in Table 5 and Table 6, respectively. And the line arts are shown in Fig. 6 and Fig. 7, respectively. For each class, its quantity is attached following the label.

TABLE 5. Reasoning results of BA-EBRB with different value of *p* in the Glass dataset.

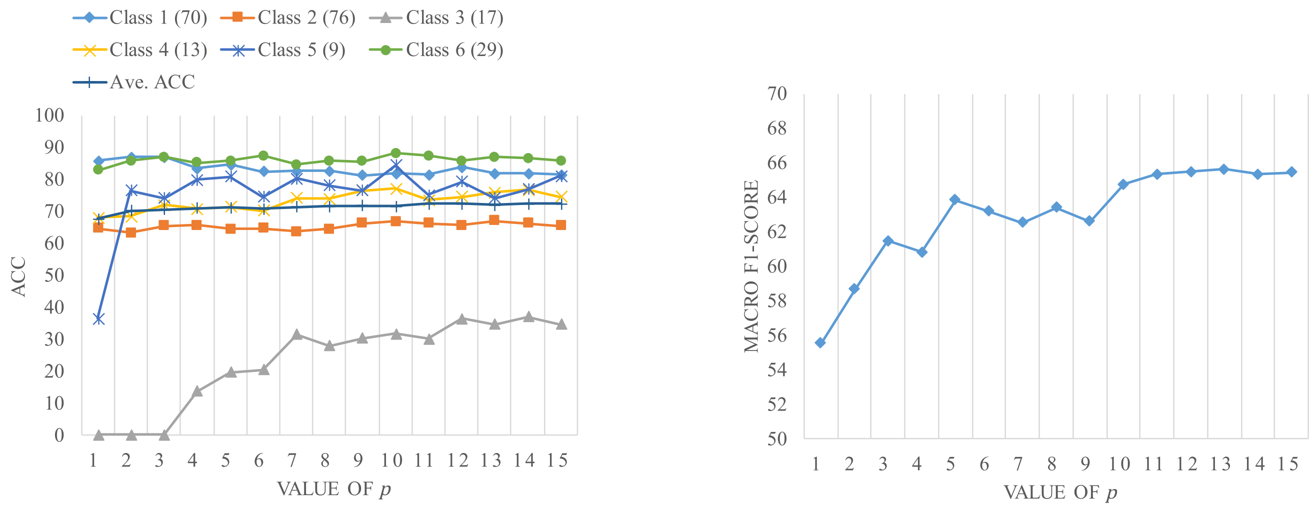
p	Class 1 (70)	Class 2 (76)	Class 3 (17)	Class 4 (13)	Class 5 (9)	Class 6 (29)	Ave. ACC	Macro F1-score
1.0	85.73	64.58	0.00	67.86	36.29	82.73	67.48	55.50
2.0	87.03	63.13	0.00	68.43	76.46	85.82	70.22	58.66
3.0	86.84	65.40	0.00	72.01	74.01	86.98	70.50	61.47
4.0	83.46	65.63	13.67	70.78	79.78	85.10	70.85	60.81
5.0	84.56	64.37	19.62	71.31	80.73	85.69	71.20	63.81
6.0	82.44	64.53	20.29	70.15	74.49	87.25	70.82	63.16
7.0	82.67	63.56	31.31	74.05	80.17	84.56	71.25	62.50
8.0	82.59	64.37	27.86	73.92	78.12	85.79	71.43	63.41
9.0	81.15	66.21	30.16	76.44	76.46	85.62	71.72	62.59
10.0	81.80	66.69	31.51	77.00	84.27	88.13	71.53	64.74
11.0	81.48	66.23	30.06	73.46	75.14	87.34	72.32	65.32
12.0	83.69	65.53	36.36	74.46	79.17	85.71	72.21	65.48
13.0	81.79	67.03	34.58	75.79	74.17	86.92	71.96	65.62
14.0	81.84	66.13	36.87	76.75	76.88	86.54	72.30	65.34
15.0	81.39	65.38	34.62	74.44	81.01	85.80	72.29	65.44

TABLE 6. Reasoning results of BA-EBRB with different value of *p* in the Bupa dataset.

p	Class 1 (145)	Class 2 (200)	Ave. ACC	Macro F1-score
1.0	32.84	89.72	64.95	59.10
2.0	51.30	81.63	69.20	65.97
3.0	57.96	77.05	68.59	66.69
4.0	57.98	75.32	67.23	65.93
5.0	59.20	74.12	67.70	65.59
6.0	61.70	73.49	68.29	66.45
7.0	60.85	73.30	67.92	67.21
8.0	61.36	72.64	67.97	66.14
9.0	60.07	72.92	67.71	66.52
10.0	60.75	71.65	67.21	65.35
11.0	58.94	71.99	66.78	65.29
12.0	59.94	71.77	66.81	64.80
13.0	59.39	71.26	66.28	64.57
14.0	58.79	70.88	66.34	64.82
15.0	59.10	71.03	66.53	64.04

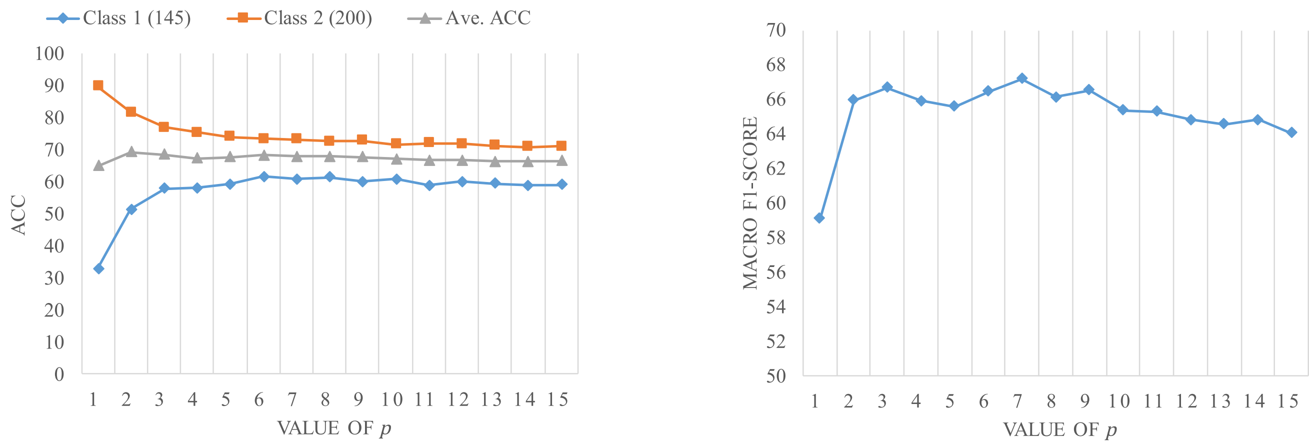
In both case studies, the average accuracies are increased by about 7% while the Macro F1-scores are increased by more than 14%. These results prove that the BA approach is an effective and reliable tool to improve the performance of EBRB in imbalanced datasets.

The relationship between the value of p and either average accuracy or Macro F1-score is hard to be concluded. It may seems to be a convergence function in Fig. 6 but a unimodal function in Fig. 7, so neither a ternary search algorithm nor set a positive infinity value for p can fit the problem.



(a) Reasoning results (b) Macro F1-score

FIGURE 6. **Reasoning result and Macro F1-score of BA-EBRB with different value of p in the Glass dataset.**



(a) Reasoning results (b) Macro F1-score

FIGURE 7. **Reasoning result and Macro F1-score of BA-EBRB with different value of p in the Bupa dataset.**

B. HOW TO DETERMINE THE VALUE OF P

In fact, the only confirmed conclusion is that increasing the value of p will lead a decrease in the accuracies of majority classes but an increase in the accuracies of minority classes. As a result, the relationship will certainly become a unimodal function in binary classification problems but may become any kind of function in multi-classification problems. Besides, the sum of rule activation weight does not only depend on the number of rules but also each single rule activation weight. Since the latter depends on testing samples and generated rules (i.e., training samples), it is impossible to calculate the value of p based on the class distribution of a dataset.

As the optimal value of p depends on the information of whole dataset, it can hardly be calculated by a mere mathematical formula. This section proposes an approximate iterative algorithm to solve the problem.

Define $E(p)$ as the evaluation function of p. $E(p)$ can be measured by average accuracy, Macro F1-score, the accuracy of a minority class or other evaluation metrics of a model. And define $[1, s]$ as the value range of p, p_{opt} as the optimal value of p in $[1, s]$, and ε as the threshold of step length. Then the objective function of the algorithm is represented as:

$$\begin{aligned} \text{maximize} \quad & E(p) \\ s.t. \quad & p \in [1, s] \\ & |p - p_{opt}| \le \varepsilon \end{aligned} \tag{22}$$

$E(p)$ in (22) also depends on the generated EBRB and testing samples, i.e., information of a whole dataset, so it has to be calculated by the 10-fold cross-validation but not a mathematical formula. The detailed algorithm is introduced as follows:

Once the value of p is determined, it is unnecessary to be changed during the reasoning process. Before applying the algorithm there are some problems worth considering:

1) What is the time complexity of this algorithm?
Let n represent the number of antecedent attributes in S and $T(nS)$ represent the time complexity of an iteration process, then the complete time complexity of

Algorithm 1 An Approximate Iterative Algorithm to Seek the Optimal Value for Balance Parameter p

Input: The complete or stratified sample dataset (S), the upperbound of value range s and the threshold of step length ε.
Output: The approximate optimal value of p for this dataset (p_{opt}).
Begin
Generate 10 EBRBs (Φ) for 10-fold cross-validation from dataset S
$inf = 1$;
$sup = s$;
$\Delta p = 1$;
while $sup - inf > \varepsilon$ **do**
  $p_{opt} = inf$;
  for $p = inf$; $p \leq sup$; $p = p + \Delta p$ **do**
    for each $\hat{\omega}$ in Φ **do**
      Adjust $\hat{\omega}$ to $\bar{\omega}$ using (21)
    end for
    Calcuate $E(p)$ using $\bar{\omega}$
    if $E(p) \geq E(p_{opt})$ **then**
      $p_{opt} = p$;
    end if
  end for
  if $p_{opt} - \Delta p \geq inf$ **then**
    $inf = p_{opt} - \Delta p$;
  end if
  if $p_{opt} + \Delta p \leq sup$ **then**
    $sup = p_{opt} + \Delta p$;
  end if
  $\Delta p = \Delta p/10$;
end while
End

Algorithm 1 is calculated as:

$$T(s + 20 * lg(\varepsilon)) * T(nS) = O(lg(\varepsilon) * T(nS)) \quad (23)$$

Eq. (23) shows that the time complexity of Algorithm 1 mainly depends on $T(nS)$, i.e., the size of S ($|S|$).

2) How to apply this algorithm to large scale datasets?
As discussed in 1), the time complexity of Algorithm 1 will increase along with $T(nS)$ and be unacceptable when the latter becomes very large. Nevertheless, the information of a dataset can be derived also from its stratified samples. $|S|$ is not recommended to be too large. Before applying this algorithm to a large scale dataset, a stratified sampling for the complete dataset will help.

3) How to determine the value of s?
Suppose a worst-case: $|S|$ is n, and there are $n-1$ rules belonging to the majority class whose weights are all equal to $\omega_{majority}$ and one rule belonging to the minority whose weight is $t * \omega_{majority}$. To adjust the rule base to be balanced, the value of s should satisfy:

$$\begin{aligned} \left(\frac{t}{n}\right)^s &\geq n - 1 * \left(\frac{1}{n}\right)^s \\ \Rightarrow \quad s &\geq log_t(n-1) \end{aligned} \quad (24)$$

As discussed in 2), $|S|$ is not recommended to be too large. Assuming $|S|$ is 2000 and $t = 1.5$, then just set $s = 20$ is enough.

V. EFFECTIVENESS VALIDATION OF THE PROPOSED APPROACH

To further validate the effectiveness of the proposed BA approach, a series of benchmark classification datasets from UCI are used in this section to test the performance of BA-EBRB, and the derived results are compared with both conventional classification approaches and recent works of EBRB system.

A. COMPARE WITH CONVENTIONAL CLASSIFICATION APPROACHES

Several public classification datasets from UCI, including balanced and imbalanced, are used in this section. The class distribution of datasets are shown in Fig. 8.

The reasoning results of BA-EBRB in this comparison are all derived from 10 independent runs with 10-fold cross-validation. Table 7 lists the comparison results. Those approaches for comparison are Decision Tree [27], Naive Bayes [28], Fuzzy Set [29], KNN [23], SVM and LDA [31], conventional Liu-EBRB, DRA-EBRB [32], SRA-EBRB [19] and VP-EBRB [33]. Since F1-score was not used as metrics in those papers, the results for comparison have to be measured by average accuracy. The comparison rank of approach is attached following each reasoning result. BA-EBRB is not always the best for all datasets, but its ranks never fell out of the top 4, and its average rank is highest. The comparison

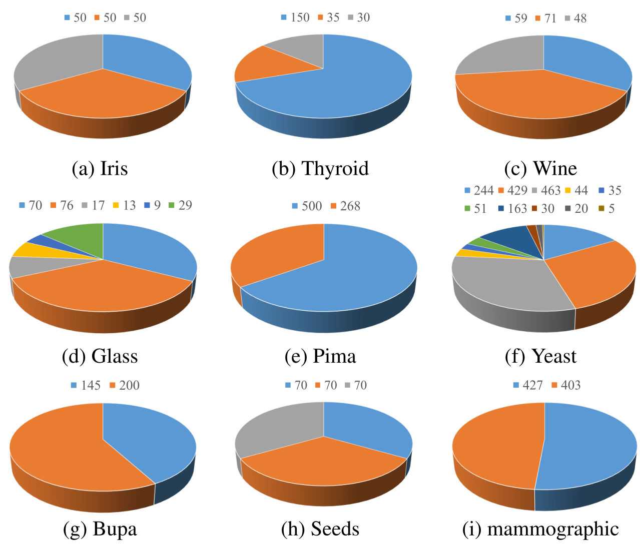


FIGURE 8. Class distribution of classification datasets.

TABLE 7. Performance of BA-EBRB compare with conventional classification approaches. (The best results are highlighted in boldface).

	Decision Tree	Naive Bayes	Fuzzy Set	KNN	SVM	LDA	Liu-EBRB	DRA-EBRB	SRA-EBRB	VP-EBRB	BA-EBRB
Iris	-	94.33 (8)	94.67 (7)	85.18 (10)	96.00 (2)	89.87 (9)	95.26 (4)	95.50 (3)	94.80 (6)	**96.27 (1)**	95.26 (4)
Thyroid	-	92.79 (5)	-	93.99 (4)	96.19 (3)	88.33 (6)	81.47 (7)	97.19 (2)	-	-	**98.00 (1)**
Wine	91.40 (8)	-	-	96.05 (6)	96.40 (4)	71.24 (9)	96.32 (5)	96.46 (3)	96.85 (2)	93.01 (7)	**97.02 (1)**
Glass	61.90 (8)	57.69 (10)	68.57 (6)	70.11 (4)	61.21 (9)	52.90 (11)	67.85 (7)	69.65 (5)	**73.08 (1)**	71.03 (3)	72.32 (2)
Pima	-	-	72.16 (5)	73.06 (3)	**77.76 (1)**	68.26 (9)	70.87 (8)	71.44 (7)	71.71 (6)	74.61 (2)	72.80 (4)
Yeast	-	-	57.77 (3)	53.17 (6)	-	-	45.61 (7)	54.13 (5)	56.85 (4)	**58.83 (1)**	58.63 (2)
Bupa	64.90 (6)	-	-	60.66 (8)	67.51 (3)	54.97 (9)	65.02 (5)	64.90 (6)	**70.46 (1)**	66.67 (4)	69.20 (2)
Seeds	-	-	90.00 (6)	-	-	-	91.33 (4)	92.02 (3)	91.24 (5)	92.48 (2)	**93.95 (1)**
Mammographic	-	-	-	80.12 (3)	-	-	79.67 (5)	78.39 (6)	**82.53 (1)**	80.95 (2)	79.81 (4)
Average Rank	7.33 (9)	7.67 (10)	5.4 (6)	5.5 (7)	3.66 (4)	8.83 (11)	5.78 (8)	4.44 (5)	3.25 (3)	2.75 (2)	**2.33 (1)**

illustrates that the BA approach is an effective and robust tool appling to various datasets.

It can be noticed that BA-EBRB not only performs well in imbalanced datasets like Thyroid and Wine, but also gets the highest rank in the balanced dataset Seeds. Such a result is reasonable. To a certain extent, the BA approach may also resolve data inconsistency like the DRA approach because of their similarity in mathematical tricks when using the conventional method to calculate rule activation weights. But the former focuses on the issue of the imbalanced rule base and will not change the visitation rule rate (VRR, usually as an evaluation metrics of efficiency) of EBRB system. As a result, it may be sometimes better than the conventional DRA approach. Furthermore, different from the DRA approach, the BA approach is also able to combine with some novel rule activation weight calculation methods which do not base on the multiplication of individual matching degrees (See [34]).

B. COMPARE WITH RECENT WORKS OF EBRB SYSTEM

To demonstrate the advancement of the BA approach, this study compares it with CABRA-EBRB [35] and NP-EBRB [36], both of them are novel and significant works proposed in the recent two years. The statistics on the extra datasets used in this section is summarized in Table 8.

Table 9 lists their comparison results in several small scale classification datasets. The reasoning results of BA-EBRB are derived from 10 independent runs with 10-fold cross-validation and still measured by average accuracy because

TABLE 8. Statistics on large scale classification datasets used for comparison.

	No. of samples	No. of attributes	No. of classes
Ecoli	336	7	8
Banana	5300	2	2
Phoneme	5404	5	2
Pageblocks	5472	10	5

TABLE 9. Comparison results in small scale classification datasets. (The best results are highlighted in boldface).

	CABRA-EBRB	NP-EBRB	BA-EBRB
Iris	**96.00**	95.73	95.20
Wine	96.63	**97.87**	97.02
Glass	72.90	**75.51**	72.32
Yeast	53.91	54.51	**58.63**
Ecoli	85.42	**87.14**	85.78
Seeds	92.38	93.24	**93.95**
Mammographic	79.52	79.57	**79.81**

TABLE 10. Comparison results in large scale classification datasets. (The best results are highlighted in boldface).

	Liu-EBRB	DRA-EBRB	NP-EBRB	BA-EBRB
Banana	63.43	87.06	87.28	**90.06**
Phoneme	71.89	76.02	85.23	**88.61**
Pageblocks	89.77	94.19	94.30	**96.01**

F1-score was not mentioned in those papers. BA-EBRB gets the highest rank in three datasets as well as NP-EBRB, and the latter performs better in Glass while the former performs better in a larger imbalanced dataset, Yeast. For other datasets, there is no much difference.

Additionally, the performance of the BA approach in large scale datasets whose sizes are greater than 5000 is compared with those listed in [36], where CABRA-EBRB did not involve in that comparison because it is too time-consuming. Their comparison results are listed in Table 10. These reasoning results are all obtained from 2-fold cross-validation.

The value of balance parameter p for these datasets are determined using Algorithm 1 with $|S| = 1000$. It can be seen from Table 10 that BA-EBRB has the best reasoning results in all the 3 datasets. Compared with NP-EBRB, the misclassification rates of BA-EBRB in these datasets are decreased by 21.9%, 22.9% and 30.0%, respectively. This result illustrates the BA approach is more effective and potential in large scale datasets. It also demonstrates that seeking the approximate optimal value of p from the stratified samples is an effectual and reliable method to apply Algorithm 1 to large scale datasets.

VI. CONCLUSION

In this study, the BA approach is proposed to improve the performance of EBRB systems applied in imbalanced classification datasets. The presented analysis and case studies show why the reasoning result of the conventional RIMER methodology can be affected by an imbalanced rule base and how the BA approach adjusts the rule activation weights to balance a rule base. A series of benchmark classification datasets validate the effectiveness of BA-EBRB compared with several conventional classification methods and novel studies about EBRB system. The further conclusions of this study are summarized as follow:

1) The rule aggregation process of EBRB system using the conventional RIMER methodology attaches more importance to the weighted combination of activated rules but neglects the importance of those rules with a higher activation weight. This issue will be magnified when the data-driven method generates an imbalanced rule base for EBRB system applied in imbalanced classification datasets. Since the number of rules belongs to minority classes will be always at a disadvantage, the reasoning ability of EBRB system, especially the accuracies of minority classes will be heavily decreased in an imbalanced rule base.
2) To illustrate the issue above, this study provides an example of the counterintuitive reasoning process of EBRB system applied in an imbalanced classification dataset about thyroid disease research, and then proposes a BA approach. The BA approach addresses the issue by simultaneously adjusting the activation weights of all rules in the rule base using a balance parameter. To determine the optimal value of the balance parameter, an approximate iterative algorithm is proposed. The BA approach can combine with various rule activation weight calculation methods and might worth to be applied as a generic process of EBRB system.
3) The case studies in imbalanced datasets demonstrate the BA approach can effectively increase the accuracies of the minority classes, and thus leads an increase in the average accuracy and the Macro F1-score over the whole dataset. And this approach never compromises the effectiveness of the system. The effectiveness and robust of BA-EBRB are further demonstrated by the comparisons with several conventional classification methods and recent works about EBRB system. Moreover, the comparison result illustrates the BA approach is more effective and potential in large scale datasets, and also proves using the proposed approximate algorithm to seek the optimal value for the balance parameter is effectual.

Future research will concentrate on a more refined balance adjusting method, e.g., adjust the rule activation weights of each class respectively by different balance parameters. An exact algorithm to seek the optimal value of these balance parameters will also be considered.

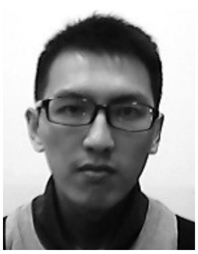
ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their helpful comments, which have been a great support to improve the quality of this paper.

REFERENCES

[1] Z. Zhou, J. Yang, and C. Hu, *Belief Rule Base Expert System and Complex System Modeling*. Beijing, China: Science Press, 2011.
[2] Z. Zhou, Z. Feng, C. Hu, X. Han, Z. Zhou, and G. Li, "A hidden fault prediction model based on the belief rule base with power set and considering attribute reliability," *Sci. China Inf. Sci.*, vol. 62, no. 10, Aug. 2019, Art. no. 202202.
[3] Y. Yang, J. Wang, G. Wang, and Y.-W. Chen, "Research and development project risk assessment using a belief rule-based system with random subspaces," *Knowl.-Based Syst.*, vol. 178, pp. 51–60, Aug. 2019.
[4] H. Wei, G.-Y. Hu, Z.-J. Zhou, P.-L. Qiao, Z.-G. Zhou, and Y.-M. Zhang, "A new BRB model for security-state assessment of cloud computing based on the impact of external and internal environments," *Comput. Secur.*, vol. 73, pp. 207–218, Mar. 2018.
[5] 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.
[6] 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.
[7] L.-H. Yang, Y.-M. Wang, Y.-X. Lan, L. Chen, and Y.-G. Fu, "A data envelopment analysis (DEA)-based method for rule reduction in extended belief-rule-based systems," *Knowl.-Based Syst.*, vol. 123, pp. 174–187, May 2017.
[8] K. AbuDahab, D.-L. Xu, and Y.-W. Chen, "A new belief rule base knowledge representation scheme and inference methodology using the evidential reasoning rule for evidence combination," *Expert Syst. Appl.*, vol. 51, pp. 218–230, Jun. 2016.
[9] R. Sun, "Robust reasoning: Integrating rule-based and similarity-based reasoning," *Artif. Intell.*, vol. 75, no. 2, pp. 241–295, Jun. 1995.
[10] C.-L. Hwang and K. Yoon, *Methods for Multiple Attribute Decision Making*. Berlin, Germany: Springer, 1981, pp. 58–191. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-642-48318-9_3
[11] L. A. Zadeh, "Fuzzy sets," *Inf. Control*, vol. 8, no. 3, pp. 338–353, 1965.
[12] D. Dua and C. Graff. (2019). *UCI Machine Learning Repository*. [Online]. Available: http://archive.ics.uci.edu/ml
[13] H. Wang, L. Yang, Y. Fu, Y. Wu, and X. Gong, "Differential evolutionary algorithm for parameter training of belief rule base under expert intervention," *Comput. Sci.*, vol. 42, no. 5, pp. 88–93, 2015.
[14] M. Li, Y. Fu, W. Liu, and Y. Wu, "Parameter training approach for belief rule base based on the cuckoo search algorithm," *J. Chin. Comput. Syst.*, vol. 39, no. 6, pp. 31–37, 2018.
[15] J.-B. Sun, J. X. Huang, L.-L. Chang, J. Jiang, and Y.-J. Tan, "BRBcast: A new approach to belief rule-based system parameter learning via extended causal strength logic," *Inf. Sci.*, vol. 444, pp. 51–71, May 2018.
[16] Q. Su, L. Yang, Y. Fu, and R. Yu, "Structure optimization framework of extended belief rule based on bk-tree," *J. Frontiers Comput. Sci. Technol.*, vol. 10, no. 2, pp. 257–267, 2016.
[17] W. Liu, C. Xiao, and Y. Fu, "Extended belief rule base inference method based on the hash index," *J. Xidian Univ.*, vol. 46, no. 2, pp. 145–151, 2019.
[18] Y. Lin and Y. Fu, "NSGA-II-based EBRB rules activation multi-objective optimization," *CAAI Trans. Intell. Syst.*, vol. 13, no. 3, pp. 422–430, 2018.
[19] 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.
[20] J.-B. Yang, "Rule and utility based evidential reasoning approach for multiattribute decision analysis under uncertainties," *Eur. J. Oper. Res.*, vol. 131, no. 1, pp. 31–61, May 2001.

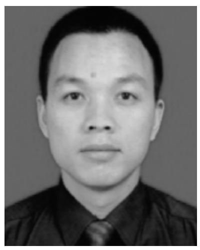
[21] J.-B. Yang, J. Liu, D.-L. Xu, J. Wang, and H. Wang, "Optimization models for training belief-rule-based systems," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 37, no. 4, pp. 569–585, Jul. 2007.
[22] J.-B. Yang and M. G. Singh, "An evidential reasoning approach for multiple-attribute decision making with uncertainty," *IEEE Trans. Syst., Man, Cybern.*, vol. 24, no. 1, pp. 1–18, Jan. 1994.
[23] A. P. Dempster, "A generalization of Bayesian inference," *J. Roy. Stat. Soc. B, Methodol.*, vol. 30, no. 2, pp. 205–232, 1968.
[24] G. Shafer, *A Mathematical Theory of Evidence*. Princeton, NJ, USA: Princeton Univ. press, 1976.
[25] J.-B. Yang and D.-L. Xu, "On the evidential reasoning algorithm for multiple attribute decision analysis under uncertainty," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 32, no. 3, pp. 289–304, May 2002.
[26] V.-N. Hu, Y. Nakamori, T.-B. Ho, and T. Murai, "Multiple-attribute decision making under uncertainty: The evidential reasoning approach revisited," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 36, no. 4, pp. 804–822, Jul. 2006.
[27] D. C. Wickramarachchi, B. L. Robertson, M. Reale, C. J. Price, and J. Brown, "HHCART: An oblique decision tree," *Comput. Statist. Data Anal.*, vol. 96, pp. 12–23, Apr. 2016.
[28] J. Wu, S. Pan, X. Zhu, Z. Cai, P. Zhang, and C. Zhang, "Self-adaptive attribute weighting for naive bayes classification," *Expert Syst. Appl.*, vol. 42, no. 3, pp. 1487–1502, Feb. 2015.
[29] 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.
[30] J. Derrac, F. Chiclana, S. García, and F. Herrera, "Evolutionary fuzzy k-nearest neighbors algorithm using interval-valued fuzzy sets," *Inf. Sci.*, vol. 329, pp. 144–163, Feb. 2016.
[31] 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.
[32] 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.
[33] 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.
[34] 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, May 2018.
[35] L.-H. Yang, Y.-M. Wang, and Y.-G. Fu, "A consistency analysis-based rule activation method for extended belief-rule-based systems," *Inf. Sci.*, vols. 445–446, pp. 50–65, Jun. 2018.
[36] L.-H. Yang, J. Liu, Y.-M. Wang, and L. Martínez, "Extended belief-rule-based system with new activation rule determination and weight calculation for classification problems," *Appl. Soft Comput.*, vol. 72, pp. 261–272, Nov. 2018.



WEIJIE FANG received the B.S. degree from Fuzhou University, in 2017, where he is currently pursuing the master's degree. His research interests include multiobjective optimization, intelligent decision technology, rule-based inference, big data analysis, and data mining.

XIAOTING GONG received the M.S. degree from Fuzhou University, in 2006. She is currently a Lecturer with the College of Economics and Management, Fuzhou University. Her research interests include multicriteria decision making under uncertainty and information hiding technology.

GENGGENG LIU 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.



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, and privacy protection.

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 decision theory and methods, data mining and machine learning, and intelligent systems.