Information Sciences 336 (2016) 75–91

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Belief rule based expert system for classification problems with new rule activation and weight calculation procedures

Leilei Chang [a], ZhiJie Zhou [a,*], Yuan You [a], Longhao Yang [b], Zhiguo Zhou [c]

[a] High-Tech Institute of Xi'an, Xi'an, Shaanxi 710025, PR China
[b] Decision Sciences Institute, Fuzhou University, Fuzhou University, Fuzhou 350116, PR China
[c] Department of Radiation Oncology, The University of Texas Southwestern Medical Center, Dallas, TX 75235, USA

ARTICLE INFO

Article history:
Received 20 June 2015
Revised 11 November 2015
Accepted 4 December 2015
Available online 21 December 2015

Keywords:
Classification problems
Belief rule base
Optimization algorithm

ABSTRACT

Classification problems are significant because they constitute a meta-model for multiple theoretical and practical applications from a wide range of fields. The belief rule based (BRB) expert system has shown potentials in dealing with both quantitative and qualitative information under uncertainty. In this study, a BRB classifier is proposed to solve the classification problem. However, two challenges must be addressed. First, the size of the BRB classifier must be controlled within a feasible range for better expert involvement. Second, the initial parameters of the BRB classifier must be optimized by learning from the experts' knowledge and/or historic data. Therefore, new rule activation and weight calculation procedures are proposed to downsize the BRB classifier while maintaining the matching degree calculation procedure. Moreover, the optimal algorithm using the evidential reasoning (ER) algorithm as the inference engine and the differential evolution (DE) algorithm as the optimization engine is proposed to identify the fittest parameters, including the referenced values of the antecedent attributes, the weights of the rules and the beliefs of the degrees in the conclusion. Five benchmarks, namely, iris, wine, glass, cancer and pima, are studied to validate the efficiency of the proposed BRB classifier. The result shows that all five benchmarks could be precisely modeled with a limited number of rules. The proposed BRB classifier has also shown superior performance in comparing it with the results in the literature.

1. Introduction

Classification problems are fundamental to solving many theoretical and practical applications, including pattern recognition [20], medical diagnosis [15], and image processing [19]. However, classification problems are complex due to the interconnected sophistications of the correlations among the high dimensions [28]. Therefore, solving classification problems is one of the most pressing challenges in these fields.

Many attempts have been made to solve the classification problem. The prevailing approaches to solving this problem include k nearest neighbors (kNN) [2], support vector machine (SVM) [22], evolutionary algorithm-based approaches [23], and the decision tree [9]. Moreover, some more interesting and non-conventional techniques were introduced, including the

* Corresponding author. Tel.: +86 29 84744954.
E-mail address: zhouzj04@mails.tsinghua.edu.cn (Z. Zhou).

http://dx.doi.org/10.1016/j.ins.2015.12.009
0020-0255/

Learning Automata (LA)-based classifier [3], gravitational inspired classifier [27], feature vector graph-based classifier [44], and the blood pressure regularization-based classifier [25].

However, three challenges must be further addressed to better understand and solve the problem. First, human knowledge must be included in the classifiers. Human knowledge plays a key role in understanding the complex correlations among high dimensions residing in classification problems. However, its vagueness and uncertainty renders human knowledge difficult to fit into any conventional classifier [13]. Second, noise data would interfere and even diminish the applicability of the classifiers. The data are the representation of the classification problems, and it always involves certain noise in either benchmarks or practical cases. Even a small fraction of noise data could cause great interference and degrade the efficiency of the classifiers [14]. The impact of noise must be reduced. Finally, the Bayesian inference approach could not be directly applied because a good knowledge of the actual probability distribution for the characteristic variables is lacking [34].

Some researchers have used expert systems to model the human inference process. Thus far, several expert-systems-based techniques have been introduced to identify the hidden structural information and interconnected correlations of the classification problem. Of these, the artificial neural network (ANN) and the fuzzy set, as well as their combinations with other techniques, have been applied in many attempts and produced promising results [13].

ANN is comprised of multiple neural units and can precisely record a system's behavior while requiring no extra information other than the initial data [24]. Moreover, multiple ANNs could be layered to integrate and reinforce the prediction results of a single ANN when the classification problems are of over-numbered attributes [1]. However, ANN is essentially a black box [14], which makes the training and learning process unreachable and further denies access to experts' knowledge and experience.

The fuzzy set was first proposed by Zadeh in 1965 [40] and uses the "IF-THEN" rule as a semantic means to represent ambiguous human knowledge [18] in various forms, including numbers, intervals, fuzzy numbers and random numbers [43]. Fuzzy set-related and -based approaches have been applied to solving the classification problem and obtained relatively satisfactory results [14,15,18,20]. However, it has been argued that fuzzy set was heavily dependent on the choice of parameters, particularly the number of fuzzy partitions and the number of rules [13].

With this, it is natural for some scholars to combine ANN with the fuzzy set and apply it to the classification problem [21,49]. Moreover, other techniques have been combined with ANN and the fuzzy set to generate new classifiers with new features [23,27].

The Belief Rule Based (BRB) expert system extends "IF-THEN" rules (which are also used in fuzzy set) from mere linguistic terms to include numerical inputs [39]. Therefore, BRB can absorb both qualitative/linguistic and quantitative/numerical information under uncertainty as well as incompleteness [38]. This extension makes BRB versatile to more applicable fields. The result of BRB is also in the same belief structure as the input, which preserves the consistency of the deduction process. Moreover, BRB is a "white box" that makes the training and learning process available for experts' involvement. Thus far, BRB has been successfully applied in various fields, including system behavior prediction [45], system readiness assessment [4] and military capabilities assessment [16].

The information handled by BRB is also of ambiguous and uncertain characteristics and shares certain similarities with "fuzziness". However, these two approaches differ from each other. The fuzzy set uses the membership function to represent the degree of an object belonging to a certain set and therefore the information could be handled in a more comprehensive and reasonable fashion [40,43]. However, BRB uses the belief function, which contains the designated scales as well as the corresponding beliefs for integrating both subjective and objective information under uncertainty [38]. In fact, Zadeh, proposed the fuzzy set theory and made the first attempt to extend the belief rules under the fuzzy environment based on his work on the concept of information granularity and the theory of possibility [41,42]. Additional studies were later conducted to further explore both theoretical and practical aspects. Readers can refer to the literature [35,36] and references therein for more information.

To utilize the BRB system in classification problems more efficiently, two challenges must be addressed. First, the BRB systems must be downsized because a BRB classifier with too many rules is impossible to construct either by experts or by using historic data [6], which is also faced by the fuzzy set [13]. Second, initial parameters of the BRB classifier must be optimized to increase their precision in solving classification problems [45,47].

For the first challenge, new rule activation and weight calculation procedures are proposed here for the first time by assuming the attributes that are "disjunctive" (while the conventional matching degree calculation procedure remains).

For the second challenge, certain BRB parameters must be trained and optimized. There have been multiple studies regarding this topic [8,45,47]. Three dilemmas must be addressed [5]: First, numerical referenced values of the antecedent attributes must be transformed into linguistic terms. Second, the initial solution has great impact on the optimization results. Last, BRB is not necessarily downsized, which is also encountered using the fuzzy set. In this study, The Evidential Reasoning (ER) algorithm [38,45,47] is used as the inference engine to integrate the activated belief rules, and the Differential Evolutionary (DE) algorithm is adopted as the optimization engine. Existing studies have all used deterministic means as the optimization engine.

In contrast to the literature [8,45,47], the referenced values of the attributes are included as the parameters. By using this method, the numerical referenced values of the attributes do not have to be discretized into linguistic terms. Additionally, this method also helps avoid optimization performance degradation. Moreover, the inclusion of the referenced values of the attributes as parameters can also work with the new rule activation procedure as an integrated method for helping to downsize the BRB classifier.

The DE algorithm was first proposed by Storn and Price [31]. Over the past decades, DE has become one of the most promising evolutionary algorithms due to its successful applications in many large-scale, multi-objective, multi-dimensional problems, e.g., weapons production planning [46]. Compared with other Evolutionary Algorithms, such as the Genetic Algorithm, Particle Swarm Optimization and Ant Colony Optimization, DE has a simpler structure, a faster evolving speed and fewer parameters to set [29], which was further validated by the results of multiple computational contest results [26,30]. It is believed that the classical version of DE is sufficient to meet the requirements of this study.

The remainder of this study is organized as follows. The applicability and challenges of BRB in the classification problem are discussed in Section 2. Section 3 introduces new rule activation and weight calculation procedures while the optimization algorithm using ER and DE is proposed in Section 4. The new BRB classifier is proposed in Section 5. Five benchmarks are studied in Section 6 to validate the efficiency and the robustness of the BRB classifier. Section 7 concludes the study.

2. Applicability and challenges for the new BRB classifier

2.1. Applicability of Belief Rule Base in classification problems

The kth rule in a BRB system is described as [7]:

$$\begin{aligned} R_k: & \text{ if } \left(A_1 \text{is} x_1^k\right) \wedge \left(A_2 \text{is} x_2^k\right) \wedge \cdots \wedge \left(A_M \text{is} x_M^k\right), \\ & \text{then } \{(D_1, \beta_{1,k}), \ldots, (D_N, \beta_{N,k})\} \end{aligned} \tag{1}$$

where $A_m (m = 1, \cdots M)$ denotes the mth attribute, $x_m^k (m = 1, \cdots M; k = 1, \cdots K)$ denotes the referenced values of the mth attribute, M denotes the number of the attributes, $\beta_{n,k} (n = 1, \cdots N)$ denotes the belief for the nth degree, D_n, N denotes the number of the degrees.

Let a classification benchmark, iris [32], demonstrate the applicability of BRB in the classification problem.

Consider that there is one set of data, as follows:

$$\{8.1, 4.0, 2.4, 0.3, setosa\} \tag{2}$$

For this set of data, x_1^k is 8.1, x_2^k is 4.0, x_3^k is 2.4, x_4^k is 0.3.

In the iris benchmark example, three different species of iris are to be classified, namely *setosa*, *versicolor* and *virgica*, which could be expressed as

$$D = \{setosa(D_1), versicolor(D_2), virgica(D_3)\} \tag{3}$$

Since this set of data is categorized as "*setosa*", which means that the belief of being "*setosa*" is "1" and the beliefs of being "*versicolor*" and "*virgica*" are both "0", then the belief distribution is as follows:

$$\beta = \{1, 0, 0\} \tag{4}$$

Combining the degree distribution with the corresponding beliefs, we have:

$$\{(D, \beta)\} = \{(D_1, 1), (D_2, 0), (D_3, 0)\} \tag{5}$$

It can then be represented as follows:

$$\begin{aligned} R_k: \text{ if } & (A_1 \text{is} 8.1) \wedge (A_2 \text{is} 4.0) \wedge (A_3 \text{is} 2.4) \wedge (A_4 \text{is} 0.3), \\ & \text{then}\{(D_1, 1), (D_2, 0), (D_3, 0)\} \end{aligned} \tag{6}$$

Following this translation, all sets of data can be translated into belief rules, which further form one BRB system. When there is input, the related rules are activated and integrated into one unified result in the same belief structure.

The conventional rules activation, matching degree calculation and weight calculation procedures are listed in Section 3.1. Readers can refer to Yang et al. [38] and references therein for detailed information.

The following example uses three sets of rules to show how the above procedures are illustrated:

$$\begin{aligned} &R_1: \text{if } (A_1 \text{ is } 5.1) \wedge (A_2 \text{ is } 3.2) \wedge (A_3 \text{ is } 1.0) \wedge (A_4 \text{ is } 0.1), \text{ then } \{(D_1, 1), (D_2, 0), (D_3, 0)\} \\ &R_2: \text{if } (A_1 \text{ is } 7.1) \wedge (A_2 \text{ is } 3.6) \wedge (A_3 \text{ is } 1.4) \wedge (A_4 \text{ is } 0.2), \text{ then } \{(D_1, 0), (D_2, 1), (D_3, 0)\} \\ &R_3: \text{if } (A_1 \text{ is } 8.1) \wedge (A_2 \text{ is } 4.0) \wedge (A_3 \text{ is } 2.4) \wedge (A_4 \text{ is } 0.3), \text{ then } \{(D_1, 0), (D_2, 0), (D_3, 1)\} \end{aligned} \tag{7}$$

The input reads as follows:

$$(6.1,\ 3.3,\ 1.2,\ 0.15) \tag{8}$$

Consider that the initial weights of the three rules are all set to be 1 ($\theta_1 = \theta_2 = \theta_3 = 1$) by Eq. (13) in Section 3.1, and the matching degrees of the attributes and the initial and final weights of each rule are shown in Table 1. The third rule is not activated.

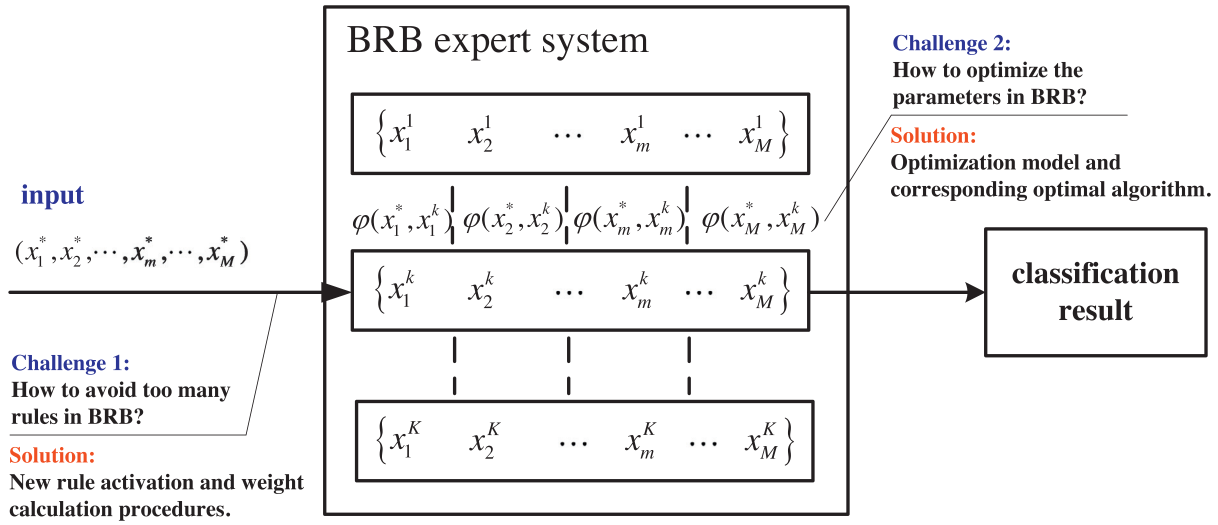

Fig. 1. Main process and challenges of the new BRB classifier.

Table 1
The matching degrees and weights.

	Initial weights θ_k	Matching degree $\varphi(x_i^*, x_{ij})$				Final weights w_k
		1	2	3	4	
1	1	0.5	0.25	0.5	0.5	0.03125
2	1	0.5	0.75	0.5	0.5	0.9375
3	1	0	0	0	0	0

2.2. Challenges of the new BRB classifier

Fig. 1 describes the main process of the new BRB classifier. There are two parts in the BRB classifier. Part I takes in the input, activates the rules and calculates the corresponding activation weights. Part II infers the activated rules and estimates the final class based on the inference result.

To meet the requirements of each specific classification case, two challenges must be addressed, which are also shown in Fig. 1.

The first challenge is that the BRB system must be downsized to avoid the combinatorial explosion because the conventional BRB system requires that all attributes and referenced values of the attributes be covered [38]. The size of the BRB-classifier is:

$$size_{BRB} = \prod_{n=1}^{N} p_n \tag{9}$$

where N denotes the number of the attributes and p_n denotes the number of the referenced values for the nth attribute. The size of the BRB classifier, $size_{BRB}$, would then grow exponentially with the increase of the number of the attributes and/or the number of the referenced values of the attributes. Suppose that there are 5 attributes in the classifier and each attribute has 3 referenced values. The size of the BRB classifier obtained by Eq. (9) would then be 243($=3^5$). The oversized problem is also encountered when using the fuzzy systems [18]. Therefore, new rule activation and the corresponding weight calculation procedures should be proposed.

The second challenge is that the parameters of the BRB classifier must be optimized. The BRB classifier can be initially given by experts or by using historic data or randomly generated if the former two means are not available, and it can be adjusted with experts' intervention. However, it is not guaranteed that the BRB classifier can achieve a high classification precision with any initial parameters. Using any initial parameter requires additional training and learning. Therefore, the new optimal algorithm for the BRB classifier should be proposed.

3. New rule activation and weight calculation procedures

3.1. Conventional rule activation, matching degree calculation and weight calculation

Consider that there is one classification example with two attributes, namely A and B [8]. Each of the two attributes has three referenced values. To cover all of the possible conditions, a traversal BRB system that consists of 9 ($=3\times3$) rules

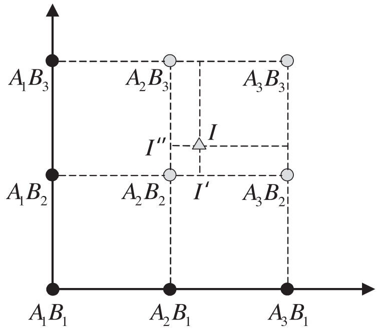

Fig. 2. Conventional rule activation.

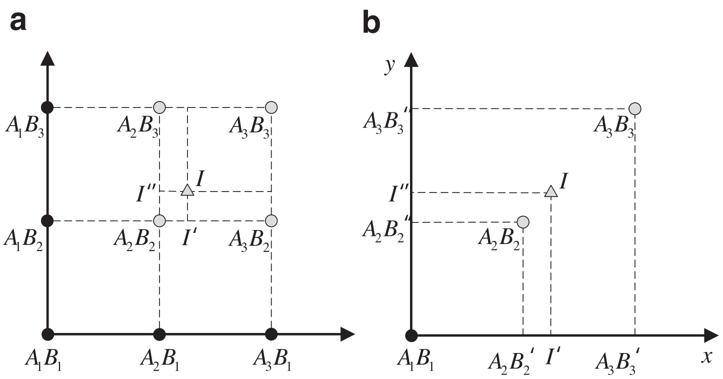

Fig. 3. Conventional and new rule activation.

is constructed. In this condition, suppose that there is one input *I*, as shown in Fig. 2, and four adjacent points/rules are activated by each piece of input data, namely A_2B_3, A_3B_3, A_2B_2 and A_3B_2.

For the input, $(x_1^*, \ldots, x_m^*, \ldots, x_M^*)$, the matching degree of the *i*th attribute in the *j*th rule is calculated by Eq. (10),

$$\varphi(x_i^*, x_{ij}) = \begin{cases} \dfrac{x_{i(k+1)} - x_i^*}{x_{i(k+1)} - x_{ik}} & j = k(x_{ik} \le x_i \le x_{i(k+1)}) \\ \dfrac{x_i^* - x_{ik}}{x_{i(k+1)} - x_{ik}} & j = k+1 \\ 0 & j = 1, 2, \ldots |x_i|, j \ne k, k+1 \end{cases} \tag{10}$$

where x_i^* denotes the value of the *i*th attribute in the input data. $x_{i(k+1)}$ and x_{ik} denote the values of the *i*th attribute in two adjacent activated rules.

The integrated matching degree for the *i*th attribute in the *j*th rule is calculated by Eq. (11),

$$\alpha_{ij} = \frac{\varphi(x_i^*, x_{ij})\varepsilon_i}{\sum \varphi(x_i^*, x_{ij})} \tag{11}$$

where ε_i denotes the confidence of the *i*th attribute being assessed as x_i^*. If there is no incomplete information ($\sum \varphi(x_i^*, x_{ij}) = 1$) and the value of the *i*th attribute is of 100% confidence ($\varepsilon_i = 1$), then:

$$\alpha_{ij} = \varphi(x_i^*, x_{ij}) \tag{12}$$

The activated weight for the *k*th rule is calculated by Eq. (13):

$$w_k = \frac{\theta_k \prod_{i=1}^{M} \alpha_{ik}}{\sum_{k=1}^{K} \theta_k \prod_{i=1}^{M} \alpha_{ik}} \tag{13}$$

where θ_k represents the initial weight of the *k*th rule and α_k represents the matching degree of the *i*th attribute between the input and the *k*th rule.

3.2. New rule activation procedure

The new rule activation procedure is shown in Fig. 3(b), in which only three points/rules are kept, including the two that represent the maximum/minimum values of the attributes (upper right, A_3B_3, and lower left, A_1B_1, in Fig. 3(b)) to define

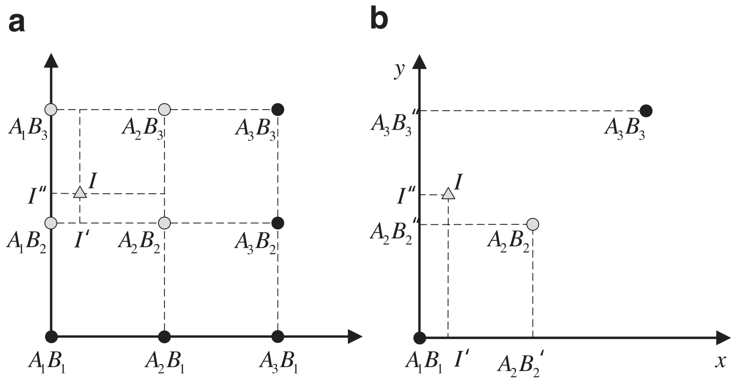

Fig. 4. Conventional and new rule activation—a further discussion.

the boundaries. With this, the input data (as well as the three points/rules) is projected on the *x*/*y* axis first to determine which (two) points/rules are most adjacent and should be activated.

A_2B_2 and A_3B_3 are activated for the specific condition shown in Fig. 3(b) because their projections in the *x* axis, A_2B_2' and A_3B_3', are closer to I', which is the *x* axis projection of *I*. A_2B_2 and A_3B_3 are also activated because their projections in the *y* axis, A_2B_2'' and A_3B_3'', are closer to I'', which is the *y* axis projection of *I*. To conclude, the two points/rules, A_2B_2 and A_3B_3, would be activated.

Eq. (14) can be derived from Fig. 3(a/b) and it shows that the matching degrees would produce the same values using either the conventional or proposed rule activation procedures. Note that the left expression in each equation in Eq. (14) is from Fig. 3(a) and the right expression in each equation is from Fig. 3(b), as in Eq. (14).

$$\begin{cases} A_2B_2I' = A_2B_2'I', I'A_3B_2 = I'A_3B_3' \\ A_2B_2I'' = A_2B_2''I'', A_2B_3I'' = A_3B_3''I'' \end{cases} \tag{14}$$

Fig. 4 describes a more complex condition, following the new rule activation procedure and rendering the projections on both axes first.

For the specific condition shown in Fig. 4(b), A_1B_1 and A_2B_2 are activated because their projections on the *x* axis, A_1B_1 and A_2B_2', are closer to I', which is the *x* axis projection of *I*. A_2B_2 and A_3B_3 are activated because their projections on the *y* axis, A_2B_2'' and A_3B_3'', are closer to I'', which is the *y* axis projection of *I*. To conclude, the three points/rules, A_1B_1, A_2B_2 and A_3B_3, would all be activated.

Eq. (15) can be derived from Fig. 4(a/b) and it shows that the matching degrees would produce the same values using either the conventional or proposed rule activation procedures.

$$\begin{cases} A_1B_2I' = A_1B_1I', I'A_2B_2 = I'A_2B_2' \\ A_1B_2I'' = A_2B_2''I'', I''A_1B_3 = I''A_3B_3'' \end{cases} \tag{15}$$

However, the new rule activation procedure may cause problems in weight calculation. Take the iris benchmark as an example, in which the BRB classifier consists of four rules shown in Table 2.

Suppose that there is one input data that reads as follows:

(5.0000, 2.1000, 5.0000, 2.0000)

For this input data, the corresponding activation weights are shown in Table 3 using conventional procedures, as stated in Section 3.1.

The reason for this dilemma is that the conventional weight calculation process is based on the hypothesis that all of the attributes are “conjunctive”. This would not be a problem because the conventional constructed BRB was traversal. For example, there were 56 rules both before and after the learning process in the pipeline leak detection case because there were seven and eight referenced values for the two attributes, respectively [45,47].

Table 2
The iris_BRB.

No. of rules	Rules weights	Input				Output		
		1	2	3	4	D1	D2	D3
1	0.7086	4.3000	2.0000	1.0000	0.1000	0.6082	0.1511	0.2406
2	0.7672	7.4055	3.5609	3.4109	1.6435	0.1146	0.4799	0.4055
3	0.1342	5.3937	2.2674	3.7394	1.4519	0.0387	0.5665	0.3948
4	0.9873	7.9000	4.4000	6.9000	2.5000	0.4463	0.1177	0.4360

Table 3
Rule weights following conventional weight calculation procedures.

No. of rules	Initial rule weights	Rules matching degrees				Final rule weights
		1	2	3	4	
1	0.7086	0.3600	0.6260	0.0000	0.0000	0.0000
2	0.7672	0.0000	0.0000	0.0000	0.5838	0.0000
3	0.1342	0.6400	0.3740	0.6012	0.0000	0.0000
4	0.9873	0.0000	0.0000	0.3988	0.4162	0.0000

Table 4
Rule weights following new weight calculation procedures.

No. of rules	Initial rule weights	Rule matching degrees				New weights distribution sum	Final rule weights
		1	2	3	4		
1	0.7086	0.3600	0.6260	0.0000	0.0000	0.2465	0.1747
2	0.7672	0.0000	0.0000	0.0000	0.5838	0.1459	0.1120
3	0.1342	0.6400	0.3740	0.6012	0.0000	0.4038	0.0542
4	0.9873	0.0000	0.0000	0.3988	0.4162	0.2038	0.2012

3.3. New weight calculation procedure

The new weight calculation procedure assumes that the attributes are “disjunctive”. The kth rule’s activated weight is calculated by Eq. (16):

$$w_k = \frac{\theta_k \sum_{i=1}^{M} \alpha_{ik}}{\sum_{k=1}^{K} \theta_k \sum_{i=1}^{M} \alpha_{ik}}, \tag{16}$$

where θ_k represents the initial weight of the kth rule, and α_{ik} represents the matching degree of the ith attribute between the input and the kth rule.

Table 4 shows the result using the new rule weights calculation procedure.

Remark 1. An oversized BRB classifier must be downsized due to the following three reasons [5,6,37]:

(1) An oversized BRB classifier cannot be efficiently accessed by experts, e.g., it is improbable for an expert to give hundreds of initial rules in a practical situation.
(2) An oversized BRB classifier may cost too many computational resources, considering that there are multiple parameters in the belief rules.
(3) An oversized BRB classifier may cause over-fitting problems and result in computation and modeling performance degradation.

Remark 2. The proposed rule activation procedure could produce the same matching degrees for the attributes as the conventional procedure.

Remark 3. In the context of the conventional rule weights calculation procedure, $w_k = 0$ when $\prod \alpha_{ik} = 0$, which denotes that the kth rule would **not be** activated even if there is only one matching degree **that is** 0 (one attribute **not being** activated).

Remark 4. In the context of the new rule weights calculation procedure, $w_k \neq 0$ if $\sum \alpha_{ik} \neq 0$, which denotes that the kth rule would **be** activated even if there is only one matching degree **that is not** 0 (one attribute **being** activated).

4. Optimal algorithm for the BRB classifier

4.1. Optimization model for the BRB classifier

First, the parameters to be optimized are defined, which include:

(1) The referenced values of the rules. For the kth rule, its ith-referenced value for the ith attribute, x_i^k, must meet the following restraints:

$$lb_i \le x_i^k \le ub_i, \quad k = 1, \ldots, K; i = 1, \ldots, M \tag{17a}$$

$$x_i^1 = lb_i \tag{17b}$$

$$x_i^K = ub_i \tag{17c}$$

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

(2) Initial weight of the rules. For the kth rule, its initial weight, θ_k, must meet the following restraint:

$$0 \le \theta_k \le 1 \tag{17d}$$

Note that θ_k is not the final weight w_k, which should be multiplied by the activated weight of the kth rule, $w_k = \theta_k * w_{activated,k}$.

(3) The beliefs in the conclusion part of each rule. For the kth rule, suppose that there is no inadequate information in the conclusion part, and therefore the belief, $\beta_{n,k}$, must meet the following constraints:

$$0 \le \beta_{n,k} \le 1, \quad n = 1, \ldots, N \tag{17e}$$

$$\sum_{n=1}^{N} \beta_{n,k} = 1 \tag{17f}$$

Second, the optimization objective function is given. The output of the BRB classifier exists in the same belief structure as that of the input rules:

$$S(f(x)) = \{(D_n, \beta_{n,k}), (n = 1, \ldots, N; k = 1, \ldots, K)\} \tag{18}$$

Define the mapping function between the derived class and the corresponding belief values as:

$$D(n) = \beta_n \tag{19}$$

Then, we have:

$$D^{-1}(\beta_n) = n \tag{20}$$

Select the class with the largest belief value as the final derived result:

$$\hat{n} = D^{-1}(max(\beta_n)) \tag{21}$$

Assume that the actual class in the benchmark case is n. Then, for each set of input data, the error is “0” if the derived result is the same as that from the actual result, and it is “1” if the two results disagree:

$$E = \begin{cases} 1 & if\ \hat{n} \ne n \\ 0 & if\ \hat{n} = n \end{cases} \tag{22}$$

As there is only “0” and “1” in the error for one single set of data, there would be no difference between the mean square error (MSE) and the mean absolute error (MAE), as proved by Eq. (23):

$$MSE(x_i^k, \theta_k, \beta_{n,k}) = \frac{1}{T}\sum_{t=1}^{T} E^2 = \frac{1}{T}\sum_{t=1}^{T} E = MAE(x_i^k, \theta_k, \beta_{n,k}) \tag{23}$$

where T is the size of the dataset.

Finally, the optimization model is given in Eq. (23), and the restraints are given in Eq. (17a–f):

$$\min MSE(x_i^k, \theta_k, \beta_{n,k}) \tag{23}$$

s.t.

$$lb_i \le x_i^k \le ub_i, \quad k = 1, \ldots, K; i = 1, \ldots, M \tag{17a}$$

$$x_i^1 = lb_i \tag{17b}$$

$$x_i^K = ub_i \tag{17c}$$

$$0 \le \theta_k \le 1 \tag{17d}$$

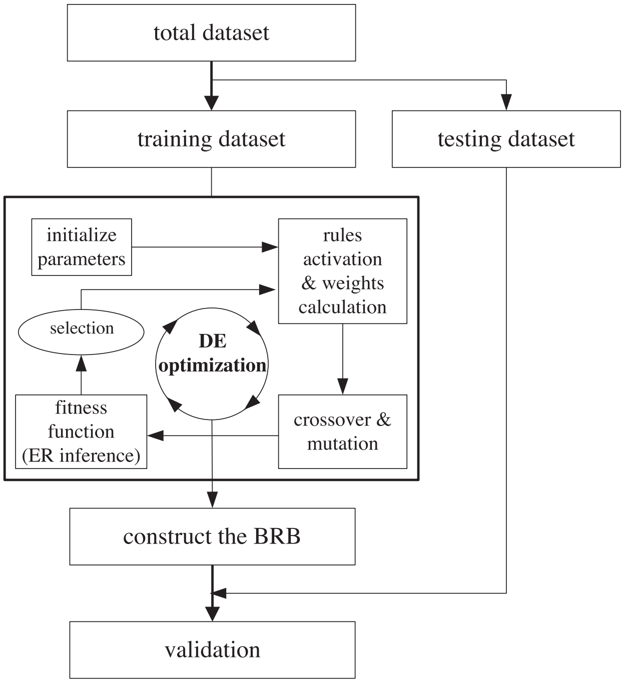

Fig. 5. Optimal algorithm for training the BRB classifier using DE.

$$0 \le \beta_{n,k} \le 1, \quad n = 1, \ldots, N \tag{17e}$$

$$\sum_{n=1}^{N} \beta_{n,k} = 1 \tag{17f}$$

Remark 5. For Eq. (17f), it is assumed that there is no inadequate information in the conclusion part in each rule and thus the sum of the belief is equal to one. Although the original definition of the D-S evidence theory assumes the sum of the beliefs to be less than or equal to one and to include incomplete information [38,45], the assumption that there is no inadequate information has been used in both theoretical studies, such as nonlinear function fitting [7] studies, and in practical studies, such as pipeline leak detection [47] studies.

Remark 6. The number of rules in BRB is not taken as the parameters to be optimized within the scope of this manuscript. Two measures could be applied to ensure the effectiveness of the rules when constructing BRB. Zhou introduced the concept of "statistical utility" to automatically calculate for each rule to decide whether a rule should exist or new rules should be added in the BRB system [45]. Yang used the DBSCAN technique to automatically reduce redundant rules when BRB is under-fitting and the error analysis technique to automatically add new rules when BRB is over-fitting [37].

4.2. Optimal algorithm for the BRB classifier using ER and DE

To conduct a more thorough comparison, the cross-validation technique is applied. It was first proposed by Efron & Tibshirani and can minimize the effects of sample characteristics, particularly in dealing with cases studies with small data samples [11]. In the l-fold cross validation, the initial samples are divided into l parts of equal size. For each training time, one of the l parts is left as the testing dataset, and the rest of the l-1 parts are taken as the training dataset. Therefore, training/testing is iterated l times in l-fold cross validation. Finally, the average of the results derived from l runs is taken as the final result to correct the optimistic nature of training error and derive a more accurate estimate of model prediction performance [12].

Kohavi [17] reviewed and compared the most common methods for accuracy estimation, including holdout, cross-validation and bootstrap, and concluded that ten-fold(10-fold) cross-validation was the most efficient for model selection because (1) it outperformed the more expensive leave-one-out and (2) more partitions (such as 15-fold, 20-fold) are too expensive. Examples of the application of cross-validation in classification problems can be found in numerous studies [2,10,14,24,25,27,28,44,48–50]. The 10/5/2-fold cross-validations are applied in this study to ensure thoroughness and comprehensiveness. (Fig. 5)

The main steps of the optimal algorithm using ER and DE as the inference and optimization engines are as follows:

Step 1: Divide the initial dataset into the training and testing datasets by 10/5/2 folds.

Step 2: Parameter initialization.

The parameters to be initiated include the parameters for the BRB classifier and the DE algorithm.

Parameters for the BRB classifier include the referenced values of the attributes, the initial weights of the rules and the beliefs of the scales in the conclusion part of all of the rules. Parameters must meet the restraints given in Eq. (17a–f).

Parameters for the DE algorithm include the number of individuals, the number of the generation and the upper and lower boundaries of the attributes/weights/degrees.

All parameters are initiated as individuals comprised of decimal coded genes and do not need to be decoded.

Step 3: Rule activation and weight calculation.

New rule activation and weight calculation procedures are introduced in Sections 3.1 and 3.2 to identify the activated rules as well as their corresponding weights.

Step 4: Crossover and mutation.

Step 4.1: Crossover

The crossover strategy states that the jth gene, $v'_{i,j}$, of a temporary individual, v, is selected by the probability of CR (or the jth gene, $x'_{i,j}$, of the current individual, x, is selected by the probability of $1-CR$) as the jth gene, $u'_{i,j}$, of the final individual, u, as shown in Eq. (24),

$$u'_{i,j} = \begin{cases} v'_{i,j} & \text{if } (rand \le CR) or (j = sn) \\ x'_{i,j} & \text{otherwise} \end{cases} \tag{24}$$

where $CR = 0.9$ is the crossover operator and $sn \in [1, 2, \ldots, n]$ is a random integer that is generated with each new individual.

Step 4.2: Mutation

The ith individual in the new generation, v'_i, can be obtained using Eq. (25):

$$v'_i = x_{r1} + F * (x_{r2} - x_{r3}) \tag{25}$$

where x_{r1}, x_{r2} and x_{r3} are three random individuals, $r1 \ne r2 \ne r3$ and $F = 0.5$ is the mutation operator.

Step 5: Fitness function calculation using ER as the optimization engine.

After certain rules are activated, the activated L rules are integrated using ER, and the analytic form of ER is given in Eqs. (26) and (27) [33]:

$$\beta_n = \frac{\mu\left[\prod_{k=1}^{L}\left(w_k\beta_{n,k} + 1 - w_k\sum_{n=1}^{N}\beta_{n,k}\right) - \prod_{k=1}^{L}\left(1 - w_k\sum_{n=1}^{N}\beta_{n,k}\right)\right]}{1 - \mu\left[\prod_{k=1}^{L}(1 - w_k)\right]} \tag{26}$$

$$\mu = \left[\sum_{n=1}^{N}\prod_{k=1}^{L}\left(w_k\beta_{n,k} + 1 - w_k\sum_{s=1}^{N}\beta_{s,k}\right) - (N-1)\prod_{k=1}^{L}\left(1 - w_k\sum_{s=1}^{N}\beta_{s,k}\right)\right]^{-1} \tag{27}$$

where β_n represents the belief for the nth degree.

Step 6: Selection

The ith individual u_i^t enters the new generation when its fitness function obtains a higher rated value, as indicated by Eq. (28):

$$x_i^{t+1} = \begin{cases} u_i^t & \text{if } f(u_i^t) \le f(x_i^t) \\ x_i^t & \text{otherwise} \end{cases} \tag{28}$$

where $f(\bullet)$ is the fitness function, which in this study is the MSE/MAE in Eq. (23).

Step 7: Check on the stop criterion.

Check on whether the stop criterion (number of generations) is met. If not met, go to Step 3; if met, the individual(s) with the smallest MSE would be selected as the ultimate optimization result.

Step 8: Construct the new BRB classifier using the derived solution.

Step 9: Efficiency validation using the testing dataset.

5. New BRB classifier

Based on the above description, Fig. 6 describes the new BRB classifier.

As introduced in Section 2.2, there are two parts to the BRB classifier. Part I (steps 1–3) determines the rules to be activated and calculates the corresponding activation weights. Part II (steps 4–5) infers the activated rules using the ER algorithm and estimates the final class based on the inference result.

Correspondingly, five steps are given as follows:

Step 1: Rule activation. For x_m^* in the input, $(x_1^*, \ldots, x_m^*, \ldots, x_M^*)$, make projections in each axis and determine the adjacent rules shown in Section 3.2.

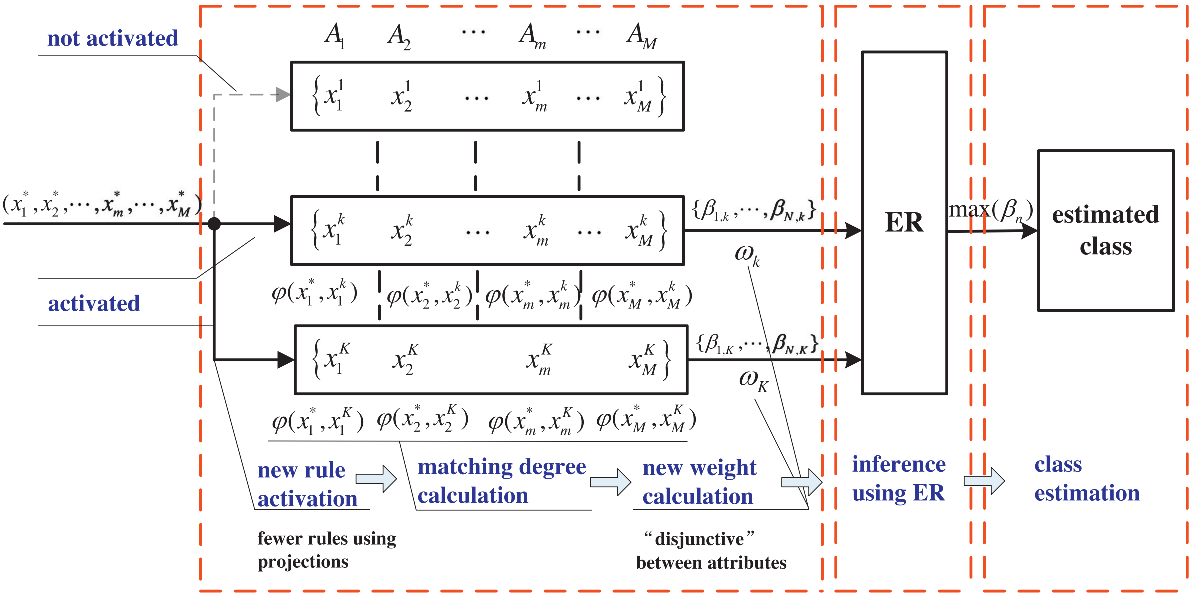

Fig. 6. New BRB classifier.

Table 5
Statistics on iris/wine/glass benchmarks.

Names	No. of attributes	No. of classes	No. of samples	Size of BRB before downsizing	Size of BRB after downsizing
Iris	4	3	150	4^4	4
Wine	13	3	178	3^{13}	3
Glass	9	7	214	3^9	3
Cancer	30	2	569	3^{30}	3
Pima	8	2	768	8^4	4

Step 2: Matching degree calculation. For the kth activated rules, calculate the matching degrees, $\varphi(x_m^*, x_m^k)$ by Eq. (10).

Step 3: Weight calculation. For the kth activated rules, calculate the activation weight, ω_k, with consideration of the initial weights of the rules, θ_k, by Eq. (16) in Section 3.3.

Step 4: Inference using ER by Eqs. (26) and (27) in Section 4.2. The inputs of ER are the activated weights, ω_k, and the conclusion part of the activated rules, $\beta_{n,k}(n = 1, \ldots, N)$. The output of ER is $\beta_n(n = 1, \ldots, N)$, which is still in the same belief structure as the input. The parameters are determined by using the optimal algorithm derived in Section 4.2.

Step 5: Class estimation. The class in the inference result with the biggest belief is selected as the final estimated result by Eq. (21) in Section 4.1.

Remark 7. The new BRB classifier is given in a general sense rather than only based on the DE algorithm which is the optimization engine in this study. Other deterministic approaches, DE-based variants or evolutionary algorithms can be tested in future comparative studies.

6. Case study

Five benchmarks obtained from the University of California at Irvine [32], including iris, wine, glass, cancer and pima, are studied to validate the efficiency of the proposed BRB classifier. Table 5 summarizes the number of attributes, classes, samples and sizes before and after downsizing BRB classifier. Under the assumption that each attribute has three referenced values including their upper and lower boundaries, Table 5 shows that the size of BRB after downsizing is significantly smaller when using the proposed rule activation and weight calculation procedures.

Note that it is neither the intent nor motive of this study to identify the optimal size of the BRB model, which calls for future research efforts.

For each instance, 30 independent runs with 40 individuals and 2000 generations are performed. The proposed BRB classifier is implemented in Matlab R2010b on Core (TM) i5-2450 M CPU 2.50 GHz with Windows 7.

Table 6a
Results statistics of iris/wine/glass/cancer/pima with 10-fold cross-validation.

						Iris						
Fold	1	2	3	4	5	6	7	8	9	10	Average	Var.
Acc.	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	0
Err.	0	0	0	0	0	0	0	0	0	0	0	
						Wine						
Fold	1	2	3	4	5	6	7	8	9	10	Average	Var.
Acc.	100%	100%	100%	100%	100%	94%	100%	100%	100%	100%	99%	3.09E−04
Err.	0	0	0	0	0	10	0	0	0	0	1	
						Glass						
Fold	1	2	3	4	5	6	7	8	9	10	Average	Var.
Acc.	71%	76%	71%	71%	76%	67%	67%	67%	67%	67%	70%	1.54E−03
Err.	61	51	61	61	51	71	71	71	71	71	64.20	
						Cancer						
Fold	1	2	3	4	5	6	7	8	9	10	Average	Var.
Acc.	98%	96%	96%	100%	98%	100%	100%	98%	100%	96%	98%	2.36E−04
Err.	1	2	2	0	1	0	0	1	0	2	0.9	
						Pima						
Fold	1	2	3	4	5	6	7	8	9	10	Average	Var.
Acc.	82%	73%	84%	75%	73%	81%	81%	79%	86%	78%	79%	1.99E−03
Err.	14	21	12	19	21	15	15	16	11	17	16.10	

Note: Acc. = accuracy; Err. = error; Var. = variance. Accuracy is in percentage and error is in quantity.

Table 6b
Results statistics of iris/wine/glass/cancer/pima with 5-fold cross-validation.

				Iris			
Fold	1	2	3	4	5	Average	Var.
Acc.	100%	100%	100%	100%	100%	100%	0
Err.	0	0	0	0	0	0	
				Wine			
Fold	1	2	3	4	5	Average	Var.
Acc.	97%	100%	97%	97%	91%	97%	9.80E−04
Err.	5	0	5	5	15	6.10	
				Glass			
Fold	1	2	3	4	5	Average	Var.
Acc.	70%	70%	72%	72%	63%	69%	1.46E−03
Err.	65	65	60	60	80	65.69	
				Cancer			
Fold	1	2	3	4	5	Average	Var.
Acc.	96%	98%	95%	95%	97%	96%	2.46E−04
Err.	4	2	6	6	3	4.2	
				Pima			
Fold	1	2	3	4	5	Average	Var.
Acc.	78%	78%	71%	80%	85%	78%	2.53E−03
Err.	34	34	43	31	23	33	

6.1. Benchmark results

Tables 6a–6c, show the results of the five benchmarks with 10/5/2-fold cross-validation, respectively.

For iris, 10/5-fold cross-validations can produce 100% accuracy. Even with 2-fold cross-validation, 99% accuracy is achieved. A BRB classifier for iris with 100% accuracy, namely iris_BRB, is shown in Table 2.

For wine, results with 10/5/2-fold cross-validations produce more than 90% accuracy, with the highest accuracy being 99% with 10-fold cross-validation. For glass, 70, 67 and 60% accuracies are produced with10/5/2-fold cross-validations, respectively. For cancer, 98, 96 and 95% accuracies are produced with10/5/2-fold cross-validations, respectively. For pima, 79, 78 and 77% accuracies are produced with10/5/2-fold cross-validations, respectively.

Regarding the results with different cross-validations, the accuracy of each benchmark increases with the increasing of partitions. The results derived from the 10-fold cross-validation surpass those derived from the 5-fold cross-validation and further surpass those derived from the 2-fold cross-validation. These results are partially due to the use of a larger amount of training data (90% of the full dataset), which improves the accuracy of the BRB classifier.

Regarding the variances of the accuracy of each partition with all cross-validations, each benchmark with each cross-validation reaches a small variance compared with the accuracy. As for different benchmarks and cross-validations, no novel differences are observed because they are all independent experiments.

Table 6c
Result statistics of iris/wine/glass/cancer/pima with 2-fold cross-validation.

		Iris		
Fold	1	2	Average	Var.
Acc.	99%	99%	99%	0
Err.	2	2	2	
		Wine		
Fold	1	2	Average	Var.
Acc.	89%	93%	91%	1.01E−03
Err.	20	12	16.00	
		Glass		
fold	1	2	Average	Var.
Acc.	64%	57%	60%	2.14E−03
Err.	78	92	85.00	
		Cancer		
fold	1	2	Average	Var.
Acc.	95%	96%	95%	9.85E−05
Err.	15	11	13	
		Pima		
fold	1	2	Average	Var.
Acc.	75%	78%	77%	8.00E−04
Err.	81	99	90.00	

Table 7
Comparison of results of selected studies and the BRB classifier.

Name		Benchmark				
Abbreviations	Core supporting theory	Iris	Wine	Glass	Cancer	Pima
KNN[2]	*K* nearest neighbor	95.33%	72.47%	74.29%		
Spectral CAT[10]	Categorical spectral clustering	97.00%		70.00%	97.00%	
HFM[14]	Fuzzy set	97.33%	**99.44%**			
LST-KSVC[24]	Neural network	99.27%	94.27%	65.76%		
BSREC[25]	Blood sugar regulation based evolutionary algorithm	96.62%	97.31%			
FGGCA[27]	Fuzzy set	97.22%	97.10%	**93.65%**		
WLTSVM[28]	Support vector machine	98.00%	96.40%	49.91%		
CMQFS[44]	Feature vector graph		98.99%	70.06%		
LMT[48]	Logistic model tree	94.67%			96.49%	77.22%
SMO[49]	Fuzzy & neural network	96.00%			97.07%	77.08%
HHONC[50]	Neural network	97.33%			97.22%	**79.15%**
BRB classifier	Belief rule base	**100**%	**99.44**%	70.09%	**98.42%**	79.10%

6.2. Comparison with conventional approaches

To further verify the validity of the proposed BRB classifier, the results derived in Section 4.2 (with 10-fold cross-validation) are further compared with studies that included at least two of the five benchmarks, as shown in Table 7.

By comparing the results of selected studies and the BRB classifier in Table 7, it has been validated that the proposed BRB classifier can produce satisfactory results. For iris and cancer, the 100 and 98.42% accuracies have outperformed all listed studies. For wine, 99.44% accuracy is also the best result, which is derived by HFM (fuzzy set). For glass and pima, although it may seem less satisfactory, they still reach the third and the second best accuracies, respectively, compared to other studies.

Note that HFM and FGGCA, two approaches supported by fuzzy set, have both produced satisfactory results. FGGCA produced 93.65% accuracy for the glass dataset, whereas the BRB classifier only produced 70.09%. This is partially because there are more classes in the glass dataset than in the iris/wine datasets (7 classes versus 3 classes). And BRB has shown inferior performance to the fuzzy set in handling this condition.

Remark 8. There is no universally applicable approach/technique that can reach the best results for all benchmarks, partially because there are different internal structures among different benchmarks. Future research into the combination of multiple classifiers with superior performance is necessary.

6.3. Robustness study of the BRB classifier

Because the DE algorithm is applied as the optimization engine, Fig. 7(a–c) shows the evolving process of 2000 generations for the five benchmarks with 10/5/2-fold cross-validations. It is shown that the error can rapidly diminish to a satisfactory value, which verifies the efficiency of the proposed BRB classifier. Fig. 7(a–c) also shows the last 1000 generations for the iris/wine/glass benchmarks with different cross-validations.

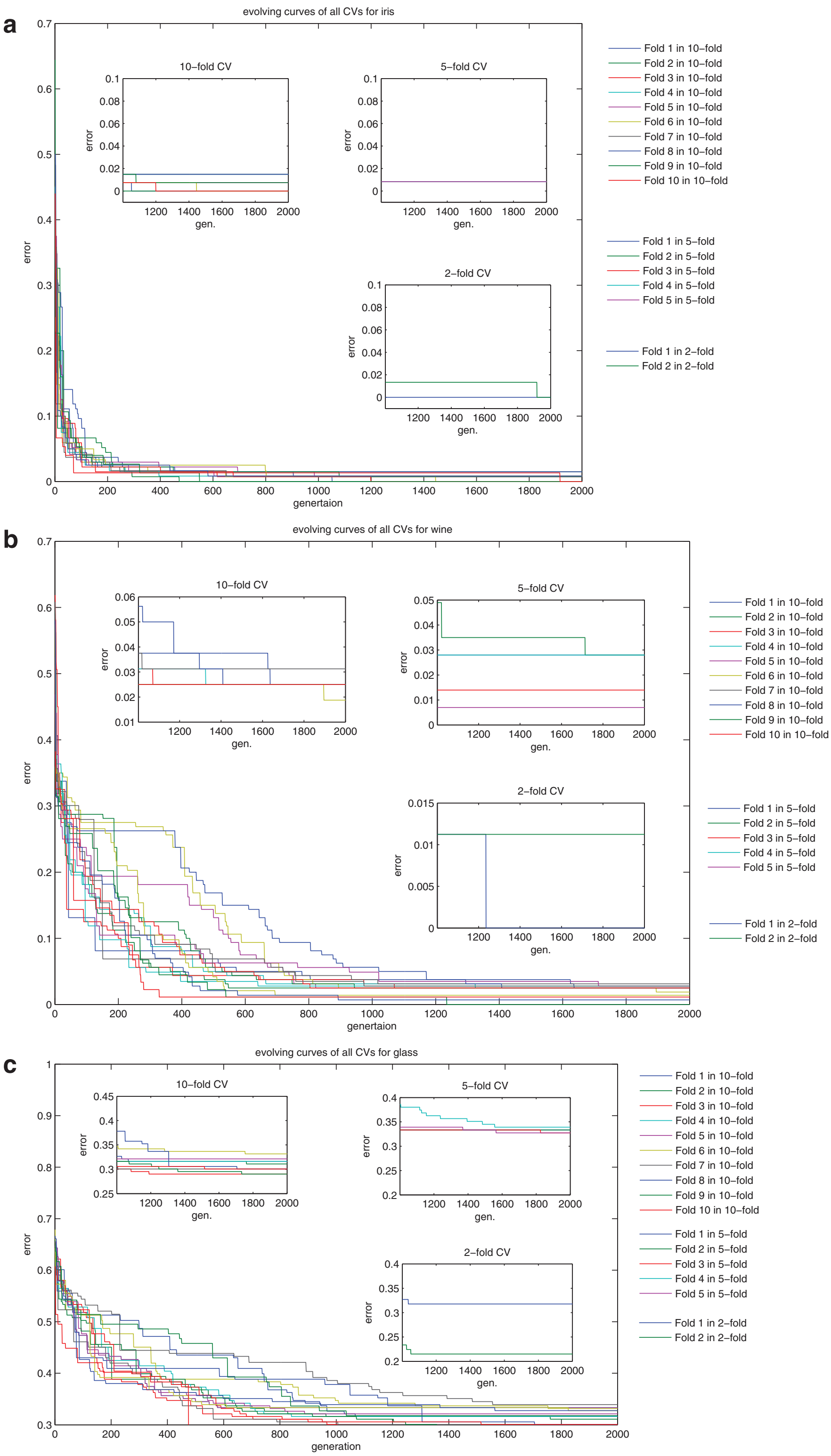

Fig. 7. (a) The evolving process using DE on the iris dataset. (b) The evolving process using DE on the wine dataset. (c) The evolving process using DE on the glass dataset.

To further verify the robustness and stability of the proposed BRB classifier, additional statistics on the iris/wine/glass benchmarks are listed in Tables A.1–A.3 in the Appendix. These statistics indicate that the variances of the accuracy are far smaller than the maximum/minimum values of the accuracy in the five benchmarks. To summarize, the fact that the benchmark experiment accuracies are stable proves the validity of the DE algorithm as the optimization engine in the BRB classifier.

6.4. Discussion

On the glass dataset, the derived accuracy is comparatively less satisfying, partially because there are more classes in this dataset (7 classes) than in the other dataset (2 or 3 classes). When there are more classes to be recognized, the parameters to be optimized grow according to which parameters would compromise the learning efficiency and further lower the accuracy. Comparatively, the increase in the number of attributes and samples does not seem to be a problem because the accuracy of the cancer (with more attributes) and pima (with more samples) benchmarks stays very high. However, it is nearly impossible for one single classifier to achieve the highest accuracy on all benchmarks, partially because these benchmarks are derived from different backgrounds and have different features. Therefore, specialized strategies for classification problems are necessary

The case study results demonstrate and verify the efficiency of the proposed BRB classifier. Among the selected five benchmarks, the iris/wine/cancer benchmarks have relatively satisfactory results, whereas the result for the glass benchmark requires improvement. The fuzzy set and the traditional K nearest neighbor demonstrated better performance, leading to the possibility of combining these different approaches to generate a new and more powerful classifier with advantages derived from different components.

7. Conclusion

In this study, a BRB classifier is proposed to solve the classification problem. With new rule activation and weight calculation procedures, the BRB classifier employs the ER and DE algorithms as the inference and the optimization engines, respectively. Five benchmarks, namely iris, wine, glass, cancer and pima, are tested with 10/5/2-fold cross-validations to validate the efficiency of the proposed BRB classifier.

The case study result shows that all benchmarks can be precisely classified by the BRB classifier in very small sizes (three to four rules for each benchmark), which are further confirmed by comparisons with the results of existing studies using conventional approaches and/or techniques. It is also shown that the proposed BRB classifier achieves better results with more partitions (10-fold cross-validation) than with fewer partitions (5/2-fold cross-validations).

Robustness studies on the efficiency of DE as the optimization engine validate DE as an appropriate and feasible choice for the optimization goal. Statistics show that errors decrease sharply in the evolving process in which a rather small error value can be reached within the first few hundred generations and remain stable in later generations. Moreover, the variances of the errors are far smaller than their maximum/minimum values, demonstrating that there is little aberration between the maximum and minimum error in each run with 10/5/2 cross-validations. This proves the robustness of DE applied in the BRB classifier.

Additional classification benchmarks should be studied in the future to more comprehensively verify the efficiency of the BRB classifier. With this research possibility in mind, other variants of DE must be proposed because the complexity of each benchmark may grow exponentially with more attributes. Moreover, extensive studies into the combination of more expert systems-based approaches and techniques may help discover new classifiers that are applicable to more conditions.

Acknowledgments

Lei-lei Chang thanks the NSFC for its partial support under grant 71201168, grant 61305084 and grant 61403404, and the open funding program of the Joint Laboratory of Flight Vehicle Ocean-based Measurement and Control under grant FOM2015OF017. Zhi-Jie Zhou thanks the NSFC for its partial support under grant 61370031 and grant 61374138, the Postdoctoral Science Foundation of China under grant 2015M570847, the Natural Science Foundation of Shaanxi Province under grant 2015JM6354, the open funding program of the Joint Laboratory of Flight Vehicle Ocean-based Measurement and Control under grant FOM2014OF14, and the Assembly Research Foundation under grant 9140A19030314JB47276.

Appendix

A.1. Statistics on the min/max/avg/var of iris/wine/glass

Table A.1
Statistics on the min/max/avg/var of iris.

		Min	Max	Avg	Var
10-fold	1	0	6.6667E−02	3.4921E−02	1.1640E−03
	2	0	1.3333E−01	5.0794E−02	2.1799E−03
	3	0	1.3333E−01	3.8095E−02	2.0317E−03
	4	0	1.3333E−01	3.3333E−02	1.6374E−03
	5	0	1.3333E−01	2.1333E−02	1.3778E−03
	6	0	2.6667E−01	7.1264E−02	4.4226E−03
	7	0	1.3333E−01	3.5556E−02	2.4550E−03
	8	0	2.0000E−01	4.1667E−02	3.4058E−03
	9	0	1.3333E−01	3.2000E−02	2.2667E−03
	10	0	1.3333E−01	4.1379E−02	1.4012E−03
5-fold	1	0	1.3333E−01	6.6667E−02	1.3072E−03
	2	0	1.0000E−01	4.0909E−02	9.4517E−04
	3	0	1.0000E−01	4.0000E−02	1.3567E−03
	4	0	1.6667E−01	5.2874E−02	1.5490E−03
	5	0	1.3333E−01	4.2857E−02	1.3492E−03
2-fold	1	1.3333E−02	1.2000E−01	5.7012E−02	5.9551E−04
	2	1.3333E−02	1.2000E−01	6.1609E−02	8.8145E−04

Table A.2
Statistics on the min/max/avg/var of wine.

		min	max	avg	var
10-fold	1	0	3.8889E−01	1.1852E−01	9.0961E−03
	2	0	4.4444E−01	1.0556E−01	9.6530E−03
	3	0	2.7778E−01	1.0556E−01	5.3959E−03
	4	0	3.8889E−01	1.4259E−01	1.2278E−02
	5	0	2.7778E−01	1.1482E−01	5.7330E−03
	6	5.5556E−02	4.4444E−01	1.5556E−01	1.1154E−02
	7	0	5.0000E−01	1.6111E−01	1.6677E−02
	8	0	2.7778E−01	1.3333E−01	8.4290E−03
	9	0	3.3333E−01	1.3519E−01	7.8080E−03
	10	0	2.7778E−01	1.3333E−01	5.2360E−03
5-fold	1	2.8571E−02	4.0000E−01	1.5810E−01	1.0400E−02
	2	0	3.4286E−01	1.3048E−01	4.9927E−03
	3	2.8571E−02	1.4286E−01	8.6667E−02	1.0969E−03
	4	2.8571E−02	2.0000E−01	1.1429E−01	2.0830E−03
	5	8.5714E−02	2.2857E−01	1.2286E−01	1.4722E−03
2-fold	1	1.1236E−01	3.1461E−01	1.7079E−01	2.9028E−03
	2	6.7416E−02	2.4719E−01	1.2584E−01	2.1105E−03

Table A.3
statistics on the min/max/avg/var of glass.

		min	max	avg	var
10-fold	1	2.8571E−01	5.7143E−01	4.4127E−01	6.7141E−03
	2	2.3810E−01	5.2381E−01	4.0000E−01	8.6950E−03
	3	2.8571E−01	5.2381E−01	4.1905E−01	5.8488E−03
	4	2.8571E−01	6.1905E−01	4.5397E−01	1.0905E−02
	5	2.3810E−01	6.1905E−01	4.1270E−01	1.4283E−02
	6	3.3333E−01	5.7143E−01	4.8254E−01	6.0573E−03
	7	3.3333E−01	6.1905E−01	4.6825E−01	7.6760E−03
	8	3.3333E−01	6.6667E−01	4.8413E−01	8.6140E−03
	9	3.3333E−01	5.2381E−01	4.3016E−01	3.9850E−03
	10	3.3333E−01	5.7143E−01	4.8254E−01	4.6500E−03
5-fold	1	3.0233E−01	6.7442E−01	4.7829E−01	7.5226E−03
	2	3.0233E−01	6.9767E−01	4.6589E−01	1.3930E−02
	3	2.7907E−01	5.8140E−01	4.2713E−01	5.0161E−03
	4	2.7907E−01	6.2791E−01	4.7209E−01	8.9573E−03
	5	3.7209E−01	6.5116E−01	4.5659E−01	5.7620E−03
2-fold	1	3.6449E−01	6.2617E−01	5.0062E−01	4.4617E−03
	2	4.2991E−01	6.1682E−01	5.0467E−01	2.9938E−03

References

[1] S. Abbasbandy, M. Otadi, M. Mosleh, Numerical solution of a system of fuzzy polynomial by fuzzy network, Inf. Sci. 178 (8) (2008) 1948–1960.
[2] M. Aci, M. Avci, *K* nearest neighbor reinforced expectation maximization method, Expert Syst. Appl. 38 (2011) 12585–12591.
[3] S. Afshar, M. Mosleh, M. Kheyrandish, Presenting a new multiclass classifier based on learning automata, Neurocomputing 104 (2013) 97–104.
[4] L.L. Chang, M.J. Li, B. Cheng, P. Zeng, Integration-centric approach to system readiness assessment based on evidential reasoning, J. Syst. Eng. Electron. 23 (6) (2012) 881–890.
[5] L.L. Chang, J.B. Sun, J. Jiang, M.J. Li, Parameter learning for the belief rule base system in the residual life probability prediction of metalized film capacitor, Knowl.-Based Syst. 73 (2015) 69–80.
[6] L.L. Chang, Y. Zhou, J. Jiang, M.J. Li, X.H. Zhang, Structure learning for belief rule base expert system: a comparative study, Knowl.-Based Syst. 39 (2013) 159–172.
[7] Y.W. Chen, J.B. Yang, C.C. Pan, D.L. Xu, Z.J. Zhou, Identification of uncertain nonlinear systems: constructing belief rule-based models, Knowl.-Based Syst. 73 (2015) 124–133.
[8] Y.W. Chen, J.B. Yang, D.L. Xu, S.L. Yang, On the inference and approximation properties of belief rule based systems, Inf. Sci. 38 (2013) 121–135.
[9] G.R. Dattatreya, V.V. Sarma, Bayesian and decision tree approaches for pattern recognition including feature measurement costs, IEEE Trans. Pattern Anal. Mach. Intell. 3 (1981) 293–298.
[10] G. David, A. Averbuch, SpectralCAT: categorical spectral clustering of numerical and nominal data, Pattern Recognit. 45 (2012) 416–433.
[11] B. Efron, R. Tibshirani, An Introduction to the Bootstrap, Chapman & Hall, 1993.
[12] R. Grossman, G. Seni, J. Elder, N. Agarwal, H. Liu, Ensemble Methods in Data Mining: Improving Accuracy Through Combining Predictions, Morgan & Claypool, 2010.
[13] N.R. Guo, T.H.S. Li, Construction of a neuron-fuzzy classification model based on feature-extraction approach, Expert Syst. Appl. 38 (2011) 682–691.
[14] N. R.Guo, T.S. Li, C.L. Kuo, Design of hierarchical fuzzy model for classification problem using gas, Comput. Ind. Eng. 50 (2006) 90–104.
[15] P. Jaganathan, R. Kuppuchamy, A threshold fuzzy entropy based feature selection for medical database classification, Comput. Biol. Med. 43 (21) (2013) 2222–2229.
[16] J. Jiang, X. Li, Z.J. Zhou, D.L. Xu, Y.W. Chen, Weapon system capability assessment under uncertainty based on the evidential reasoning approach, Expert Syst. Appl. 38 (2011) 13773–13784.
[17] R. Kohavi, A study of cross-validation and bootstrap for accuracy estimation and model selection, in: Proceedings of the Joint Conference on Artificial Intelligence, 1995, pp. 1137–1145.
[18] X.D. Liu, X.C. Wang, W. Pedrycz, Fuzzy clustering with semantic interpretation, Appl. Soft Comput. 26 (2015) 21–30.
[19] Z.W. Lu, L.W. Wang, Learning descriptive visual representation for image classification and annotation, Pattern Recognit. 48 (2015) 498–508.
[20] P. Melin, O. Castillo, A review on type-2 fuzzy logic applications in clustering, classification and pattern recognition, Appl. Soft Comput. 21 (2014) 568–577.
[21] R. Meyer, S. O'Keefe, A fuzzy binary neural network for interpretable classifications, Neurocomputing 121 (2013) 401–415.
[22] J.A. Nasiri, N.M. Charkari, S. Jalili, Least squares twin multi-class classification support vector machine, Pattern Recognit. 48 (2015) 984–992.
[23] T. Nguyen, A. Khosravi, D. Creighton, S. Nahavandi, Classification of healthcare data using genetic fuzzy logic system and wavelets, Expert Syst. Appl. 42 (2015) 2184–2197.
[24] Q.F. Nie, L.Z. Jin, S.M. Fei, J.Y. Ma, Neural network for multi-class classification by boosting composite stumps, Neurocomputing 149 (2015) 949–956.
[25] S.C. Pandey, C.N. Nandi, Blood sugar regulation based evolutionary algorithm for data classification, Appl. Soft Comput. 12 (2012) 2266–2273.
[26] K. Price, Differential evolution vs. the functions of the 2 ICEO, in: Proceedings of 1997 IEEE International Conference on Evolutionary Computation, 1997, pp. 153–157.
[27] M.A. Sanchez, O. Castillo, J.R. Castro, P. Melin, Fuzzy granular gravitational clustering algorithm for multivariate data, Inf. Sci. 279 (2014) 498–511.
[28] Y.H. Shao, W.J. Chen, Z. Wang, C.N. Li, N.Y Deng, Weighted linear loss twin support vector machine for large-scale classification, Knowl.-Based Syst. 73 (2015) 276–288.
[29] R. Storn, On the usage of differential evolution for function optimization, in: Proceedings of Biennial Conference of the North American Fuzzy Information Processing Society (AFIPS), 1996, pp. 519–523.
[30] R. Storn, K. Price, Minimizing the real functions of the ICEC 1996 contest by differential evolution, in: Proceedings of 1996 IEEE International Conference on Evolutionary Computation, 1996, pp. 842–844.
[31] R. Storn, K. Price, Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces, J. Global Optim. 11 (4) (1997) 341–359.
[32] UCI repository of machine learning database, http://www.ics.uci.edu/~learn/MLRepository.html/ (accessed Nov. 2015).
[33] Y.M. Wang, J.B. Yang, D.L. Xu, Environmental impact assessment using the evidential reasoning approach, Eur. J. Oper. Res. 174 (3) (2006) 1885–1913.
[34] C.F. Wu, C.J. L, C.Y. Lee, A functional neural fuzzy network for classification applications, Expert Syst. Appl. 38 (2011) 6202–6208.
[35] R. Yager, Generalized probabilities of fuzzy events from fuzzy belief structures, Inf. Sci. 28 (1982) 45–62.
[36] J. Yen, Computing generalized belief functions for continuous fuzzy sets, Int. J. Approx. Reason. 6 (1) (1992) 1–31.
[37] L.H. Yang, Y.G. Fu, Y.J. Wu, Structure learning approach of belief rule base for best decision structure, J. Front. Comput. Sci. Technol. 10 (8) (2014) 1216–1230.
[38] J.B. Yang, J. Liu, J. Wang, H.S. Sii, H.W. Wang, Belief rule-base inference methodology using the evidential reasoning approach-RIMER, IEEE Trans. Syst. Man Cybernet. Part A-Syst. Humans 36 (2006) 266–285.
[39] J.B. Yang, D.L. Xu., Evidential reasoning rule for evidence combination, Artif. Intell. 205 (2013) 1–29.
[40] L.A. Zadeh, Fuzzy sets, Inf. Control 8 (1965) 338–353.
[41] L.A. Zadeh, Fuzzy sets and information granularity, in: Proceedings of Advances in Fuzzy Set Theory and Applications, 1979, pp. 3–18.
[42] L.A. Zadeh, Fuzzy sets as a basis for a theory of possibility, Fuzzy Sets Syst. 1 (1978) 3–28.
[43] L.A. Zadeh, A note on Z-numbers, Inf. Sci. 181 (14) (2011) 2923–2932.
[44] G.D. Zhao, Y. Wu, F.Q. Chen, J.M. Zhang, J. Bai, Effective feature selection using feature vector graph for classification, Neurocomputing 151 (2015) 376–389.
[45] Z.J. Zhou, C.H. Hu, J.B. Yang, D.L. Xu, D.H. Zhou, New model for system behavior prediction based on belief rule based systems, Inf. Sci. 180 (24) (2010) 4834–4864.
[46] Y. Zhou, J. Jiang, Z.Y. Yang, Y.J. Tan, A hybrid approach for multi-weapon production planning with large-dimensional multi-objective in defense manufacturing, Proc. Inst. Mech. Eng., Part B: J. Eng. Manuf. 228 (2) (2014) 302–316.
[47] Z.J. Zhou, J.B. Yang, D.L. Xu, D.H. Zhou, Online updating belief-rule-base using the RIMER approach, IEEE Trans. Syst. Man Cybernet.-Part A: Syst. Humans 41 (6) (2011) 1225–1243.
[48] N. Landwehr, M. Hall, E. Frank, Logistic model trees, in: Proceedings of the Joint Conference on Artificial Intelligence, 2003, pp. 241–252.
[49] S.K. Pal, R.K. De, J. Bsask, Unsupervised feather selection: a neuro-fuzzy approach, IEEE Trans. Neural Netw. 11 (2) (2000) 366–376.
[50] M. Fallahnezhad, M.H. Moradi, S. Zaferanlouei, A hybrid high order neural classifier for handling classification problems, Expert Syst. Appl. 38 (2011) 386–393.