Expert Systems With Applications 146 (2020) 113161

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Ensemble belief rule base modeling with diverse attribute selection and cautious conjunctive rule for classification problems

Long-Hao Yang[a], Fei-Fei Ye[a], Ying-Ming Wang[a,b,*]

[a] Decision Sciences Institute, Fuzhou University, Fuzhou, Fujian 350116, China
[b] Key Laboratory of Spatial Data Mining & Information Sharing of Ministry of Education, Fuzhou University, Fuzhou, Fujian 350116, China

ARTICLE INFO

Article history:
Received 24 June 2019
Revised 18 December 2019
Accepted 21 December 2019
Available online 24 December 2019

Keywords:
Belief rule base
Attribute selection
Classification
Ensemble learning
Cautious conjunctive rule

ABSTRACT

Belief rule-based systems have demonstrated its advantages in solving complicated problems with uncertain information. However, the rule combinatorial explosion problem is still a great challenge for belief rule bases (BRBs) when a problem involves a large number of attributes, because existing attempts have not addressed this challenge adequately, *e.g.*, utilization of single attribute selection method to downsize BRBs without considering its inherent weakness, or adjustment of referential values to optimize BRBs without attribute selection. Thus, inspired by ensemble learning, the objective of this paper is to propose an ensemble BRB modeling method to deal with classification problems. First, six attribute selection methods that have different advantages are introduced to select diverse sets of antecedent attributes for constructing multiple BRBs, and all of these BRBs are further trained by parameter learning for diverse belief rule-based systems. Second, due to the fact that each belief rule-based system has different importance and hardly satisfies the assumption of independence, a weight learning method is proposed to determine the weight of each belief rule-based system, and a new analytical cautious conjunctive rule (CCR) is deduced from the recursive CCR, that is suitable for the combination of non- independent individuals, to combine the outputs of all belief rule-based systems. Eight classification datasets from the well- known UCI database are adopted to verify the effectiveness of the proposed BRB modeling method in comparison with the belief rule-based systems constructed by single attribute selection, conventional fuzzy rule-based classifiers, and machine learning-based classifiers.

1. Introduction

Among various rule-based systems for addressing complicated problems, belief rule-based systems (Yang, Liu, Wang, Sii & Wang, 2006) have shown its powerful capabilities to represent uncertain information, *e.g.*, fuzzy terms used to describe each belief rule, and incomplete and probabilistic uncertainties coexisting in belief distributions. Moreover, evidential reasoning (ER) algorithm (Wang, Yang & Xu, 2006; Yang & Xu, 2002) is regarded as the inference engine of belief rule-based systems to integrate uncertain information. Hence, belief rule-based systems have been successfully used in many fields, such as bridge risk assessment (Yang, Wang, Liu & Martínez, 2018), pipeline leak detection (Xu et al., 2007), consumer preference prediction (Yang, Wang, Xu, Chin & Chatton, 2012), classification problems (Zhou et al., 2015), and system fault diagnosis (Zhou et al., 2018).

As the knowledge base of belief rule-based systems, belief rule bases (BRBs) contain all possible belief rules so that it has complete rule representation for any given problems. But there is also a flipside that shows users the challenge of rule combinatorial explosion (Chang, Zhou, Jiang, Li & Zhang, 2013). For example, while a given problem has M attributes with J_i referential value to describe the ith attribute, the total number of belief rules would be $\prod_{i=1}^{M} J_i$ that is associated with the combination of all referential values of each attribute. In this case, BRB modeling must be improved before utilizing belief rule-based systems to deal with complicated problems.

In the past decade, many attempts have been made to address the challenge of rule combinatorial explosion, such as attribute selection, referential value adjustment, BRB modeling under disjunctive assumption, and the extension of BRB modeling, where the details of these attempts can be found in Section 2.2. However, the rule combinatorial explosion is still a great challenge for BRB modeling because (1): the existing studies of attribute selection

* Corresponding author at: Decision Sciences Institute, Fuzhou University, Fuzhou, Fujian 350116, China.
E-mail address: msymwang@hotmail.com (Y.-M. Wang).

https://doi.org/10.1016/j.eswa.2019.113161
0957-4174/

were just based on single method without considering its inherent weakness; (2) the existing studies of referential value adjustment neglected the necessity of attribute selection; (3) the existing studies of disjunctive assumption and extension are actually out of the scope of the conventional BRB modeling. Therefore, it is necessary to develop a more effective method of attribute selection to further improve the conventional BRB modeling while facing the challenge of rule combinatorial explosion.

Ensemble learning is one of prolific disciplines in machine learning owing to the assumption that the combination of the output of multiple independent models is better than the output of any single model. On the basis of the same principle, ensemble learning has been applied to attribute selection based on the homogeneous approach that uses the same attribute selection method with different training data and the heterogeneous approach that uses different attribute selection methods with the same training data (Bolón-Canedo et al., 2019). From the previous studies of ensemble learning on attribute selection (Seijo-Pardo, Porto-Díaz, Bolón-Canedo & Alonso-Betanzos, 2017), both approaches have been widely used for classification problems to select relevant attributes and discard irrelevant or redundant attributes. Hence, ensemble learning can be a solution for the aforementioned challenge of BRB modeling.

According to the above-mentioned idea of ensemble learning with attribute selection, a new ensemble BRB modeling method with diverse attribute selection and cautious conjunctive rule (CCR) (Denoeux, 2008) is proposed for classification problems to construct an ensemble belief rule-based system. The main components of the proposed BRB modeling method can be summarized as the following two aspects.

First, through reviewing a large number of existing attribute selection methods with the standard of diversity that means the greatest possible change to select attributes based on their inherent strengths, six different methods on the basis of information gain, gain ratio, chi-squared, ReliefF, OneR classifier, and support vector machine are therefore applied to select diverse sets of antecedent attributes for constructing multiple BRBs with sufficient diversity. However, the parameter values of any BRB constructed by the selected set of attributes are still unavailable to guarantee the desired classification performance of a belief rule-based system. Hence, according to the existing global parameter learning model for regression problems (Chen, Yang, Xu, Zhou & Tang, 2011), a new parameter learning model aimed to classification problems is developed to optimize those BRBs independently for improving the classification performance of all belief rule-based systems.

Second, considering that all the belief rule-based systems are not independent of each other since some of them not only have the same antecedent attributes, but also are trained by using the same training data, leading to the dilemma that many output combination methods widely used in ensemble learning lost the ability to combine the outputs of all the belief rule-based systems because these methods have to work under the assumption of independence. For the non-independence of those belief rule-based systems, CCR is introduced to propose a new output combination method for the combination of the outputs of all the belief rule-based systems. Meanwhile, it is undoubted that each belief rule-based system usually has different importance in an ensemble belief rule-based system. Facing this situation, a weight learning method is developed to determine the weight of each belief rule-based system in the ensemble belief rule-based system.

The CCR proposed by Denoeux is a extension of Dempster's rule and it opens the doors to new ways of combining the evidences that may not independent. However, the current CCR is based on a recursive process to combine evidences. In some situations such as in optimization problems, an explicit combination process is more desirable. Hence, an analytical CCR is investigated for the first time to combine evidences for classification problems.

To verify the effectiveness of the proposed BRB modeling method, the process of constructing and optimizing BRB is used to illustrate the model development of an ensemble belief rule-based system in modeling classification problems. Moreover, two kinds of classifiers based on conventional rule-based systems and machine learning algorithm are applied to compare the performance of the ensemble belief rule-based system.

The remainder of this paper is organized as follows: Section 2 briefly reviews the background and challenges of the conventional BRB modeling used in complex problems. Section 3 introduces new ensemble BRB modeling method with diverse attribute selection and CCR for classification problems. Section 4 provides an experiment study to demonstrate the effectiveness of the proposed BRB modeling method, and finally Section 5 concludes the study.

2. Background of BRB modeling and its challenges for complex problems

In this section, the conventional BRB modeling is introduced to provide the basic knowledge of this study, followed by the related works on BRB modeling to show past developments. Finally, several challenges of BRB modeling for complex problems are therefore summarized to illustrate the purpose of this study.

2.1. Conventional BRB modeling

The BRB modeling is a crucial bridge that links complex problems and belief rule-based systems. Typically, a BRB can be assumed to have M antecedent attributes $\{U_1, U_2,\ldots, U_M\}$ and one consequent attribute D. Each antecedent attribute U_i ($i = 1,\ldots, M$) is described by using J_i referential values $\{A_{i,j}; j = 1,\ldots, J_i\}$ and the consequent attribute D is described by using N consequents $\{D_n, n = 1,\ldots, N\}$. As a result, the belief rule in the BRB is defined as the following representation (Yang et al., 2006):

$$\begin{aligned} R_l: & \ IF\ U_1\ is\ A_1^l\ \wedge U_2\ is\ A_2^l\ \wedge \cdots \wedge\ U_M\ is\ A_M^l, \\ & THEN\ D\ is\ \left\{(D_n, \beta_{n,l}); n = 1, \ldots, N\right\}, \\ & with\ rule\ weight\ \theta_l\ and\ attribute\ \ weights\ \{\delta_1, \ldots, \delta_M\} \end{aligned} \quad (1)$$

where A_i^l denotes the referential value of the ith ($i = 1,\ldots, M$) antecedent attribute in the lth ($l = 1,\ldots, L$) belief rule, namely, $A_i^l \in \{A_{i,j}; j = 1,\ldots, J_i\}$; L is the total number of belief rules in the BRB; and $\beta_{n,l}$ denotes the belief degree to which the consequent D_n is believed to be true. Fig. 1 gives the graphical representation of the lth belief rule.

From Fig. 1, each antecedent attribute is represented as a point and its referential values are represented as multiple adjacent

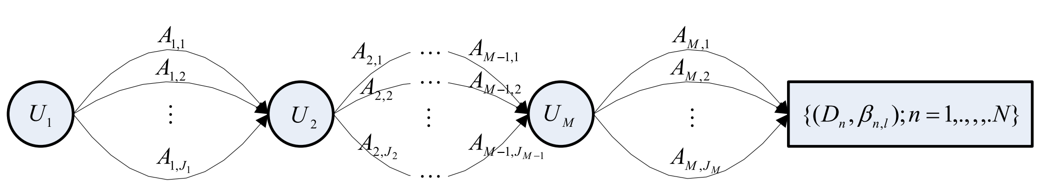

Fig. 1. Graphical representation of the lth belief rule.

edges. Hence, the construction of a complete BRB can be understood as the combination of each edge from U_1 to D and each possible path is therefore represented as one kind of belief rules. As a result, a total of $\prod_{i=1}^{M} J_i$ belief rules would be constructed for the complete BRB while each antecedent attribute U_i (i = 1,..., M) includes J_i referential values.

After constructing BRB, parameter learning should be applied to optimize the parameter values of all belief rules (Yang et al., 2007). The core of parameter learning is that all parameter values are updated iteratively for obtaining their optimal values on the basis of collected data, so that the corresponding belief rule-based system can produce a desired output for each input data by using ER-based inference method, which is the inference engine of belief rule-based systems and its detailed steps are as follows:

Step 1: Calculation of activation weights for each belief rule.

Suppose there is an input data $\boldsymbol{x} = (x_1, x_2, \ldots, x_M)$, and x_i represents the input of the ith antecedent attribute. First, the individual matching degree is transformed using utility-based equivalence transformation techniques (Yang, 2001).

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

where

$$\alpha_{i,j} = \frac{u\left(A_{i,j+1}\right) - x_i}{u\left(A_{i,j+1}\right) - u\left(A_{i,j}\right)} \text{ and } \alpha_{i,j+1} = 1 - \alpha_{i,j}, \text{ if } u\left(A_{i,j}\right) \le x_i \le u\left(A_{i,j+1}\right), \tag{3}$$

$$\alpha_{i,k} = 0, \text{ for } k = 1, \ldots, J_i \text{ and } k \ne j, j+1, \tag{4}$$

where $u(A_{i,j})$ denotes the utility value of referential value $A_{i,j}$; $\alpha_{i,j}$denotes the individual matching degree of the given input x_i to referential value $A_{i,j}$.

Next, the activation weight for the lth rule is calculated as follows:

$$w_l = \frac{\theta_l \prod_{i=1}^{M} \left(\alpha_i^l\right)^{\bar{\delta}_i}}{\sum_{k=1}^{L} \left(\theta_k \prod_{i=1}^{M} \left(\alpha_i^k\right)^{\bar{\delta}_i} \right)}, \bar{\delta}_i = \frac{\delta_i}{\max_{k=1,\ldots,M}\{\delta_k\}}, \tag{5}$$

where θ_l is the rule weight of the lth rule, δ_i is the attribute weight of the ith antecedent attribute.

Step 2: Integration of activated belief rules using ER algorithm.

Suppose there are L belief rules and their activation weights are greater than 0. Thus, using the analytical ER algorithm (Wang et al., 2006), the combined belief degree β_i can be obtained by integrating these L belief rules as follows:

$$\beta_i = \frac{\prod_{l=1}^{L} \left(w_l \beta_{i,l} + 1 - w_l \sum_{n=1}^{N} \beta_{n,l} \right) - \prod_{l=1}^{L} \left(1 - w_l \sum_{n=1}^{N} \beta_{n,l} \right)}{\sum_{n=1}^{N} \prod_{l=1}^{L} \left(w_l \beta_{n,l} + 1 - w_l \sum_{j=1}^{N} \beta_{j,l} \right) - (N-1) \prod_{l=1}^{L} \left(1 - w_l \sum_{j=1}^{N} \beta_{j,l} \right) - \prod_{l=1}^{L} (1 - w_l)}. \tag{6}$$

While the D_n denotes the nth class, the inferential output for the input data $\boldsymbol{x}$ is as follows:

$$f(\mathbf{x}) = D_t, t = \underset{n=1,\ldots,N}{\arg\max}(\beta_n). \tag{7}$$

2.2. Related works on BRB modeling

In order to improve the conventional BRB modeling, many endeavors have been undertaken in the past decade. The most popular endeavors are about parameter learning, and lots of models and algorithms have been developed to obtain the optimal parameter values of BRBs. In this respect, the recent literatures (Tang, Xiao, Liang, Zhu & Li, 2019; Wang, Yang, Fu, Chang & Chin, 2016; Yang et al., 2018) have provided a detailed overview. As another important process of BRB modeling, the construction of BRBs is lack of sufficient attentions, because BRBs are usually under some assumptions that antecedent attributes and their referential values are provided using expert knowledge. Hence, this section aims at reviewing the previous studies of constructing BRBs and these studies can be categorized into the following four types.

(1) Attribute selection for BRB modeling, which selects the most representative attributes that are regarded as antecedent attributes for constructing BRBs. Wang, Yang, Xu and Chin (2009) aimed to avoid building an over-large BRB for consumer preference prediction, thus the first two or three principal components were regarded as antecedent attributes to construct a downsized BRB. A similar research was done in Yang et al. (2012). Chang et al. (2013) introduced four kinds of attribute reduction techniques, namely gray target, multidimensional scaling, Isomap, and principle component analysis, to reduce the number of antecedent attributes and further compared the performance of these techniques on BRB modeling. (Yang et al., 2016b) extracted a small number of antecedent attributes from product attributes by conducting both exploratory and confirmatory factor analysis. The results showed that the size of BRB is reduced from 152 to 25 belief rules.
(2) Referential value adjustment for BRB modeling, which selects the most representative referential values for each antecedent attribute. Zhou et al. (2010) proposed a statistical utility to perform the addition or pruning of belief rules, accompanied by the adjustment of referential values. Thereafter, the statistical utility was extended for the application of BRBs on delayed coking unit (Yu, Huang, Jiang & Jin, 2012). Wang et al. (2016) utilized density analysis and error analysis to determine whether it needs to add or prune referential values for obtaining a complete BRB. Ke, Ma and Wang (2017) proposed a reduction method for removing possibly redundant referential values, and the obtained BRB was proven to have good performance on Box-Jenkins gas furnace problem. Yang et al. (2018) was based on generalization error to propose a heuristic strategy for selecting suitable referential values without facing the combinatorial explosion problem of BRB modeling. (Chang et al., 2018) proposed an empirical optimization path search strategy based on Akaike Information Criterion to optimize the sets of referential values.
(3) Disjunctive assumption for BRB modeling, which is based on disjunctive assumption, instead of conjunctive assumption, to construct belief rules for a BRB. The first BRB with disjunctive assumption was proposed by Yang et al. (2006). However, no study was carried out until it was extended by proposing new activation weight calculation to address classification problems (Chang, Zhou, You, Yang & Zhou, 2016). As a sign of the new openness on BRB modeling, many scholars applied the BRBs with disjunctive assumption to many fields, such as bridge risk assessment (Yang et al., 2017a), nonlinear complex system modeling (Chang et al., 2019b), and threat level assessment (Chang et al., 2019a).
(4) Derivation or extension of BRB modeling, which is based on new or improved mechanism to construct BRBs. Chen et al. (2015) discussed the structure of Bayesian network modeled by conditional probability table and then proposed a data-driven approximate causal model, which only constructs $\sum_{i=1}^{M} J_i$ belief rules for a BRB. A similar study was proposed in Sun et al. (2018). Jiao, Pan, Denœux, Liang and

Feng (2015) combined BRBs with fuzzy rule-based classification system and then proposed a calculation model to generate belief rules from collected data, so that the size of a BRB is the smaller value of $\prod_{i=1}^{M} J_i$ and the number of collected data. The most popular extension of BRB modeling is the extended BRB (EBRB) that embedded belief structure into all antecedent attributes and it can generate extended belief rules from collected data directly (Liu, Martinez, Calzada & Wang, 2013). Owing to its advantages comparing to BRB modeling, EBRBs have attracted many attentions of BRB researchers (Calzada, Liu, Wang & Kashyap, 2015; Lin, Fu, Su, Wang & Gong, 2017; Yang et al., 2016a, Yang et al. 2017b).

To summarize, the above types of studies have enhanced the process of constructing belief rules in BRB modeling to different degrees. Comparatively, the former two types were performed within the scope of conventional BRB modeling, and the latter two types were aiming to develop new BRB modeling.

2.3. Challenges of BRB modeling for complex problems

In this section, some challenges of BRB modeling are summarized while addressing complex problems according to the above-mentioned conventional BRB modeling and its related works. First, the rule combinatorial explosion challenge that has been commonly discussed in many previous studies should be provided as follows:

Challenge 1 (***Rule combinatorial explosion in BRB modeling***): When there is a large number of antecedent attributes or referential values for each antecedent attribute, the size of a complete BRB can be very large because the construction of complete belief rules is required to cover all combinations of each referential value for each antecedent attribute.

To address Challenge 1, four types of studies shown in Section 2.2 have been carried out in the past decade and all of them can achieve the improvement of BRB modeling. Comparing with the first and second types of attempts, the third and forth ones were trying to propose a new mechanism to construct belief rules so that these are actually out of the scope of the conventional BRB modeling. For example, previous studies mainly focused on the complex problems which consist of attributes with conjunctive logical relationship. Obviously, the BRB modeling under disjunctive assumption requires attributes to have disjunctive logical relationship. Hence, based on the first two types of previous studies, the following new challenges for complex problems can be identified and prompt the further development of BRB modeling.

Challenge 2 (***"No free lunch" of antecedent attribute selection***): Although previous studies have shown the importance of attribute selection and many methods were applied to select the most representative antecedent attributes for the construction of belief rules, none of them paid attention to the usability of an attribute selection method that actually has its strengths and weaknesses, and its performance usually depends on the type of dataset.

Challenge 3 (***Hercules task of referential value adjustment***): The referential value adjustment is another solution to deal with Challenge 1. However, it is a much more difficult task to obtain a desired set of referential values for complex problems because the size of a complete BRB $\prod_{i=1}^{M} J_i$ would still be very large under the situation of a large number of antecedent attributes M.

The above-mentioned three challenges clearly discuss the drawbacks of existing BRB modeling and also clarify the necessity of a smarter attribute selection method to improve the construction of belief rules. Therefore, in the next section, an ensemble method is proposed to overcome the three challenges to deal with complex classification problems.

3. Ensemble BRB modeling based on diverse attribute selection and CCR for classification problems

According to the challenges pointed out in Section 2.3, this section aims to propose an ensemble BRB modeling method based on diverse attribute selection and CCR. Firstly, the framework of proposed modeling method is introduced, followed by attribute selection and parameter learning for constructing a number of belief rule-based systems, and output combination and weight learning for aggregating these systems.

3.1. Framework of diverse attribute selection and CCR based ensemble BRB modeling

In this section, the framework of the ensemble BRB modeling method based on diverse attribute selection and CCR is introduced to illustrate how to construct an ensemble belief rule-based system. As shown in Fig. 2, the proposed modeling method mainly includes the following two processes.

First, different attribute selection methods are applied to collaboratively select diverse sets of representative antecedent attributes from different perspectives. The features of these diverse sets of attributes are that neither of them has the same attributes as another set of attributes and the union of all sets is equal to the entire set of attributes. In other words, any attribute is selected at least once by some attribute selection methods. Based on the selected diverse sets, a number of BRBs can be constructed and their parameter values are optimized by parameter learning.

Second, due to the fact that each BRB together with ER-based inference method (called belief rule-based system) can produce an inferential output for each same input data, it is necessary to achieve a combination of the inferential output of all belief rule-based systems, which usually have different importance and are not independent of each other. Hence, a weight learning method is applied to determine the importance of each system and the CCR which do not require the independence of individuals is adopted to combine the inferential output of all belief rule-based systems.

To further illustrate the framework of the proposed BRB modeling method, its pseudo-code is provided as follows:

Algorithm: Pseudo-code of ensemble BRB modeling method based on diverse attribute selection and CCR

Data: S denotes number of attribute selection methods used to obtain attribute rankings
M denotes number of attributes selected for constructing a BRB.
K denotes number of BRBs or belief rule-based systems in an ensemble belief rule-based system

```
Obtain S rankings of all attributes by using S attribute selection
methods.
While not all attributes have been selected at least once do
    For each s from 1 to S do
        Select top M attributes as a set of antecedent attributes
        based on the sth attribute rankings.
        If the set of antecedent attributes has not yet been reused do
            Construct a BRB by combining each referential value of the M
            selected attributes.
            Obtain optimal parameter values of the BRB by parameter
            learning.
        End if
        Delete the M selected attributes from the sth attribute rankings.
    End for
End while
Obtain K optimal weights for each belief rule-based system using
weight learning method.
Aggregate K inferential belief distributions and optimal weights using
CCR-based combination method.
```

For the above-mentioned framework for the ensemble BRB modeling method, the following remark can be given:

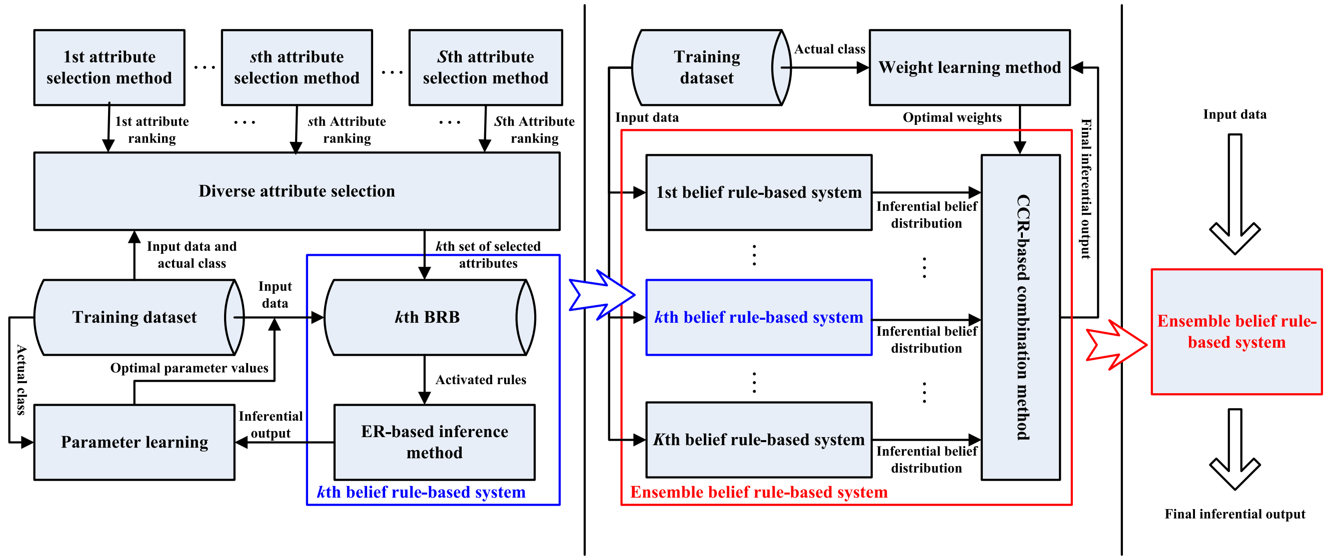

Fig. 2. Framework of the proposed ensemble BRB modeling method.

Table 1
Introduction of six different attribute selection methods (Alcala-Fdez et al., 2011).

Core of attribute selection	Abbr.	Description
Information gain	AS-IG	To evaluate the worth of each attribute by measuring its information gain with respect to each class
Gain ratio	AS-GR	To evaluate the worth of each attribute by measuring its gain ratio with respect to each class
Chi-squared	AS-CS	To evaluate the worth of each attribute by computing the value of chi-squared statistic with respect to each class
ReliefF	AS-RF	To evaluate the worth of each attribute by repeatedly sampling a data and considering the value of a given attribute for the nearest data of the same and different classes
OneR classifier	AS-ORC	The current set of attributes is applied to train an oneR classifier iteratively by removing attributes, so that the performance can evaluate the worth of each attribute
Support vector machine	AS-SVM	This method is similar to AS-ORC and the difference is that the application of classifier on attribute selection is based on support vector machine

(1) There are some differences between the ensemble BRB modeling method and some well-known ensemble learning algorithms, *e.g.*, random forest (Breiman, 2001) which utilizes random way to perform attribute selection, but the ensemble BRB modeling method is based on the importance among different attributes.
(2) In the proposed framework, it is necessary to guarantee a high diversity among attribute selection methods, because the attribute selection methods with a small diversity usually fail to select diverse sets of attributes owing to the fact that their attribute rankings are the same to each other.

3.2. Attribute selection and parameter learning for diverse BRBs

According to the framework shown in Section 3.1, the first process of the proposed BRB modeling method is based on attribute selection and parameter learning to construct a number of BRBs with sufficient diversity, respectively, and their details are provided in the coming subsections.

3.2.1. Diverse attribute selection to construct BRB

One of the main challenges for the existing BRB modeling is selecting an appropriate attribute selection method for determining the most representative antecedent attributes, because any kind of attribute selection method inevitably has its strengths and weaknesses. Hence, a smart strategy is to select attributes based on various kinds of methods so that diverse sets of representative antecedent attributes can be selected from various perspectives. From this point of view, six different attribute selection methods, as shown in Table 1, are adopted to construct diverse BRBs.

To show the different performances among the six attribute selection methods, a small diversity study using Glass classification dataset is conducted to measure Spearman correlation coefficient of the attribute rankings obtained by using these methods, in which each correlation coefficient is ranged from $[-1, 1]$, and 1 denotes that the obtained rankings are equal absolutely. The results are shown in Table 2.

From Table 2, it can be seen that most of correlation coefficients are far from 1, which indicates that there are significant differences between the paired attribute selection methods. Hence, this experiment demonstrates that the attribute selection methods chosen for constructing BRBs can ensure sufficient diversity.

Thereafter, on the basis of the selected attributes, experts should be invited to determine referential values for each antecedent attribute and finally multiple BRBs with different attributes can be obtained. Note that these BRBs are less likely to suffer from the challenge of rule combinatorial explosion because only the selected attributes are used to construct BRBs. Undoubtedly, when the number of selected attributes is still a large value, the size of each BRB will also be very large. Thus, it is important to determine a reasonable number of selected attributes according to prior knowledge, *e.g.* number of original attributes, size of collected data, or domain experts' knowledge.

3.2.2. Parameter learning to optimize BRB

The BRBs derived from Section 3.2.1 may not produce inferential outputs with desired accuracy if their initial parameter values are subjectively set by experts. This is because expert knowledge is inherently incomplete and inaccurate. To address this situation, a training dataset should be used to obtain optimal parameter val-

Table 2
Spearman correlation coefficient of the attribute rankings in Glass dataset.

Methods	AS-IG	AS-GR	AS-CS	AS-RF	AS-ORC	AS-SVM
AS-IG	1.0000	0.3167	0.7667	0.3667	0.0500	0.1333
AS-GR	0.3167	1.0000	0.4333	−0.3667	−0.3500	0.4333
AS-CS	0.7667	0.4333	1.0000	0.2333	−0.0833	−0.2833
AS-RF	0.3667	−0.3667	0.2333	1.0000	0.7833	−0.1000
AS-ORC	0.0500	−0.3500	−0.0833	0.7833	1.0000	0.0333
AS-SVM	0.1333	0.4333	−0.2833	−0.1000	0.0333	1.0000

ues for each belief rule in the BRBs. According to the existing study (Chen et al., 2011) which proposed a global parameter learning model for regression problems, a revised parameter learning model for classification problems is developed to optimize the parameter values of the BRBs. Note that each BRB is optimized through parameter learning independently.

Taking the kth (k = 1,..., K) BRB as an example, assume this BRB is composed of J_i utility values $\{u(A_{i,j})^k, j = 1,..., J_i\}$ for each antecedent attribute U_i (i = 1,..., M), L^k rule weights $\{\theta_l^k, l = 1,..., L^k\}$, M attribute weights $\{\delta_i^k, i = 1,..., M\}$, and N belief degrees $\{\beta_{n,l}^k, n = 1,..., N\}$ for each belief rule R_l (l = 1,..., L^k). Additionally, when $f_k(\boldsymbol{x_t})$ is assumed to be the inferential output of the kth belief rule-based system for the given input data $\boldsymbol{x_t}$ and y_t to be the actual class of $\boldsymbol{x_t}$, the criterion E_t to evaluate the classification correctness of the kth belief rule-based system can be denoted as follows:

$$E_t = \begin{cases} 1, & if\ f_k(x_t) \neq y_t \\ 0, & otherwise \end{cases}. \tag{8}$$

According to the above-mentioned assumption, the details of the revised parameter learning model are as follows:

$$Minimize\ \ \xi(u(A_{i,j})^k, \theta_l^k, \delta_i^k, \beta_{n,l}^k) = \sum_{t=1}^{T} E_t, \tag{9a}$$

$$Subject\ to \sum_{n=1}^{N} \beta_{n,l}^k = 1,\ l = 1, ..., L^k, \tag{9b}$$

$$0 \leq \beta_{n,l}^k \leq 1,\ \ n = 1, ..., N;\ l = 1, ..., L^k, \tag{9c}$$

$$0 \leq \theta_l^k \leq 1,\ l = 1, ..., L^k, \tag{9d}$$

$$0 \leq \delta_i^k \leq 1,\ i = 1, ..., M, \tag{9e}$$

$$u(A_{i,j})^k - u(A_{i,j+1})^k \leq V_i^k,\ i = 1, ..., M;\ j = 1, ..., J_i - 1, \tag{9f}$$

where Eq. (9a) is a target function used to minimize classification error; Eqs. (9b) and (9c) are constraints on $\beta_{n,l}^k$ in the kth BRB; Eqs. (9d) and (9e) are constraints on δ_i^k and θ_l^k in the kth BRB, respectively; Eq. (9f) is a constraint on $u(A_{i,j})^k$ in the kth BRB; V_i^kdenotes a small value to differentiate between two adjacent referential values for the ith antecedent attribute in the kth BRB; T is the number of training data to optimize the kth BRB.

To solve the optimization model shown in Eqs. (9a)–(9f), many techniques, such as MATLAB, differential evolution (DE), and genetic algorithm (GA)-based optimization techniques, have been proposed in previous studies. All of them can optimize the parameters of a BRB to different degrees and finally obtain a high-quality solution for the parameter learning of the BRB. More details of these techniques can be found in Tang et al., (2019).

3.3. Output combination and weight learning for ensemble belief rule-based systems

After constructing and optimizing BRBs individually, a CCR-based combination method is proposed to achieve the combination of the inferential outputs of each belief rule-based system. Moreover, a weight learning method is proposed to determine the importance of each belief rule-based system.

3.3.1. CCR-based combination method to aggregate outputs of belief rule-based systems

To make full use of diverse belief rule-based systems to deal with classification problems, it is necessary to provide an effective output combination method for these belief rule-based systems. Ensemble learning (Zhou et al., 2012) indicates that the most widely used output combination methods include averaging, voting, and combining by learning and almost all of them have to work under the assumption of independence. In other words, all classifiers in an ensemble classifier must be independent of each other.

However, it is not easy to guarantee the independence of all the belief rule-based systems in an ensemble belief rule- based system since all of them potentially have same antecedent attributes and they are based on the same training data to perform parameter learning for optimizing BRBs. In such a situation, CCR (Denoeux, 2008) is introduced to propose an output combination method for combining those belief rule-based systems. Note that CCR is one of extensions of Dempster 's rule but it does not require the combination of different evidences under the assumption of independence. Traditionally, CCR is based on a recursive process to combine various piece of evidence individually. The advantage of doing so is its clarity in concept. In situations where an explicit combination process is required such as in optimization, an analytical CCR would be desirable for the combination of evidences. In view of this, the analytical CCR is developed in Appendix A.

Additionally, all the belief rule-based systems that are constructed by diverse attribute selection usually have different importance on final inferential outputs. Thus, the importance of each belief rule-based system together with the analytical CCR is used to propose a CCR-based combination method to aggregate the inferential outputs of all the belief rule-based systems. The detailed steps of the proposed method are introduced as follows.

Step 1: To generate the inferential outputs of all belief rule-based systems. Suppose there are K belief rule-based systems and their inferential outputs are denoted as $f_k(\boldsymbol{x})$=$\{(D_n,\beta_n^k)\}$ (k = 1,..., K) while classifying input data $\boldsymbol{x}$. Hence, the inferential output $g(\boldsymbol{x})$ of all belief rule-based systems can be expressed as follows:

$$g(\boldsymbol{x}) = \begin{array}{c} \\ f_1(\boldsymbol{x}) \\ f_2(\boldsymbol{x}) \\ \vdots \\ f_K(\boldsymbol{x}) \end{array} \!\!\begin{array}{c} \begin{array}{cccc} D_1 & D_2 & \cdots & D_N \end{array} \\ \begin{bmatrix} \beta_1^1 & \beta_2^1 & \cdots & \beta_N^1 \\ \beta_1^2 & \beta_2^2 & \cdots & \beta_N^2 \\ \vdots & \vdots & \cdots & \vdots \\ \beta_1^K & \beta_2^K & \cdots & \beta_N^K \end{bmatrix} \end{array}. \tag{10}$$

Step 2: To aggregate the inferential outputs of all belief rule-based systems. For the purpose of considering importance and non-independency of each belief rule-based system, the weight ω_k is used to represent the importance of the kth belief rule-based systems and the analytical CCR are applied to aggregate the inferential outputs of all belief rule-based systems as follows:

$$m(D_n) = \prod_{i=1, i\neq n}^{N} \varpi(D_i) - \prod_{i=1}^{N} \varpi(D_i), \tag{11}$$

where $m(D_n)$ denotes the integrated basic belief assignment on the nth (n = 1,..., N) class, $\varpi(D_n)$denotes the combination of conjunctive weight weights for the nth class and its calculation formula is as follows:

$$\varpi(D_n) = \min_{k=1,\ldots,K} \left\{ \frac{1-\omega_k \sum_{n=1}^{N} \beta_n^k}{\beta_n^k \omega_k + 1 - \omega_k \sum_{n=1}^{N} \beta_n^k} \right\}. \tag{12}$$

Step 3: To produce final inferential outputs. On the basis of Eq. (11), a final inferential output $G(\boldsymbol{x})$ can be produced according to the biggest value from N integrated basic belief assignments as follows:

$$G(\boldsymbol{x}) = D_t, t = \arg\max_{n=1,\ldots,N}\{m(D_n)\}. \tag{13}$$

3.3.2. Weight learning method to determine importance of belief rule-based systems

In Section 3.3.1, K weights $\{\omega_1,..., \omega_K\}$ are used to describe different importance of each belief rule-based system while dealing with classification problems. Commonly, it is difficult to accurately determine the optimal values of these K weights by expert knowledge because a slight change of weights may have significant influences on an ensemble belief rule -based system. To address this situation, a weight learning method based on DE algorithm is proposed to determine the optimal value of these K weights.

DE algorithm was proposed by Storn and Price (1997) and has demonstrated its excellent performance in dealing with optimization problems in past decades. The core of DE algorithm is to maintain a group of candidate solutions and creating new candidate solutions by combining existing ones according to evolution strategy. On the basis of DE algorithm, the detailed steps of the weight learning method are introduced as follows:

Step 1: To initialize a group of candidate solutions. Assume that there are C candidate sets of K weights and the cth candidate set is expressed as follows:

$$\omega_{\mathbf{c}} = \{\omega_{c,k}; k = 1, ..., K\}; c = 1, ..., C, \tag{14}$$

where the initial value of $C \times K$ weights is determined by using a random value between 0 and 1, namely,

$$\omega_{c,k} = random(0,\ 1); c = 1, ..., C; k = 1, ..., K. \tag{15}$$

Step 2: To evolve all candidate solutions. For each candidate set $\boldsymbol{\omega}_c$ in the C candidate sets, a new candidate set $\omega_{\mathbf{c}}^0$ is generated by using the three different candidate sets randomly selected from the C candidate sets, in which these three sets are signed as $\omega_{\mathbf{c}}^1$, $\omega_{\mathbf{c}}^2$, and $\omega_{\mathbf{c}}^3$, respectively. The weight values in the candidate set $\omega_{\mathbf{c}}^0$ are assigned according to the classical evolution strategy (Storn et al., 1997) as follows:

$$\omega_{c,k}^0 = \begin{cases} \omega_{c,k}, & if\ random(0,1) > CR \\ \omega_{c,k}^1 + F \times (\omega_{c,k}^2 - \omega_{c,k}^3), & otherwise \end{cases}, c = 1, ..., C; k = 1, ..., K, \tag{16}$$

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

Step 3: To update all candidate solutions. When any weight value in the new candidate set $\boldsymbol{\omega}_{\mathbf{c}}^0$ is out of the range [0, 1], a new weight value should be produced using Eq. (15). Then, according to the CCR-based combination method, the classification error regarding the candidate set $\boldsymbol{\omega}_{\mathbf{c}}^0$, denoted as $\xi(\boldsymbol{\omega}_{\mathbf{c}}^0)$, can be calculated by using all training data, and the weight values of the candidate set $\boldsymbol{\omega}_c$ should be updated as follows:

$$\boldsymbol{\omega}_{\mathbf{c}} = \begin{cases} \boldsymbol{\omega}_c^0; & if\ \xi(\boldsymbol{\omega}_{\mathbf{c}}^0) < \xi(\boldsymbol{\omega}_{\mathbf{c}}) \\ \boldsymbol{\omega}_{\mathbf{c}}; & otherwise \end{cases}, c = 1, ..., C, \tag{17}$$

where

$$\xi(\boldsymbol{\omega}_{\mathbf{c}}) = \sum_{t=1}^{T} E_t, E_t = \begin{cases} 1, & if\ G(\boldsymbol{x}_t) \neq y_t \\ 0, & otherwise \end{cases}, \tag{18}$$

where $G(\boldsymbol{x}_t)$ denotes the final inferential output of an ensemble belief rule-based system for the tth (t = 1,..., T) input data $\boldsymbol{x}_t$, and y_t is the actual class of the tth input data;

Step 4: To select the best candidate solution. When the number of iterations to perform Steps 2 and 3 is equal to S, the candidate set with the minimum classification error is selected as the best one and its weights are regarded as the optimal weights used for the combination of K belief rule-based systems.

In the above-mentioned method, there are some following remarks:

(1) Step 3 guarantees that all candidate sets of weights can get better or remain unchanged, but never deteriorate at each iteration. Hence, the process of weight learning is able to converge to an optimal set of weights after performing S iterations with C candidate sets.
(2) If C and S are large, the classification performance of an ensemble belief rule-based system will be better, but the resulting weight learning is more time-consuming. Therefore, these two parameters should be chose carefully according to the expected system performance.

4. Experimental study

In this section, the performance of the proposed BRB modeling method is empirically verified through experimental study with eight datasets from the well-known UCI database. Moreover, the belief rule-based systems with single attribute selection and conventional classifiers are used to compare with ensemble belief rule-based systems.

4.1. Classification datasets and experimental setting

Eight classification datasets obtained from the UCI database are used to evaluate the performance of the proposed BRB modeling method. The main characteristics of these eight classification datasets are summarized in Table 3.

From Table 3, the eight datasets have different numbers of attributes which are in range from 6 to 30. Conventionally, the size of complete BRBs would be 3^6=729 to 3^{30}=2.0589 × 10^{14} when each antecedent attribute only has three referential values. It is obvious that attribute selection must be performed to downsize those complete BRBs. Hence, the number of attributes to be selected is defined as 3, 4, and 5, respectively, for all those classification datasets.

To develop comparison in multiple aspects, 10-fold cross-validation (10-CV) is considered in the experimental study, where each dataset is divided into 10 blocks with 9 blocks as training data and the remaining one as testing data. Moreover, the non-parametric statistical analysis based on the Friedman and Holland tests is applied to investigate whether significant differences exist among different classifiers at a level of significance of α=0.1. More details of these two tests can be found in García, Fernández, Luengo and Herrera, (2010). Additionally, the number of referential values for each antecedent attribute is set as 3. The DE -based parameter learning, algorithm proposed in (Chang et al., 2015) is used to optimize the parameter values of BRBs, in which the number of generations and individuals are set as 1000 and 100, respectively. For the weight learning method shown in Section 3.3.2, the number of iterations and sets are set as 200 and 20, respectively.

Table 3
Statistics on eight classification datasets (Dua & Graff, 2019).

No.	Dataset	No. of data	No. of attributes	No. of classes	No. of selected attributes
1	Monks	432	6	2	3/4/5
2	Diabetes	393	8	2	3/4/5
3	Pima	768	8	2	3/4/5
4	Glass	214	9	7	3/4/5
5	Wine	178	13	3	3/4/5
6	Heart	270	13	2	3/4/5
7	Cleveland	297	15	5	3/4/5
8	Cancer	569	30	2	3/4/5

Table 4
Attribute rankings obtained by six different attribute selection methods.

Attribute selection method	Attribute rankings								
	No. 1	No. 2	No.3	No. 4	No. 5	No. 6	No. 7	No. 8	No. 9
AS-IG	Al	Mg	K	Ca	Ba	RI	Na	Fe	Si
AS-GR	Ba	Mg	Na	Al	K	Ca	RI	Fe	Si
AS-CS	K	Al	Ca	Ba	Mg	Na	RI	Fe	Si
AS-RF	Mg	Al	Ba	Na	Ca	RI	Si	K	Fe
AS-ORC	Al	K	RI	Ca	Ba	Na	Mg	Si	Fe
AS-SVM	Al	Na	K	Fe	Ba	Mg	Si	RI	Ca

4.2. Model development of ensemble belief rule-based system

In order to illustrate how the proposed BRB modeling method is applied to establish an ensemble belief rule-based system for classification problems, the dataset Glass with setting 3 selected attributes is taken for example, in which Glass includes 9 attributes, namely, refractive index (RI), sodium (Na), magnesium (Mg), aluminum (Al), Silicon (Si), Potassium (K), Calcium (Ca), Barium (Ba), and Iron (Fe). Additionally, the following model development of the ensemble belief rule- based system is based on 90% Glass data as training data and the remaining 10% data as testing data.

In the first step, six different attribute selection methods are independently used to obtain the attribute rankings of 9 attributes. All attribute rankings are shown in Table 4. From Table 4 which also highlights the repeated set of attributes using the same type of underlines, 15 different sets of attributes can be selected as the representative antecedent attributes for constructing diverse BRBs, namely {Al, Mg, K}, {Ba, Mg, Na}, {K, Al, Ba}, {Mg, Al, Ba}, {Al, K, RI}, {Al, Na, K}, {Ca, Ba, RI}, {Na, Ca, RI}, {Ca, Ba, Na}, {Fe, Ba, Mg}, {Na, Fe, Si}, {RI, Fe, Si}, {Si, K, Fe}, {Mg, Si, Fe}, and {RI, Si, Ca}. It is clear from these sets of attributes that any attribute can be used to construct BRBs at least once so that none of input data would be wasted. By covering the traversal combinations of three referential values of three selected attributes, 27 belief rules would be constructed for 15 diverse BRBs. For convenience, the belief rule-based system related with these 15 diverse BRBs is abbreviated as BRB-{Al, Mg, K}, BRB-{Ba, Mg, Na}, BRB-{K, Al, Ba}, BRB-{Mg, Al, Ba}, BRB- {Al, K, RI}, BRB-{Al, Na, K}, BRB-{Ca, Ba, RI}, BRB-{Na, Ca, RI}, BRB-{Ca, Ba, Na}, BRB-{Fe, Ba, Mg}, BRB-{Na, Fe, Si}, BRB-{RI, Fe, Si}, BRB-{Si, K, Fe}, BRB-{Mg, Si, Fe}, and BRB-{RI, Si, Ca}, respectively, and the combination of these 15 belief rule-based systems is abbreviated as Ensemble-BRB.

In the second step, the parameter learning model shown in Eqs. (9a)–(9f) is applied to optimize the parameter values of all belief rules in the 15 BRBs independently. Fig. 3 shows the change of classification error of the 15 belief rule-based systems for training data within 1000 generations. It is clear from Fig. 3 that the classification error of all belief rule-based systems decreases with the increase of generations, indicating that the trained BRBs have better parameter values than the initial BRBs and the corresponding belief rule-based systems are able to classify input data with desired accuracy.

In the last step, the weight learning method shown in Section 3.3.2 is applied to determine the weight of each belief rule-based system for a minimum classification error of Ensemble-BRB. Fig. 4 shows the change of classification error of Ensemble-BRB for training data within 200 iterations and Table 5 lists the obtained weights for 15 belief rule-based systems. From Fig. 4, the classification error of Ensemble-BRB for training data continues to be decreased within 200 iterations of the weight learning method, indicating that although Fig. 3 shows that the classification accuracy of each belief rule-based system has been improved through parameter learning, this does not mean those belief rule-based systems are perfect classifiers due to the inevitable truth that single attribute selection method fails to provide an effective set of attributes for constructing BRBs, leading to impacting the classification accuracy of belief rule-based systems. Meanwhile, it is clear from Table 5 that each belief rule-based system has different importance on Ensemble-BRB, *i.e.*, BRB-{Ca, Ba, RI} is much more important than BRB-{Mg, Al, Ba} according to the weight of these two belief rule-based systems, namely 0.9294 and 0.0246, thus the final inferential output of Ensemble-BRB depends more upon the inferential output of BRB- {Ca, Ba, RI} comparing to BRB-{Mg, Al, Ba}.

To illustrate the effectiveness of the proposed BRB modeling method, Table 6 shows the classification results on testing data of 15 belief rule-based systems and Ensemble-BRB. For the belief rule-based systems with single attribute selection, their numbers of classification errors are greater than Ensemble-BRB since the input data that are misclassifies by some members of Ensemble-BRB have chance to be correctly classified by others, so that the final number of classification errors is decreased. In other words, the classification accuracy of Ensemble-BRB is better than that of other belief rule-based systems.

4.3. Comparative analysis with different numbers of selected attributes

In order to investigate the influences of the selected attribute's numbers on the proposed BRB modeling method, the number of selected attributes is set as 3, 4, and 5, respectively, to perform sensitivity analysis for Ensemble-BRB and the belief rule-based systems constructed by single attribute selection. The experimental results are shown in Tables 7, 8, and 9, respectively.

In the study of sensitivity analysis, six belief rule-based systems constructed by singe attribute selection shown in Table 1 are

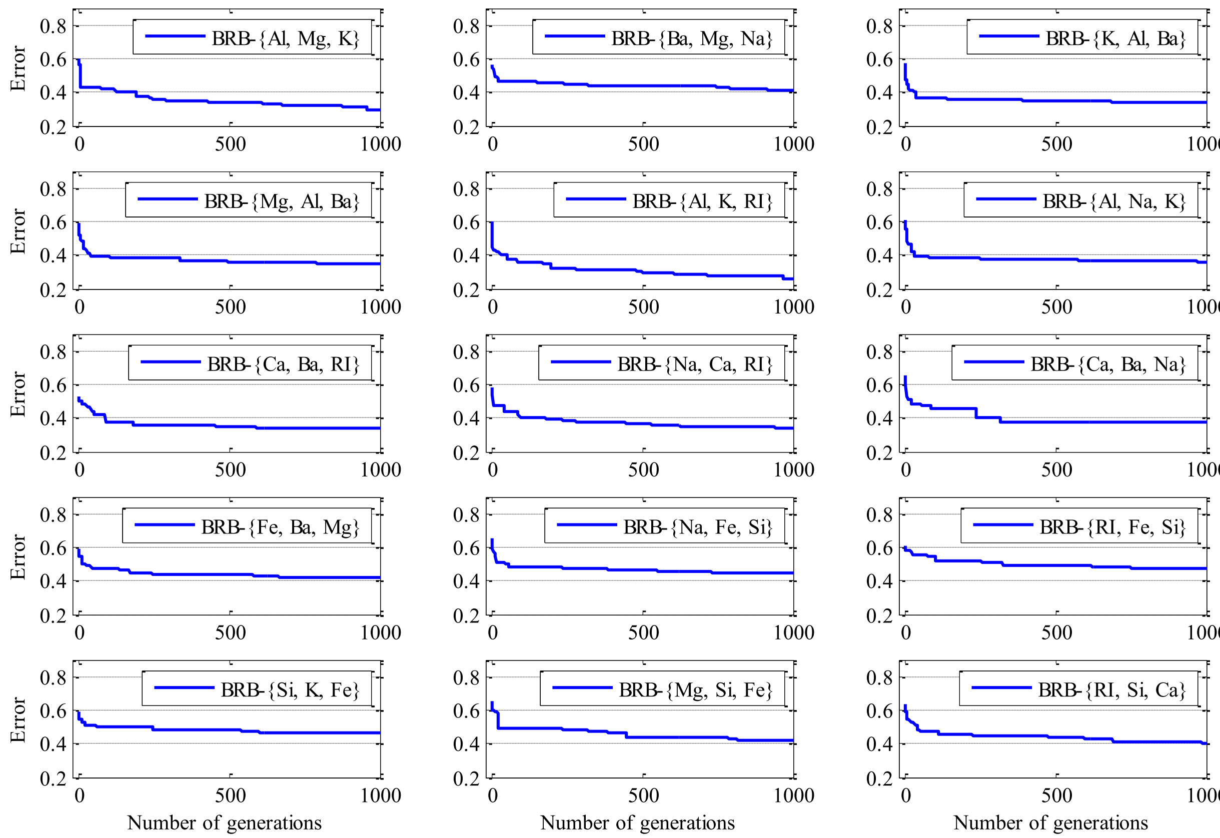

Fig. 3. Classification errors of 15 belief rule-based systems for training data.

Table 5
Weights of 15 belief rule-based systems.

BRB type	Weight	BRB type	Weight	BRB type	Weight
BRB-{Al, Mg, K}	0.1050	BRB-{Ba, Mg, Na}	0.9791	BRB-{K, Al, Ba}	0.7933
BRB-{Mg, Al, Ba}	0.0246	BRB-{Al, K, RI}	0.9799	BRB-{Al, Na, K}	0.6913
BRB-{Ca, Ba, RI}	0.9294	BRB-{Na, Ca, RI}	0.6858	BRB-{Ca, Ba, Na}	0.3113
BRB-{Fe, Ba, Mg}	0.5621	BRB-{Na, Fe, Si}	0.6336	BRB-{RI, Fe, Si}	0.3298
BRB-{Si, K, Fe}	0.8136	BRB-{Mg, Si, Fe}	0.7403	BRB-{RI, Si, Ca}	0.2640

Table 6
Classification results of different belief rule-based systems for testing data.

BRB type	Error	Accuracy	BRB type	Error	Accuracy
BRB-{Al, Mg, K}	9	60.87%	BRB-{Ca, Ba, Na}	11	52.17%
BRB-{Ba, Mg, Na}	13	43.48%	BRB-{Fe, Ba, Mg}	11	52.17%
BRB-{K, Al, Ba}	9	60.87%	BRB-{Na, Fe, Si}	12	47.83%
BRB-{Mg, Al, Ba}	11	52.17%	BRB-{RI, Fe, Si}	12	47.83%
BRB-{Al, K, RI}	8	65.22%	BRB-{Si, K, Fe}	12	47.83%
BRB-{Al, Na, K}	10	56.52%	BRB-{Mg, Si, Fe}	14	39.13%
BRB-{Ca, Ba, RI}	9	60.87%	BRB-{RI, Si, Ca}	12	47.83%
BRB-{Na, Ca, RI}	9	60.87%	Ensemble-BRB	7	69.57%

Table 7
Comparison of Ensemble-BRB and other belief rule-based systems (3 selected attributes).

Dataset	AS-CS-BRB	AS-GR-BRB	AS-IG-BRB	AS-ORC-BRB	AS-RF-BRB	AS-SVM-BRB	Ensemble-BRB
Diabetes	77.86% (4.5)	78.88% (2)	77.86% (4.5)	78.12% (3)	76.08% (7)	76.59% (6)	79.64% (1)
Cancer	91.21% (7)	93.15% (5)	91.56% (6)	93.32% (4)	94.02% (2)	93.85% (3)	95.43% (1)
Wine	97.19% (1)	91.01% (6)	93.26% (4.5)	93.26% (4.5)	89.32% (7)	95.51% (2)	94.93% (3)
Glass	63.08% (4)	53.74% (7)	64.49% (2.5)	59.81% (5)	64.49% (2.5)	59.35% (6)	66.82% (1)
Cleveland	59.60% (1.5)	55.22% (5)	59.60% (1.5)	51.52% (7)	55.89% (4)	54.21% (6)	58.59% (3)
Heart	84.44% (3)	81.85% (5)	84.44% (3)	84.44% (3)	76.67% (6)	75.19% (7)	85.19% (1)
Monks	97.22% (5.5)	97.22% (5.5)	97.22% (5.5)	97.45% (3)	97.22% (5.5)	99.07% (1)	98.84% (2)
Pima	73.96% (3.5)	73.96% (3.5)	73.83% (5)	72.79% (7)	74.61% (1)	73.70% (6)	74.22% (2)
Average rank	3.75	4.875	4.0625	4.5625	4.375	4.625	1.75

Table 8
Comparison of Ensemble-BRB and other belief rule-based systems (4 selected attributes).

Dataset	AS-CS-BRB	AS-GR-BRB	AS-IG-BRB	AS-ORC-BRB	AS-RF-BRB	AS-SVM-BRB	Ensemble-BRB
Diabetes	78.37% (3)	78.12% (4.5)	77.10% (6)	78.88% (2)	75.32% (7)	78.12% (4.5)	80.66% (1)
Cancer	94.20% (4.5)	95.25% (1)	94.02% (6)	93.50% (7)	94.55% (3)	94.20% (4.5)	94.90% (2)
Wine	93.82% (3.5)	91.57% (7)	93.82% (3.5)	93.26% (5)	94.94% (2)	92.70% (6)	96.07% (1)
Glass	62.62% (1.5)	59.81% (6.5)	61.22% (4)	59.81% (6.5)	61.68% (3)	60.75% (5)	62.62% (1.5)
Cleveland	56.57% (1.5)	54.21% (5.5)	55.89% (4)	54.21% (5.5)	56.57% (1)	53.54% (7)	56.23% (3)
Heart	80.00% (3.5)	79.26% (5)	80.00% (3.5)	78.89% (6)	77.04% (7)	81.48% (2)	82.22% (1)
Monks	96.53% (6)	96.53% (6)	96.53% (6)	97.45% (3)	97.22% (4)	99.31% (2)	100% (1)
Pima	73.05% (6)	73.96% (4)	73.05% (6.5)	74.87% (3)	73.31% (5)	77.21% (1)	76.82% (2)
Average rank	3.75	4.9375	4.9375	4.75	4.0625	4	1.5625

Table 9
Comparison of Ensemble-BRB and other belief rule-based systems (5 selected attributes).

Dataset	AS-CS-BRB	AS-GR-BRB	AS-IG-BRB	AS-ORC-BRB	AS-RF-BRB	AS-SVM-BRB	Ensemble-BRB
Diabetes	79.64% (2)	79.64% (2)	79.39% (4)	77.61% (6)	75.83% (7)	79.13% (5)	79.64% (2)
Cancer	93.15% (5.5)	94.90% (3)	93.15% (5.5)	92.62% (7)	94.20% (4)	96.31% (2)	96.49% (1)
Wine	95.51% (2)	91.57% (7)	96.63% (1)	92.13% (6)	92.70% (5)	93.26% (4)	94.94% (3)
Glass	59.81% (3)	64.02% (1)	59.81% (3)	59.35% (5.5)	58.88% (7)	59.35% (5.5)	59.81% (3)
Cleveland	56.57% (3)	57.91% (1)	55.89% (5)	53.87% (7)	54.88% (6)	56.23% (4)	56.90% (2)
Heart	78.52% (5.5)	78.89% (3.5)	78.89% (3.5)	77.41% (7)	78.52% (5.5)	84.44% (1)	81.11% (2)
Monks	95.60% (6)	95.60% (6)	95.60% (6)	97.22% (4)	99.53% (2)	99.31% (3)	100% (1)
Pima	74.48% (4.5)	74.48% (4.5)	74.87% (3)	73.70% (7)	74.09% (6)	77.34% (1)	75.52% (2)
Average rank	3.9375	3.5	3.875	6.1875	5.3125	3.1875	2

Table 10
Friedman and Holland tests to compare different belief rule-based systems (α=0.1).

No. of attributes	Indicator	AS-CS-BRB	AS-GR-BRB	AS-IG-BRB	AS-ORC-BRB	AS-RF-BRB	AS-SVM-BRB
3 selected attributes	*p* value	0.0641	0.0038	0.0323	0.0092	0.0151	0.0078
	Critical value	0.1000	0.4686	0.1900	0.3439	0.2710	0.4095
	Hypothesis	**Rejected**	**Rejected**	**Rejected**	**Rejected**	**Rejected**	**Rejected**
4 selected attributes	*p* value	0.0428	0.0018	0.0018	0.0032	0.0206	0.0240
	Critical value	0.1000	0.4686	0.4095	0.3439	0.2710	0.1900
	Hypothesis	**Rejected**	**Rejected**	**Rejected**	**Rejected**	**Rejected**	**Rejected**
5 selected attributes	*p* value	0.0728	0.1649	0.0826	0.0001	0.0022	0.2716
	Critical value	0.3439	0.1900	0.2710	0.4686	0.4095	0.1000
	Hypothesis	**Rejected**	**Rejected**	**Rejected**	**Rejected**	**Rejected**	Accepted

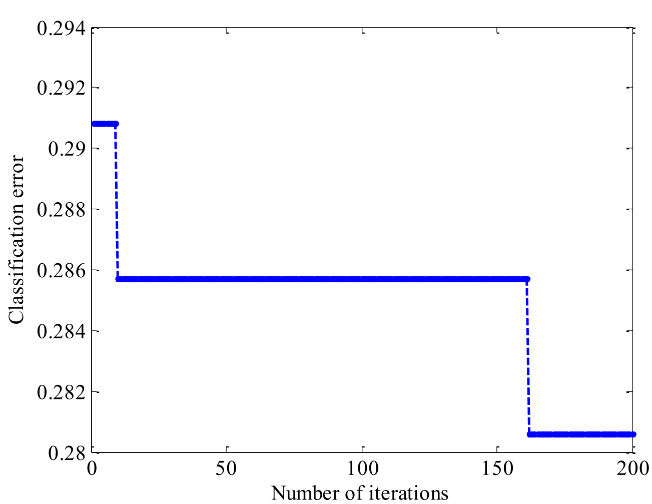

Fig. 4. Classification errors of Ensemble-BRB for training data.

related with AS-IG, AS-GR, AS-CS, AS-RF, AS-ORC, and AS-SVM, thus all of them are abbreviated as AS-IG-BRB, AS-GR-BRB, AS-CS-BRB, AS-RF-BRB, AS-ORC-BRB, and AS-SVM-BRB, respectively. Tables 7, 8, and 9 show their accuracies and ranks for the eight classification datasets, in which Ensemble-BRB is able to have the best accuracies in four of eight classification datasets when setting the number of selected attributes as 3 (namely 79.64%, 95.43%, 66.82%, and 85.19% for Diabetes, Cancer, Glass, and Heart) and 4 (namely 80.66%, 96.07%, 82.22%, and 100% for Diabetes, Wine, Heart, and Monks), and have the best accuracies on two classification datasets (namely 96.49% and 100% for Cancer and Monks). Although Ensemble BRB fails to have the best accuracies on all of datasets, the average rank of Ensemble BRB is better than other belief rule-based system for 3, 4, and 5 selected attributes.

To further compare the classification accuracy of Ensemble-BRB and other six belief rule-based systems, Friedman and Holland tests are used to provide the statistical analysis when Ensemble-BRB is selected as the target classifier. The results can be found in Table 10. From Table 10, apart from AS-SVM-BRB under 5 selected attributes, all hypotheses are in favor of the significant difference between Ensemble-BRB and other six belief rule-based systems, which reveals that Ensemble- BRB can significantly better than the belief rule-based systems constructed by singe attribute selection, but its advantages may disappear because there is few difference when considering more numbers of selected attributes.

In summary, for the study of sensitivity analysis on different numbers of selected attributes, the experiment results have showed that Ensemble-BRB has much higher classification accuracy than the belief rule-based system under single attribute selection. Moreover, the statistical analysis based on Friedman and Holland tests has demonstrated the superiority of the ensemble BRB modeling method.

Table 11
Comparison of Ensemble-BRB and conventional rule-based classifiers.

Dataset	SLAVE	FH-GBML	SGERD	FARC–HD	Chi-FRBCS	Micro-EBRB	Ensemble-BRB
Diabetes	77.10% (2.5)	70.23% (7)	76.59 (4)	77.10% (2.5)	72.80% (6)	74.91% (5)	80.66% (1)
Cancer	92.33% (4)	92.26% (6)	90.68% (7)	95.25% (3)	92.32% (5)	96.49% (1.5)	96.49% (1.5)
Wine	89.47% (7)	92.61% (4)	91.88% (5)	94.35% (3)	90.17% (6)	95.84% (2)	96.07% (1)
Glass	58.05% (5)	57.99% (6)	58.49% (4)	70.24% (1)	50.37% (7)	63.32% (3)	66.82% (2)
Cleveland	48.82% (6)	53.51% (4)	51.59% (5)	55.24% (3)	37.71% (7)	56.57% (2)	58.59% (1)
Heart	71.36% (6)	75.93% (4)	73.21% (5)	84.88% (2)	49.63% (7)	78.89% (3)	85.19% (1)
Monks	97.26% (4)	98.18% (3)	80.65% (5)	99.77% (2)	42.82% (7)	75.00% (6)	100% (1)
Pima	73.71% (4)	75.26% (3)	73.37% (5)	75.66% (2)	71.74% (6)	67.19% (7)	76.82% (1)
Average rank	4.8125	4.625	5	2.3125	6.375	3.6875	1.1875

Table 12
Friedman and Holland tests to compare the conventional rule-based classifiers ($\alpha = 0.1$).

Indicator	SLAVE	FH-GBML	SGERD	FARC–HD	Chi-FRBCS	Micro-EBRB
p value	0.0008	0.0015	0.0004	0.2976	0.0000	0.0206
Critical value	0.3439	0.2710	0.4095	0.1000	0.4686	0.1900
Hypothesis	**Rejected**	**Rejected**	**Rejected**	Accepted	**Rejected**	**Rejected**

Table 13
Comparison of Ensemble-BRB and conventional machine learning-based classifiers.

Dataset	KNN	Bagging	DT	RF	ANN	SVM	Ensemble-BRB
Diabetes	74.30% (6)	75.39% (5)	77.61% (4)	77.86% (3)	79.13% (2)	66.92% (7)	80.66% (1)
Cancer	94.02% (5)	94.20% (4)	93.32% (6)	95.61% (3)	96.31% (2)	62.74% (7)	96.49% (1)
Wine	97.19% (1.5)	96.07% (3.5)	92.13% (6)	94.38% (5)	97.19% (1.5)	44.38% (7)	96.07% (3.5)
Glass	61.21% (6)	72.43% (1)	67.76% (4)	71.50% (2)	39.72% (7)	69.16% (3)	66.82% (5)
Cleveland	56.23% (3)	56.57% (2)	55.56% (4)	54.21% (5)	51.85% (7)	53.87% (6)	58.59% (1)
Heart	83.70% (2)	81.85% (4)	82.96% (3)	81.11% (5)	80.74% (6)	55.56% (7)	85.19% (1)
Monks	84.26% (7)	100% (2)	97.22% (5.5)	99.77% (4)	100% (2)	97.22% (5.5)	100% (2)
Pima	70.96% (6)	77.21% (1)	75.39% (4)	75.13% (5)	76.04% (3)	65.10% (7)	76.82% (2)
Average rank	4.5625	2.8125	4.5625	4	3.8125	6.1875	2.0625

Table 14
Friedman and Holland tests to compare the conventional machine learning-based classifiers ($\alpha = 0.1$).

Indicator	KNN	Bagging	DT	RF	ANN	SVM
p value	0.0206	0.4875	0.0206	0.0728	0.1052	0.0001
Critical value	0.4095	0.1000	0.3439	0.2710	0.1900	0.4686
Hypothesis	**Rejected**	Accepted	**Rejected**	**Rejected**	**Rejected**	**Rejected**

4.4. Comparative analysis with conventional classifiers

To further verify the validity of the proposed BRB modeling method, two kinds of comparative experiments are carried out in this section by using the datasets shown in Table 3. The relevant experimental results can be found in Tables 11–13 and 14, respectively.

In the first experiment, six conventional rule-based classifiers are introduced to compare with Ensemble-BRB. These classifiers include: fuzzy rule-based classification system proposed by Chi et al. (Chi-FRBCS); structural learning algorithm on vague environment (SLAVE), fuzzy hybrid genetic-based machine learning algorithm (FH-GBML), steady-state genetics algorithm for extracting fuzzy classification rule from data (SGERD), fuzzy association rule-based classification method for high-dimensional problems (FARC–HD), and Micro-EBRB. Table 11 shows the accuracy of Ensemble-BRB in comparison with six rule-based classifiers, whose experimental results are derived from the previous studies (Alcala-Fdez et al., 2011; Yang, Liu, Wang & Martínez, 2019). In such group of comparison, FARC–HD obtains the best accuracy for Glass better than Ensemble-BRB, which has the second best accuracy. Apart from this dataset, Ensemble-BRB obtains the best accuracies on Diabetes, Cancer, Wine, Cleveland, Heart, Monks, and Pima over SLAVE, FH-GBML, SGERD, FARC–HD, Chi-FRBCS, and Micro-EBRB classifiers. Therefore, the average rank of Ensemble-BRB is 1.1875, smaller than the average rank of other classifiers.

Table 12 shows the statistical analysis of classification accuracy when Ensemble-BRB is selected as the target classifier for the Friedman and Holland tests. It is clear from Table 12 that the hypotheses regarding SLAVE, FG-GBML, SGERD, Chi-FRBCS, and Micro-EBRB are rejected, namely Ensemble-BRB has better accuracy then these classifiers. Although the hypothesis of FRAC–HD is accepted, the rank of these two classifiers shown in Table 11 can clearly reveal the superiority of Ensemble-BRB, *i.e.*, Ensemble-BRB has the best accuracies at almost all datasets.

In the 2nd experiment, Ensemble-BRB continues to be compared with conventional machine learning-based classifiers, including k Nearest Neighbor (KNN), Bagging, Decision Tree (DT), Random Forest (RF), Artificial Neural Network (ANN), and Support Vector Machine (SVM). Note that all these machine learning-based classifiers are implemented by the WEKA software with default setting, *i.e.*, 20% number of training data is defined as the neighbors for KNN, 5% number of training data is defined as the minimum number of data per leaf for DT, the number of random trees is defined as 5 for RF, and the number of iterations is defined as 10 for ANN. Table 13 shows the accuracy of Ensemble-BRB in comparison with six other classifiers, whose experimental results are derived from the previous study (Yang et al., 2019). Ensemble-

BRB still has the best accuracies at Diabetes, Cancer, Cleveland, and Heart, respectively. For Glass, the commonly used ensemble learning classifiers, named Bagging and RF, show their powerful performances and obtain 72.43% and 71.50% accuracies better than Ensemble-BRB. Despite this fact, it is still possible to see considerable rank of Ensemble-BRB which is superior to KNN and ANN on Glass.

Table 14 shows the statistical analysis of classification accuracy while Ensemble-BRB is selected as the target classifier for the Friedman and Holland tests. It is clear from Table 14 that the hypotheses regarding KNN, DT, RF, ANN, and SVM are rejected, which supports that Ensemble-BRB has better classification accuracy than these classifiers. But beyond those, the hypothesis of Bagging is accepted, which means an insignificant difference between Ensemble-BRB and Bagging. However, the rank of these two classifiers shown in Table 13 can prove the superiority of Ensemble-BRB because Ensemble -BRB has the best accuracies at four of eight datasets, *e.g.*, Diabetes, Cancer, Cleveland, and Heart.

In summary, for the comparison of Ensemble-BRB and conventional classifiers, the results have demonstrated that the proposed BRB modeling method is able to improve the classification accuracy of belief rule-based systems. Furthermore, it is clear that the proposed BRB modeling method provides an effective approach to address the classification problems with a large number of attributes.

5. Conclusions

This study proposed an ensemble BRB modeling method with diverse attribute selection and CCR for classification problems. The main motive of proposing this method is to overcome the rule combinatorial explosion challenge while there is a large number of antecedent attributes in a BRB. Eight classification datasets were used to validate the effectiveness of the proposed BRB modeling method and compare with some popular classifiers. The main contributions can be summarized into three aspects below:

(1) Previous studies on the rule combinatorial explosion challenge just focused on single attribute selection method, leading to an inevitable situation that BRB modeling has to suffer from the inherent weakness of the attribute selection method. Therefore, multiple different attribute selection methods are used to perform diverse attribute selection for taking full advantages of each method in BRB modeling.
(2) CCR is one of the few combination methods that are suitable for combining non-independent evidences but it has not yet been investigated for an analytical CCR which has an advantage of explicit combination processes. Therefore, the analytical CCR is deduced from the recursive CCR to propose a new output combination method for combining the output of belief rule-based systems. Due to this method, it is unnecessary to require the independence of belief rule-based systems while constructing ensemble belief rule-based systems for complex problems.
(3) The detailed procedure of BRB modeling, sensitivity analysis, and two comparative experiments were provided for verification. It was validated that the proposed BRB modeling method can facilitate downsizing BRBs and guarantee desirable classification accuracy for belief rule-based systems. More importantly, the classification accuracy of ensemble belief rule-based systems has significant advantage better than other conventional classifiers on some classification datasets.

For future research, referential values adjustment as well as complex practical problems should be further studied to improve BRB modeling, which would promote the application of belief rule-based systems.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Credit authorship contribution statement

Long-Hao Yang: Writing - original draft, Conceptualization, Methodology, Data curation, Formal analysis. **Fei-Fei Ye:** Investigation, Writing - review & editing, Supervision. **Ying-Ming Wang:** Writing - review & editing, Supervision, Conceptualization, Investigation, Validation.

Acknowledgments

This research was supported by the Social Science Planning Fund Project of Fujian Province (Nos. FJ2019C032 and FJ2018C014), the National Natural Science Foundation of China (Nos. 61773123, 71371053, 71501047, 71701050, 71801050, 71403055, and 71872047), the Humanities and Social Science Foundation of the Ministry of Education, China (No. 14YJC630056), and the Natural Science Foundation of Fujian Province, China (No. 2016J05171 and 2015J01248).

Appendix A. The derivation of the analytical CCR for classification problems

CCR is one of extensions of Dempster‘s rule which assumes the combined evidence should be independent. However, this assumption of Dempster‘s rule is not always verified in practice. To overcome this dilemma, CCR was proposed by Denoeux (2008) for the combination of non-independent evidences.

According to Denoeus (2008), CCR can be represented as the following proposition and definitions.

Proposition 1. *Let A_1,..., A_N be N subset of $\mathbf{\Omega}$ such that $A_i \cap A_j = \emptyset$ for all i, $j \in \{1,..., N\}$, and let m be a basic belief assignment on $\mathbf{\Omega}$ with focal sets A_1,..., A_N, and $\mathbf{\Omega}$. Assume that $m(\mathbf{\Omega}) + \sum_{n=1}^{N} m(A_n) \leq 1$ and $\emptyset$ is a focal set. The conjunctive weight function associated to m is:*

$$\varpi(A) = \begin{cases} \frac{m(\mathbf{\Omega})}{m(A_n)+m(\mathbf{\Omega})}, & if\ A = A_n \\ m(\mathbf{\Omega}) \prod_{n=1}^{N} \left(1 + \frac{m(A_n)}{m(\mathbf{\Omega})}\right), & if\ A = \emptyset \\ 1, & otherwise \end{cases}. \tag{B1}$$

Definition 1 (CCR)**.** Let m_1 and m_2 be nondogmatic basic belief assignments. Their combined basic belief assignment $m_{1,2}$ using CCR is calculated using transferable belief model conjunctive rule (TBMCR) for the combination of generalized sample basic belief assignment (GSBBA) $A^{\varpi_1(A) \wedge \varpi_2(A)}$ for all $A \in \mathbf{\Omega}$ such that $\varpi_1(A) \wedge \varpi_2(A) \neq 1$.

Definition 2 (TBMCR)**.** Let m_1 and m_2 be basic belief assignments, $\mathbf{\Omega}$ be a finite frame of discernment. The combined basic belief assignment of m_1 and m_2 using TBMCR is as follows:

$$m_{1,2}(A) = \sum_{B \cap C = A} m_1(B) m_2(C), \forall A \subseteq \mathbf{\Omega}. \tag{B2}$$

To deduce the analytical CCR, suppose that there is a classification problem with N classes, thus the finite frame of discernment can be denoted as $\mathbf{\Omega} = \{D_1,..., D_N\}$. While the weight and the belief distribution of the kth ($k = 1,..., K$) belief rule-based system are ω_k and $f_k(\boldsymbol{x})=\{(D_n, \beta_n^k)\}$ ($k = 1,..., K$), respectively, the basic belief assignment in the D_n and $\mathbf{\Omega}$ can be calculated as follows:

$$m_k(D_n) = \omega_k \beta_n^k;\ n = 1, ..., N. \tag{B3}$$

$$m_k(\mathbf{\Omega}) = \omega_k\left(1-\sum_{n=1}^{N}\beta_n^k\right)+1-\omega_k = 1-\omega_k\sum_{n=1}^{N}\beta_n^k. \tag{B4}$$

According to Proposition 1, the conjunctive weight function associated to the D_n can be obtained as follows:

$$\varpi_k(D_n) = \frac{m_k(\mathbf{\Omega})}{m_k(D_n)+m_k(\mathbf{\Omega})} = \frac{1-\omega_k\sum_{i=1}^{N}\beta_i^k}{\omega_k\beta_n^k+1-\omega_k\sum_{i=1}^{N}\beta_i^k}. \tag{B5}$$

Thereafter, the combination of these conjunctive weight functions for the D_n, denoted as $\varpi(D_n)$, is obtained as follows:

$$\varpi(D_n) = \varpi_1(D_n)\wedge\cdots\wedge\varpi_K(D_n) = \min_{k=1,\ldots,K}\{\varpi_k(D_n)\}. \tag{B6}$$

Next, based on Definition 1, the GSBBA on the D_n and $\mathbf{\Omega}$ for the nth conjunctive weight function $\varpi(D_n)$ can be denoted as follows:

$$m_n^{\varpi(D_n)}(D_i) = \begin{cases} 1-\varpi(D_i); & if\ i=n \\ 0 & otherwise \end{cases}. \tag{B7}$$

$$m_n^{\varpi(D_n)}(\mathbf{\Omega}) = \varpi(D_n). \tag{B8}$$

While combining the GSBBA of the 1st and the 2nd conjunctive weight function based on Definition 2, the integrated GSBBA on the D_1 and $\mathbf{\Omega}$ can be calculated as follows:

$$\begin{aligned} m_{1-2}^{\varpi(D_{1-2})}(D_1) &= \sum_{B\cap C=D_1} m_1^{\varpi(D_1)}(B)\cdot m_2^{\varpi(D_2)}(C) \\ &= m_1^{\varpi(D_1)}(D_1)\cdot m_2^{\varpi(D_2)}(D_1) + m_1^{\varpi(D_1)}(D_1)\cdot m_2^{\varpi(D_2)}(\mathbf{\Omega}) \\ &\quad + m_1^{\varpi(D_1)}(\mathbf{\Omega})\cdot m_2^{\varpi(D_2)}(D_1) \\ &= m_1^{\varpi(D_1)}(D_1)\cdot 0 + m_1^{\varpi(D_1)}(D_1)\cdot m_2^{\varpi(D_2)}(\mathbf{\Omega}) \\ &\quad + m_1^{\varpi(D_1)}(\mathbf{\Omega})\cdot 0 \\ &= m_1^{\varpi(D_1)}(D_1)\cdot m_2^{\varpi(D_2)}(\mathbf{\Omega}) \\ &= \varpi(D_2) - \varpi(D_1)\cdot\varpi(D_2) \\ &= \prod_{i=1,i\neq 1}^{2}\varpi(D_i) - \prod_{i=1}^{2}\varpi(D_i). \end{aligned} \tag{B9}$$

$$\begin{aligned} m_{1-2}^{\varpi(D_{1-2})}(\mathbf{\Omega}) &= \sum_{B\cap C=\mathbf{\Omega}} m_1^{\varpi(D_1)}(B)\cdot m_2^{\varpi(D_2)}(C) \\ &= m_1^{\varpi(D_1)}(\mathbf{\Omega})\cdot m_2^{\varpi(D_2)}(\mathbf{\Omega}) \\ &= \varpi(D_1)\cdot\varpi(D_2) \\ &= \prod_{i=1}^{2}\varpi(D_i). \end{aligned} \tag{B10}$$

Suppose the following equations are true for combining the first N-1 conjunctive weight function and let N_1=N-1

$$m_{1-N_{1'}}^{\varpi(D_{1-N_{1'}})}(D_1) = \prod_{i=1,i\neq 1}^{N-1}\varpi(D_i) - \prod_{i=1}^{N-1}\varpi(D_i). \tag{B11}$$

$$m_{1-N_{1'}}^{\varpi(D_{1-N_{1'}})}(\mathbf{\Omega}) = \prod_{i=1}^{N-1}\varpi(D_i). \tag{B12}$$

The above combined GSBBA are further aggregated with the Nth conjunctive weight function. The combined GSBBA is then given below:

$$\begin{aligned} m_{1-N}^{\varpi(D_{1-N})}(D_1) &= \sum_{B\cap C=D_1} m_{1-N_{1'}}^{\varpi(D_{1-N_1})}(B)\, m_N^{\varpi(D_N)}(C) \\ &= m_{1-N_{1'}}^{\varpi(D_{1-N_1})}(D_1)\, m_N^{\varpi(D_N)}(D_1) + m_{1-N_{1'}}^{\varpi(D_{1-N_1})}(D_1)\, m_N^{\varpi(D_N)}(\mathbf{\Omega}) \\ &\quad + m_{1-N_{1'}}^{\varpi(D_{1-N_1})}(\mathbf{\Omega})\, m_N^{\varpi(D_N)}(D_1) \\ &= m_{1-N_{1'}}^{\varpi(D_{1-N_1})}(D_1)\cdot 0 + m_{1-N_{1'}}^{\varpi(D_{1-N_1})}(D_1)\, m_N^{\varpi(D_N)}(\mathbf{\Omega}) \\ &\quad + m_{1-N_{1'}}^{\varpi(D_{1-N_1})}(\mathbf{\Omega})\cdot 0 \\ &= m_{1-N_{1'}}^{\varpi(D_{1-N_1})}(D_1)\, m_N^{\varpi(D_N)}(\mathbf{\Omega}) \\ &= \left(\prod_{i=1,i\neq 1}^{N-1}\varpi(D_i) - \prod_{i=1}^{N-1}\varpi(D_i)\right)\cdot\varpi(D_N) \\ &= \prod_{i=1,i\neq 1}^{N}\varpi(D_i) - \prod_{i=1}^{N}\varpi(D_i). \end{aligned} \tag{B13}$$

$$\begin{aligned} m_{1-N}^{\varpi(D_{1-N})}(\mathbf{\Omega}) &= \sum_{B\cap C=\mathbf{\Omega}} m_{1-N_1}^{\varpi(D_{1-N_1})}(B)\cdot m_N^{\varpi(D_N)}(C) \\ &= m_{1-N_1}^{\varpi(D_{1-N_1})}(\mathbf{\Omega})\cdot m_N^{\varpi(D_N)}(\mathbf{\Omega}) \\ &= \prod_{i=1}^{N-1}\varpi(D_i)\cdot\varpi(D_N) \\ &= \prod_{i=1}^{N}\varpi(D_i). \end{aligned} \tag{B14}$$

Therefore, the above equations are true for the integrated GSBBA on the D_1 and $\mathbf{\Omega}$ by aggregating N conjunctive weight functions. Similarly, the integrated GSBBA on the D_n aggregating N conjunctive weight functions can be deduced and denoted as follows:

$$m_{1-N}^{\varpi(D_{1-N})}(D_n) = \prod_{i=1,i\neq n}^{N}\varpi(D_i) - \prod_{i=1}^{N}\varpi(D_i). \tag{B15}$$

References

Alcala-Fdez, J., Fernández, A., Luengo, J., Derrac, J., García, S., Sanchez, L., et al. (2011). KEEL data-mining software tool-data set repository, integration of algorithms and experimental analysis framework. *Journal of Multiple- Valued Logic and Soft Computing, 17*, 255–287.

Bolón-Canedo, V., & Alonso-Betanzos, A. (2019). Ensembles for feature selection: A review and future trends. *Information Fusion, 52*, 1–12.

Breiman, L. (2001). Random forests. *Machine Learning, 45*, 5–32.

Calzada, A., Liu, J., Wang, H., & Kashyap, A. (2015). A new dynamic rule activation method for extended belief rule-based systems. *IEEE Transactions on Knowledge and Data Engineering, 27*(4), 880–894.

Chang, L. L., Jiang, J., Sun, J. B., Chen, Y. W., Zhou, Z. J., & Xu, X. B. (2019a). Disjunctive belief rule base spreading for threat level assessment with heterogeneous, insufficient, and missing information. *Information Sciences, 476*, 106–131.

Chang, L. L., Zhou, Y., Jiang, J., Li, M. J., & Zhang, X. H. (2013). Structure learning for belief rule base expert system: A comparative study. *Knowledge-Based Systems, 39*, 159–172.

Chang, L. L., Zhou, Z. J., Liao, H. C., Chen, Y. W., Tan, X., & Herrera, F. (2019b). Generic disjunctive belief rule base modeling, inferencing, and optimization. *IEEE Transactions on Fuzzy Systems, 27*(9), 1866–1880.

Chang, L. L., Zhou, Z. J., You, Y., Yang, L. H., & Zhou, Z. G. (2016). Belief rule based expert system for classification problems with new rule activation and weight calculation procedures. *Information Sciences, 336*, 75–91.

Chang, L. L., Zhou, Z. J., Chen, Y. W., Xu, X. B., Sun, J. B., Liao, T. J., & Tan, X. (2018). Akaike Information Criterion-based conjunctive belief rule base learning for complex system modeling. *Knowledge-Based Systems, 161*, 47–64.

Chen, Y., Chen, Y. W., Xu, X. B., Pan, C. C., Yang, J. B., & Yang, G. K. (2015). A data–driven approximate causal inference model using the evidential reasoning rule. *Knowledge-Based Systems, 88*, 264–272.

Chen, Y. W., Yang, J. B., Xu, D. L., Zhou, Z. J., & Tang, D. W. (2011). Inference analysis and adaptive training for belief rule based systems. *Expert Systems with Applications, 38*(10), 12845–12860.

Denœux, T. (2008). Conjunctive and disjunctive combination of belief functions induced by nondistinct bodies of evidence. *Artificial Intelligence, 172*(2–3), 234–264.

Dua, D., & Graff, C. (2019). UCI machine learning repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.

García, S., Fernández, A., Luengo, J., & Herrera, F. (2010). Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining-Experimental analysis of power. *Information Sciences, 180*, 2044–2064.

Jiao, L. M., Pan, Q., Denœux, T., Liang, Y., & Feng, X. X. (2015). Belief rule-based classification system: Extension of FRBCS in belief functions framework. *Information Sciences, 309*, 26–49.

Ke, X. L., Ma, L. Y., & Wang, Y. (2017). A modified belief rule based model for uncertain nonlinear systems identification. *Journal of Intelligent & Fuzzy Systems, 32*(6), 3879–3891.

Lin, Y. Q., Fu, Y. G., Su, Q., Wang, Y. M., & Gong, X. T. (2017). A rule activation method for extended belief rule base with VP-tree and MVP-tree. *Journal of Intelligent & Fuzzy Systems, 33*(6), 3695–3705.

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

Seijo-Pardo, B., Porto-Díaz, I., Bolón-Canedo, V., & Alonso-Betanzos, A. (2017). Ensemble feature selection: Homogeneous and heterogeneous approaches. *Knowledge-Based Systems, 118*, 124–139.

Storn, R., & Price, K. (1997). Differential evolution-a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization, 11*, 341–359.

Sun, J. B., Huang, J. X., Chang, L. L., Jiang, J., & Tan, Y. J. (2018). BRBcast: A new approach to belief rule-based system parameter learning via extended causal strength logic. *Information Sciences, 444*, 51–71.

Tang, X.L., .Xiao, M.Q., .Liang, Y.J., .Zhu, H.Z., .& Li, J.F. (.2019).Online updating belief-rule-base using Bayesian estimation, Knowledge-based Systems,, In Press.

Wang, Y. M., Yang, J. B., & Xu, D. L. (2006). Environmental impact assessment using the evidential reasoning approach. *European Journal of Operational Research, 174*, 1885–1913.

Wang, Y. M., Yang, J. B., Xu, D. L., & Chin, K. S. (2009). Consumer preference prediction by using a hybrid evidential reasoning and belief rule-based methodology. *Expert Systems with Applications, 36*(4), 8421–8430.

Wang, Y. M., Yang, L. H., Fu, Y. G., Chang, L. L., & Chin, K. S. (2016). Dynamic rule adjustment approach for optimizing belief rule-base expert system. *Knowledge-Based Systems, 96*, 40–60.

Xu, D. L., Liu, J., Yang, J. B., Liu, G. P., Wang, J., Jenkinson, I., et al. (2007). Inference and learning methodology of belief- rule-based expert system for pipeline leak detection. *Expert Systems with Applications, 32*(1), 103–113.

Yang, J. B. (2001). Rule and utility based evidential reasoning approach for multiattribute decision analysis under uncertainties. *European Journal of Operational Research, 131*(1), 31–61.

Yang, J. B., Liu, J., Wang, J., Sii, H. S., & Wang, H. W. (2006). Belief rule-base inference methodology using the evidential reasoning approach-RIMER. *IEEE Transaction on Systems, Man, and Cybernetics-Part A: Systems and Humans, 37*(4), 569–585.

Yang, J. B., Liu, J., Xu, D. L., Wang, J., & Wang, H. W. (2007). Optimization models for training belief-rule-based systems. *IEEE Transactions on Systems Man and Cybernetics-Part A: Systems and Humans, 37*(4), 569–585.

Yang, J. B., Wang, Y. M., Xu, D. L., Chin, K. S., & Chatton, L. (2012). Belief rule-based methodology for mapping consumer preferences and setting product targets. *Expert Systems with Applications, 39*(5), 4749–4759.

Yang, J. B., & Xu, D. L. (2002). On the evidential reasoning algorithm for multiple attribute decision analysis under uncertainty. *IEEE Transactions on Systems, Man, and Cybernetics -Part A: Systems and Humans, 32*(3), 289–304 .

Yang, L. H., Liu, J., Wang, Y. M., & Martínez, L. (2019). A micro-extended belief rule-based system for Big Data multiclass classification problems. *IEEE transactions on systems, man, and cybernetics: Systems*. In Press.

Yang, L. H., Wang, Y. M., Chang, L. L., & Fu, Y. G. (2017a). A disjunctive belief rule-based expert system for bridge risk assessment with dynamic parameter optimization model. *Computers & Industrial Engineering, 113*, 459–474.

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

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

Yang, L. H., Wang, Y. M., Su, Q., Fu, Y. G., & Chin, K. S. (2016a). Multi-attribute search framework for optimizing extended belief rule-based systems. *Information Sciences, 370-371*, 159–183.

Yang, Y., Fu, C., Chen, Y. W., Xu, D. L., & Yang, S. L. (2016b). A belief rule based expert system for predicting consumer preference in new product development. *Knowledge-Based Systems, 94*, 105–113.

Yu, X. D., Huang, D. X., Jiang, Y. H., & Jin, Y. H. (2012). Iterative learning belief rule-base inference methodology using evidential reasoning for delayed coking unit. *Control Engineering Practice, 20*(10), 1005–1015.

Zhou, Z. G., Liu, F., Li, L. L., Jiao, L. C., Zhou, Z. J., & Yang, J. B. (2015). A cooperative belief rule based decision support system for lymph node metastasis diagnosis in gastric cancer. *Knowledge-Based Systems, 85*, 62–70.

Zhou, Z. H. (2012). *Ensemble methods - Foundations and algorithms*. Taylor & Francis.

Zhou, Z. J., Hu, C. H., Yang, J. B., Xu, D. L., Chen, M. Y., & Zhou, D. H. (2010). A sequential learning algorithm for online constructing belief-rule-based systems. *Expert Systems with Applications, 37*(2), 1790–1799.

Zhou, Z. J., Hu, G. Y., Zhang, B. C., Hu, C. H., Zhou, Z. G., & Qiao, P. L. (2018). A model for hidden behavior prediction of complex systems based on BRB and power set. *IEEE Transactions on Systems, Man, and Cybernetics: Systems, 48*(9), 1649–1655.