**Ensemble Belief Rule-Based Model for complex system classification and prediction**

Yaqian Youa, Jianbin Suna, Yu-wang Chenb, Caiyun Niua, Jiang Jianga,\*

a. College of Systems Engineering, National University of Defense Technology

109 Deya Road, Kaifu District, Changsha, Hunan, 410073

b. Alliance Manchester Business School, The University of Manchester

Manchester M15 6PB, UK

E-mail:{youyaqian13, sunjianbin, niucaiyun14, jiangjiangnudt }@nudt.edu.cn, yu-wang.chen@manchester.ac.uk

Abstract

Belief Rule-Based (BRB) model has been widely used for complex system classification and prediction. However, excessive antecedent attributes will cause the combinatorial explosion problem, which restricts the applicability of the BRB model to high-dimensional problems. In this paper, we propose an Ensemble-BRB model with the use of the bagging framework to downsize the belief rule base and avoid the combinatorial explosion problem. The kernel of the Ensemble-BRB model is to generate several weak BRBs orderly, each of which only consists of a subset of antecedent attributes. Different combination methods can be used to integrate these weak BRBs coherently for classification and prediction respectively. Four benchmark problems are tested to validate the efficiency of the proposed Ensemble-BRB model in classification, and a real case on the health index prediction of engines proves the feasibility of the Ensemble-BRB model in prediction. The results on both classification and prediction show that the Ensemble-BRB model can effectively downsize the BRB as well as reach a high modeling accuracy.

***Key words:*** BRB model; Bagging framework; Ensemble learning; Classification; Prediction.

1. Introduction

Classification and prediction have been widely implemented for complex system modeling and analysis in engineering design [1], disease diagnosis [2] and image processing [3] and other practical applications. Thus in consideration of the system characteristics, including nonlinearity, uncertainty and complexity, many approaches have been proposed for complex system classification and prediction, such as support vector machine (SVM) [4] [5] [6], fuzzy rule based (FRB) system [7] [8].

SVM is a classical machine learning method for classification and prediction, which focuses on minimizing empirical and expected risks to realize the classification and regression. The modeling of SVM need to determine an appropriate kernel function with parameter adjusting, which requires specific knowledge of the field. The computational complexity of SVM depends on the number of support vectors rather than the dimension of the sample space, which is suitable for high-dimensional classification and regression of complex systems. The mechanism of SVM theoretically guarantees its excellent generalization performance [9]. However, SVM only considers the processing of quantitative data under certain conditions, which make it incapable of modeling both qualitative information and quantitative data under mixed uncertainties.

FRB system is a combination of fuzzy sets and knowledge-based systems [10], which can be categorized in terms of different consequent parts into Mamdani fuzzy system, TSK fuzzy system, etc. [11]. As the FRB systems take the fuzzy sets to describe uncertain information, the knowledge representation ability is mainly concerned with the circumstance in which fuzziness is taken into account in representing uncertain knowledge [12]. The belief rule-based (BRB) model is built in a similar form of the FRB model. Traditional “IF-THEN” rules [17] are extended to belief rules by embedding belief degrees in the consequents of rules [18], which provides an informative scheme of formulating expert experience, uncertain knowledge and hybrid information. The evidential reasoning (ER) algorithm is used aggregate the activated belief rules with the activation of both qualitative and quantitative information. Both the FRB and the BRB are essentially a rule-based system suitable for describing and processing uncertain information, but the forms and inference methods of rules are different. The BRB model extends the “IF-THEN” rules to take into consideration both qualitative/linguistic and quantitative/numerical information under uncertainty and incompleteness, which also differs from the FRB with mere linguistic terms. Since introduced, the BRB model has been successful applied in the fields about classification and prediction, such as medical diagnosis [2], risk analysis [19], preference prediction [20] and so on.

There are two key advantages of the BRB model in classification and prediction compared with other models: first, the modeling and reasoning processes based on “IF-THEN” rules are traceable, which can incorporate experts’ knowledge for further updating; Second, both qualitative and quantitative information under uncertainty and incompleteness in the modeling process can be effectively integrated. However, when constructing a BRB model, it is necessary to define a set of reference values for each of the antecedent attributes. As a result, the combination of multiple reference values over multiple attributes may cause the combinatorial explosion problem. Currently, two main methods have been included in constructing BRB: knowledge-based approach and data-driven approach. The knowledge-based approach directly gives the belief rules based on experts’ expertise and knowledge, which is suitable for the problems lack of historic data. The data-driven approach determines the belief rules from historic data based on optimal learning algorithms, as proposed by Yang et al. [21]. It determines the parameters of the BRB model by minimizing the error (e.g., measured by MSE or MAE) between the predicted and actual values of the training dataset. So far, there are many optimization techniques employed in BRB parameter learning, such as particle swarm optimization (PSO) [22], gradient descent method [23], and differential evolution (DE) algorithm [24]. However, excessive rules will increase the time complexity of optimization. A BRB model with too many rules is almost impossible to be constructed either by experts or using historic data. Therefore, solving the combinatorial explosion problem is a central topic to further promote the application and development of the BRB methodology.

In literature, a great deal of research has been conducted to address the combinatorial explosion problem of BRB models. Among these, two perspectives have been demonstrated as feasible: (I) Reduction of antecedent attributes. The reduction of antecedent attributes is a classical BRB structure learning method, which reduces the belief rules by selecting a subset of key attributes from all antecedent attributes. Common antecedent attributes reduction methods include principal component analysis (PCA) [25], grey target (GT) [25], rough set theory [25], [26], etc. When the degree of influence on the consequent part varies significantly across different antecedent attributes, the antecedent attributes reduction method can effectively identify the core attributes and ensure the BRB modeling accuracy. However, when the influence degrees on the consequent part are similar, this method will cause the information loss and the accuracy reduction. (II) Constructing disjunctive BRB. Traditional belief rules are constructed based on the conjunctive assumption, where the attributes are connected by the “and” operator. The disjunctive BRB transforms the conjunctive assumption into the disjunctive assumption, where the attributes are connected by the "or" operator. The disjunctive assumption between attributes can effectively reduce the number of belief rules and avoid the problem of combinatorial explosion. However, some practical situations require that the antecedent attributes should be connected based on the conjunctive assumption. For example, in the hepatitis B virus examination, only when the five indicators, including hepatitis B surface antigen (HBsAg), hepatitis B surface antibody (HBsAb), hepatitis B e antigen (HBeAg), hepatitis B e antibody (HBeAb), and hepatitis B core antibody (HBcAb), are abnormal, a patient can be judged as the hepatitis B virus carrier. The logical constraints among the attributes limit the applicability of the disjunctive BRB in practice. Therefore, the disjunctive BRB is more suitable as a complement of the conjunctive BRB, rather than the rule reduction method. Other methods to downsize the BRB such as belief k-means (BKM) clustering algorithm and DBSCAN algorithm are introduced in [27] and [28] with details.

In order to deal with the combinatorial explosion problem and reduce the modeling complexity of the BRB, an Ensemble-BRB model based on the bootstrap aggregating (bagging) is proposed in this paper. Ensemble learning has been combined with FRB in some work, especially the boosting of fuzzy rules or classifiers [29], [30]. Literature [31] has mentioned the idea of combining BRB with ensemble learning framework in order to improve the performance of the BRB model. Literature [31] simply integrates multiple BRB models without deep exploration of the potential of ensemble learning framework to solve the combinatorial explosion problem of the BRB, in which the antecedent attributes of each weak BRB model are the same, and the differences between weak BRBs are only derived from different training sets. Bagging is an ensemble learning framework for generating multiple versions of weak learners and using them to get an integrated learner [32], [33], and it has been widely used in fields such as demand forecasting [34], classification [35], [36] and clustering [37]. This paper combines the bagging framework with the BRB model, called Ensemble-BRB, to deal with the combinatorial explosion problem as well as keep the accuracy of the BRB model. The Ensemble-BRB model is a combination of several weak BRB models. The *m* () attributes in a weak BRB model are randomly selected from all *M* antecedent attributes, and *m* is directly determined by the modelers. The increase of *m* can improve the modeling accuracy of weak BRB as well as increase the cost of modeling. Therefore, the setting of *m* requires a trade-off between modeling accuracy and modeling cost, and it can usually be set as *m*=2 or *m*=3 in the practical applications. These weak BRBs obtain the conclusion from different attributes respectively and their ensemble is equivalent to analyze the original information from multiple perspectives. Although the inference results of each weak BRB are not accurate enough, their ensemble can reach satisfactorily accurate results. The proposed Ensemble-BRB model can avoid the combinatorial explosion problem under the premise of ensuring modeling accuracy.

The rest of this paper is organized as follows: Section 2 briefly introduces the concepts of the BRB model and ensemble learning. Section 3 describes the modeling steps of the Ensemble-BRB model based on bagging. Four benchmarks and a health evaluation problem are analyzed in Section 4 respectively to illustrate the implementation process and validate the efficiency of the proposed Ensemble-BRB model in classification and prediction. Finally, some concluding remarks are presented in Section 5.

2. Preliminary

In this section, the basic knowledge of the BRB model and ensemble learning are briefly introduced as the foundation of this study. Further details on them can be found in references [1] and [32] respectively.

2.1 BRB model

(1) Basics of BRB model

The BRB model consists of a set of belief rules, and the *k*-th rule can be described as follows [1]:

 (1)

where  represent *M* antecedent attributes and  indicate the reference values of attributes in the *k*-th rule. Traditional BRB model is based on the conjunctive assumption and the attributes are directly connected by the “” operator.  denote the consequents with the corresponding belief degrees .

The activation weights  of each belief rule are determined by the input data, which can be calculated by Eqs. (2) and (3):

 (2)

 (3)

where  represents the normalized attribute weight,  denotes the matching degree of the input for the *i*-th attribute belongs to its *j*-th referential value  in the *k*-th rule,  denotes the input of the *i*-th attribute. Once is formed, the *k*-th rule is activated.

The inference of the activated belief rules is aggregated by the analytical ER algorithm using Eq. (4) [1], [21]:

 (4)

where denotes the belief degrees of the corresponding consequents .  is the activation weight of the *k*-th rule.  is an intermediate variable for normalization.

(2) Application of the BRB model

The BRB model has a range of applications in classification and prediction. In this section, we take the benchmark *Iris* dataset in the University of California at Irvine (UCI) database, to illustrate the application of the BRB model in classification. This benchmark is aimed to determine the tare of flower (i.e., Setosa, Versicolour, and Virginica) based on four characteristics, which include calyx length, calyx width, petal length, and petal width. These four characteristics correspond to the antecedent attributes in the BRB model, while the tare of flower corresponds to the consequence of the BRB model, as follows:





Take a BRB model contains three rules as an example:





Suppose that the characteristic values of an iris are known as:



The matching degrees  of the input to the attribute reference values are calculated according to Eq. (3):







Under the premise that  and , the activation weights of rules are calculated according to Eq. (2):



According to Eq. (4), calculate the matching degrees of the input to categories separately:



This result indicates that the belief of input X belongs to *Setosa* is 0.9884, the belief belongs to *Versicolour* is 0.0016, and the belief belongs to *Virginica* is 0.01. Obviously, the category *Setosa* has the greatest belief, so we can draw a conclusion that X belongs to the category *Setosa*.

The application of the BRB model in prediction is similar to in classification. Firstly, divide the object to be predicted into *N* levels in order to build a BRB model. Then, calculate the matching degrees of the input to each level. Finally, usethe utility function (given in Eq. (5)) to calculate predicted value to the input.

 (5)

where  represents the reference value corresponding to the *i*-th level.

(3) Parameter learning

As shown in Eq. (1), there are a number of parameters in the BRB model, such as the reference values of attributes, the attribute weights, the rule weights and the belief degrees of the consequents. Currently, two main methods have been included in BRB construction: the knowledge-based approach and the data-driven approach as discussed previously. Compared with the knowledge-based approach of constructing BRB, the data-driven approach is more objective, which can ensure the BRB model statistically reliable. The BRB parameter learning is a data-driven approach to construct and optimize the BRB, which was proposed by Yang et al. [22]. The parameter optimization of the BRB model is essentially an optimization problem with constraints shown in Eq. (6), which determines the parameters of the BRB model by minimizing the error, such as the mean square error (MSE) between the predicted and actual values of the training dataset.





 (6)





where  and  represent the lower and upper bounds of the referenced values of the *m*-th antecedent attribute respectively.

So far, there are many optimization techniques employed in BRB parameter learning, such as particle swarm optimization (PSO) [24], gradient descent method [25], and differential evolution (DE) algorithm [26].Among these, DE algorithm has a faster evolving speed and fewer parameters [17]. The DE algorithm involves a group of heuristic search algorithms. The parameters in the belief rule base are encoded by a floating-point vector to form an individual in the population, which are optimized through crossover, mutation and selection. The parameter learning steps of the BRB model based on DE are given as follows:

Step 1: Population initialization and coding.

Randomly generate *N* individuals meet the constraints, each individual is encoded as shown in Fig. 1:



Fig.1 Individual coding

Step 2: Calculate individual fitness.

Decode the individual into the corresponding BRB model, the MSE of the inference results and the labeling results of samples in the training set is calculated as the fitness of each individual. The smaller the fitness, the better the performance of the individual.

Step 3: Determine whether the current generation *g* is greater than the preset termination generation *G*. If satisfied, decode the best individual in current population as the learned BRB model; otherwise, go to Step 4.

Step 4: Mutation.

Randomly select three individuals in the g-th generation, denotes as , calculate the mutation vector as follow:

 （7）

where *F* represents the mutation probability.

Step5: Crossover.

Randomly select two individuals  and  in the population to generate new individuals as follow:

 （8）

where  represents the value of the r-th place of the new individual,  represents the value of the r-th place of individual  in the g-th generation, *cr* denotes the cross probability.

Step 6: Selection.

Calculate the fitness of the new individual  and the original individual . Determine the individual in the next generation according to Eq. (9), and the generation *g:=g+1*.

 （9）

(4) Combinatorial explosion

The construction of the BRB model needs to traverse all the reference values of all antecedent attributes, and the size of the BRB model (i.e., the number of rules) is determined by the number of antecedent attributes and the number of reference values for each antecedent attribute. For a BRB model with *M* antecedent attributes and the *i*-th antecedent attribute has  reference values, there are totally  rules to be constructed. Obviously, when there are numbers of antecedent attributes in a BRB model, the number of rules in this BRB will increase exponentially, which will result in the combinatorial explosion problem. Therefore, reducing the number of antecedent attributes is an effective way to avoid the combinatorial explosion problems.

However, the antecedent attributes directly affect the modeling accuracy of the BRB model. Although reducing the number of antecedent attributes can effectively reduce the modeling complexity, the reliability of the BRB model may be affected. If the requirednumber of rules in BRB model can be changed from exponential growth to linear growth, the problem of combinatorial explosion can be effectively avoided. Based on this idea, the ensemble learning framework is introduced into the BRB model. The main idea of integrated learning will be presented in Section 2.2.

2.2 Ensemble learning

Ensemble learning is a machine learning method that exploits a certain combination methods to integrate multiple weak learners to obtain an ensemble learner with higher prediction and\or classification accuracy. Bagging is a classical framework in ensemble learning, and its principle is described in Fig. 2:



Fig. 2: Principle of bagging

The bagging framework can be divided into 3 phases:

(1) Random Sampling: The samples in the training sets are randomly selected from the base dataset with replacement. The uncollected samples for each training set are called out-of-bag (OOB) samples, which are not used in the training process and can be used to test the generalization accuracy of the corresponding weak learner. Each training set is different due to the randomness of sampling. This sampling strategy not only ensures the uniqueness of each training set, but also ensures the applicability of bagging in the small base dataset.

(2) Training: Different classifiers or prediction models, such as decision trees and neural networks, are introduced as the weak learners based on specific problems. Each weak learner is independent from each other. A weak learnertrained by the corresponding training set does not need to have high classification/prediction accuracy.

(3) Combination methods [38]: The combination methods refer to the strategies of integrating the classification or prediction results of the multiple weak learners asthe final result, including:

* Voting: For classification problems, we usually use the voting strategy. Among the classification results of all weak learners, the most predicted category is chosen as the final result of the ensemble learner.
* Weighted averaging: For numerical prediction problems, the commonly used combination method is weighted averaging. Suppose the normalized weight of *l*-th weak learner is , the final result of the ensemble learner can be calculated using as follow:

 (10)

where  denotes the predicted result of the *l-*th weak learner, and *H* denotes the final result after combination.

3. Ensemble-BRB model under bagging framework

This section aims to develop an Ensemble-BRB model under bagging framework. In the proposed Ensemble-BRB model, we define the BRB applied as the weak learner as the weak BRB. The modeling steps of the proposed Ensemble-BRB are described in Section 3.1, and the characteristics of the Ensemble-BRB model are illustrated in Section 3.2.

3.1 Ensemble-BRB modeling steps

The initial input data is regarded as the base dataset, and its total number of samples and attributes are denoted as *P* and *M* respectively. There are three parameters that need to be set in the Ensemble-BRB model: the size of each training sets, the attribute number of each training set and the number of weak BRBs, which are denoted as *p*, *m*, and *L*, respectively. The samples in training sets are randomly selected from the base dataset with replacement.

The modeling of Ensemble-BRB model is described as follows:

 

Fig. 3: Ensemble-BRB modeling steps

Step. 1: Parameter setting:

Set model parameters based on actual situation. The initial number of weak BRBs is set as . When , turn to Step. 2, otherwise, turn to Step. 4.

Step. 2: Weak BRB optimization:

Step. 2.1: Randomly select *m* attributes without replacement as the antecedent attributes of the *l*-th weak BRB;

Step. 2.2: Randomly select *p* samples with replacement as the corresponding training set . Only the chosen antecedent attributes are contained in the training set.

Step. 2.3: Optimize the parameters of the *l*-th weak BRBusing the DE algorithm, which the MSE between the predicted value and the actual values is selected as the minimized objective function. See lecture [24] for details.

Step. 3: Out-of-bag (OOB) error calculation:

Step. 3.1: Calculate the OOB error of the *l*-th weak BRB in the test set, which is denoted as . The test set consists of samples that were not selected in Step. 2.

Step. 3.2: . When , turn to Step. 2; otherwise, turn to Step. 4.

Step. 4: Weak BRB combination: Each weak BRBs can give a classification or prediction result, and use the combining methods described in Section 2.2 to obtain the Ensemble-BRB model result.

For classification problem, the voting method is used to integrate the classification results of weak BRBs, the category with the highest number of votes is regarded as the final forecast.

For prediction problem, give the weight of weak BRBs based on the OOB error, i.e. , where  represents the OOB error of the *l-*th weak BRB model, and  represents the weight of the *l*-th weak BRB model. The weighted averaging method shown in Eq. (10) is utilized to integrate the prediction results of weak BRBs.

3.2 Characteristics

Compared with the traditional BRB model, the characteristics of the Ensemble-BRB model are summarized as follows:

(1) Random selection of attributes

Only parts of attributes in the original base dataset are taken into consideration in a weak BRB, which is the key of the Ensemble-BRB model to deal with the combinatorial explosion problem. The number of attributes in weak BRBs is determined by modelers, so the increase of the antecedent attributes in the base dataset will not affect the number of rules in weak BRBs. Each weak BRB only contains part of the attributes, which will not create the combinatorial explosion problem. Correspondingly, the accuracy of weak BRBs in classification and prediction will be lower compared to the traditional BRB model with all the attributes. The attributes contained in each weak BRB model are derived from the random sampling of all attributes, which ensures the diversity of the weak BRB models. Further, the aggregation of various weak BRB models is equivalent to the aggregation of analysis results from multiple perspectives, which ensures the accuracy and comprehensiveness of the Ensemble-BRB model [33].

(2) Random selection of training sets

Each training set consists of *p* samples, which are randomly selected from the base dataset with replacement. The advantages of this sampling method are summarized as follows: (I) Each weak BRB has an unique training set. Different training set will ensure the difference between a pair of weak BRBs even if they share the same antecedent attributes. (II) When the size of base dataset is small, the size of training set can be set as *p=P*. The random selected with replacement can provide enough training sets as well as ensuring the uniqueness among them. When the base dataset size is large, the training sets only need to contain a small part of the base dataset (i.e. *p*<*P*), which reduces the computational overhead of weak BRB optimization; (III) It is not necessary to take out the test set in BRB optimization specifically to get the unbiased error. The classification\prediction error of each weak BRB measured by the corresponding OOB samples is an unbiased estimate of the test error [33], which is approximated as the k-fold cross-validation.

(3) Ensemble

In order to get a better and more comprehensive classification/prediction results, multiple weak BRBs are combined in the Ensemble-BRB model. Different weak BRBs model the conclusions from different attributes, so their ensemble is equivalent to analyzing the original information from multiple perspectives. Each weak BRB only contains a few attributes, which reduces the accuracy of its classification/prediction result. The combination of weak BRBs ensures the stability and classification\prediction accuracy of the Ensemble-BRB model in different situations. Although the results of each weak BRBs are not accurate enough, their ensemble can achieve the satisfactory accuracy. In brief, the Ensemble-BRB model can avoid the combinatorial explosion problem under the premise of ensuring modeling accuracy.

4. Case study

This section includes some cases to demonstrate the feasibility and effectiveness of the proposed Ensemble-BRB model. In Section 4.1, four benchmarks obtained from the University of California at Irvine (UCI) [39], including *iris*, *wine*, *seed* and *haberman*, are studied to validate the efficiency of the Ensemble-BRB model in classification. In Section 4.2, a case for Prognostic and Health Management (PHM) [40] of turbofan engine is studied to illustrate the feasibility of the Ensemble-BRB model in prediction.

4.1 Applicability of the Ensemble-BRB model in classification

(1) Problem description

The general information about the datasets, including *Iris*, *Wine*, *Seed* and *Haberman*, are shown in Table 1 as follows:

Table 1: General information about Iris/Wine/Seed/Haberman

|  |  |  |  |
| --- | --- | --- | --- |
| Dataset | No. of attributes | No. of classes | No. of samples |
| Iris | 4 | 3 | 150 |
| Wine | 13 | 3 | 178 |
| Seed | 7 | 3 | 210 |
| Haberman | 3 | 2 | 306 |

As illustrated in Section 2.1, a number of  parameters need to be optimized in the BRB model.  indicates the number of reference values contained in the *i*-th attribute, denotes the number of attributes and denotes the ranks of consequent part. Therefore, the reduced number of parameters by the Ensemble-BRB model can be calculated as follow:

 (11)

where *L* denotes the number of weak BRBs and *m* denotes the number of attributes in weak BRBs.

Taking *Wine* dataset as an example to illustrate the principle of the Ensemble-BRB model in dealing with the combinatorial explosion problem, each attribute has four reference values, i.e. , and the consequent part is divided into 5 ranks. There are 402,653,200 () parameters need to be given in the traditional BRB model. It is almost impossible to give so many parameters either through the knowledge-based modeling approach or through the data-driven modeling approach. The Ensemble-BRB model contains 20 weak BRBs and each weak BRB owns 2 antecedent attributes in this case. The Ensemble-BRB model only contains 2,160 () parameters, which can be easily obtained based on the data-driven approach. As a result, the number of parameters in Ensemble-BRB model is significantly lesser than that in the traditional BRB model, especially for multi-attribute datasets.

(2) Modeling steps

This part takes the *Iris* dataset as an example to illustrate the modeling steps of the Ensemble-BRB model in classification. The *Iris* dataset consists of 150 samples with four antecedent attributes, each tare of iris has 50 samples. 10% samples are randomly selected from each tare of sample to form the training set of the Ensemble-BRB model, and the remaining samples are form as the test set.

Step. 1: Parameter setting.

In the *Iris* dataset, the size of the training set and the number of antecedent attributes (*m=4*) of the Ensemble-BRB model are small, the scale of the training set for weak BRB model is set to *p=P*, and the number of antecedent attributes of the weak BRB is m=2. The number of weak BRB models is set to *L=20* provisionally to observe the modeling accuracy (If the modeling accuracy is too low, increase the number of *L*, otherwise remain *L=20*).

Step. 2: Weak BRB optimization.

Take the first weak BRB model, denoted as weak BRB1 as an example:

Step. 2.1: Randomly selected 2 antecedent attributes from. In weak BRB1, *m3* and *m4* are selected.

Step. 2.2: Randomly selected 30 times from the training set with replacement to form the training set for weak BRB1, denoted as *Tr1*. The unselected samples in the training set form the test set for weak BRB1, denoted as *Te1*. Notice that the samples in only contain values of attributes *m3* and *m4*.

Step. 2.3: The DE algorithm is used to optimize the parameters of the weak BRB1 model. See Section 2.1(3) for details.

Step. 3: Repeat step 2 to build a total of 20 weak BRB models. Noe that the classification problem uses the voting method to integrate weak BRB models, so there is no need to calculate the OOB error. When applied to the prediction problem, the OOB error of each weak BRB model need to be calculated at this step.

Step. 4: Weak BRB combination.

Each weak BRB model classifies the samples in the test set independently, see Section 2.1(2) for detail. The voting method given in Section 2.2 is used to integrate the classification results of the 20 weak BRB models, the category with the highest number of votes is regarded as the final tare of iris.

(2) Numerical analysis

The division of training set and test set for the above four benchmarks are as follows: part of the samples in the base dataset are randomly selected as the training set of the Ensemble-BRB model, and the remaining samples are regarded as the test set. The size of the training set is recorded as *P*. The *l*-th weak BRB model randomly selected *P* samples from the training set as their own training set, recorded as *Trl*, seen Fig. 3 for detail. Each weak BRB model has the different training set.

For each benchmark, we set the size of antecedent attributes of each weak BRB model to 2, i.e., m=2.The differential evolution (DE) algorithm described in Section 2.1 is used to optimize the weak BRB models, in which the population size is set to 20 individuals and the iteration number is set to 500 generations. 20 weak BRB models are trained for the Ensemble-BRB model, and the classification accuracy of weak BRBs and the Ensemble-BRBs on four benchmarks are shown Fig. 4. The blue lines marked by squares indicate the classification accuracy for each weak BRB (from the first weak BRB to the 20-th weak BRB), while the orange lines marked by dots indicate the classification accuracy of the Ensemble-BRB models after the boosting of integrated weak BRBs (For example, the 10th orange dot represents the Ensemble-BRB model who integrates the first 10 weak BRBs). The X-axis indicates the number of weak BRBs in the Ensemble-BRB model, while the Y-axis indicates classification accuracy.



Fig. 4: Relationship between classification accuracy and weak BRB number

As shown in Fig. 4, the classification accuracy curves of weak BRBs are in random fluctuation, and the Ensemble-BRB curves are fluctuation with rising trend. In the early stage of ensemble, the addition of weak BRBs can significantly improve the classification accuracy of the Ensemble-BRB. With the increase of the integrated weak BRBs, the classification accuracy of the Ensemble-BRB model tends to be stable. The Iris and Haberman datasets have fewer attributes than other datasets, and the classification accuracy of weak BRB models are closer to their ensemble results. As a result, the increase of the integrated weak BRBs has less improvement to the Ensemble-BRB modeling accuracy. Intuitively, the slope of the orange lines in Fig.4(a) and Fig.4(d) are smaller than that in Fig.4(b) and Fig.4(c). This phenomenon illustrates that the more antecedent attributes in the base dataset, the better improvement of the classification accuracy of the Ensemble-BRB model with the increase of the integrated weak BRB. Base dataset with more antecedent attributes can highlight the advantages and characteristics of the Ensemble-BRB model. The comparison of classification accuracy in test set between the weak BRBs and the Ensemble-BRB is shown in Table. 2:

Table 2: Classification accuracy of Iris/Wine/Seed/Haberman

|  |  |  |
| --- | --- | --- |
| Dataset | Mean accuracy of weak BRBs | Accuracy of Ensemble-BRB |
| Iris | 0.9182 | 0.9778 |
| Wine | 0.7547 | 0.9888 |
| Seed | 0.8540 | 0.9381 |
| Haberman | 0.7270 | 0.7614 |

As shown in Table. 2, for each dataset, the classification accuracy of the Ensemble-BRB model is much higher than the weak BRBs. Each weak BRB classifies the samples only through some attributes of the original data, so its classification is not accurate. The Ensemble-BRB model is equivalent to analyze samples from multiple perspectives. Although the analysis results of each weak BRB are not accurate enough, the ensemble of them can reach a satisfactory accuracy.

(3) Comparative research

To verify the feasibility of the proposed Ensemble-BRB model, nine classifiers are taken as the comparison models, including naïve Bayes [41], Bayes net [42], decision tree learner (DT) [43], one nearest neighbor (1-NN) [44], Dempster’s combination (DC) rule-based classifier [45], ~~and~~ ER-based classifier [46], Fuzzy granular gravitational clustering algorithm (FGGCA) [47], Difference of Convex functions Algorithm (DCA) [48], and GA-based SVM [49]. Table 3 presents the classification accuracy of the Ensemble-BRB model together with the other classifiers in these four benchmarks. Spaces shown with “–” signify that results are not given in the corresponding lecture, and values in BOLD represent the best classification accuracy for each dataset.

Table 3: The classification accuracy of different classifiers [46], [47], [48], [49]

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Dataset** | **Iris** | **Wine** | **Seed** | **Haberman** |
| naïve Bayes | 0.9333 | 0.9718 | 0.881 | 0.7623 |
| Bayes net | 0.9167 | 0.9859 | 0.9048 | 0.7787 |
| DT | 0.9167 | 0.8592 | 0.881 | 0.7377 |
| 1-NN | 0.9 | 0.9437 | 0.869 | 0.6537 |
| DC | 0.9667 | 0.9069 | 0.9048 | **0.8000** |
| ER | 0.9633 | 0.9783 | 0.8956 | 0.7424 |
| FGGCA | 0.9722 | 0.9666 | **0.9557** | 0.7620 |
| IP-DCA-VNS | 0.9670 | 0.9320 | - | - |
| TPMSVM+GA | - | - | - | 0.7289 |
| Ensemble-BRB | **0.9778** | **0.9888** | 0.9381 | 0.7614 |

For *Iris* and *Wine* datasets, the Ensemble-BRB has the highest accuracy. For *Seed* and *Haberman* dataset, although the accuracy of the Ensemble-BRB model seems less satisfactory, it still reach the second and the fifth highest accuracy respectively, compared to other studies. The comparison verifies that the Ensemble-BRB model proposed in this paper can produce satisfactory results in classification.

4.2 Applicability of the Ensemble-BRB model in prediction

(1) Background

Engine is a core component of many large equipment, and its health status is critical to the normal operation of equipment. The Prognostic and Health Management (PHM) aims at improving the reliability, safety and availability of devices, which is a collection of health status prediction, fault diagnosis, and so on. Developing a reasonable PHM strategy is significant to improve the reliability of the engine and thus extending its service life.

In the field of PHM, the health status of engine is often measured by the *Health Index* (HI). HI is a continuous value defined on [0,1]. *HI=1* means the device is in a normal state, *HI=0* means the engine is completely destroyed, and  denotes the performance of engine is reduced.

Due to the following reasons, the health status of equipment is hard to be measured directly: (I) The operation of the equipment does not allow frequent shutdown to be measured; (II) In the early stage of performance degradation, the damage is often in the micro range, which is difficult to be found and measured. (III) Without damaging the equipment, some internal failures of complex parts are difficult to find. A variety of status monitoring signals contain a lot of health status information are captured from the running device through sensors, which are useful to estimate the health status of the device in real time. Excavating the health status (i.e., HI) of equipment from monitoring signals is of great significance for guiding the operation and maintenance of equipment.

Currently, there are multiple calculation methods for HI. This paper selects the Wang's [115] Mahalanobis distance-based calculation method, which realizes the calculation of HI by fusing the Mahalanobis distance of multiple parameters to its normal state. The calculation formula is as follows:

 (12)

where,  indicates the health index of the device at the *i*-th time,  is the vector of all monitored parameters at the *i*-th time, and  is the vector of all parameter values when the device is out of function.

In this section, an Ensemble-BRB model is built to predict the HI during engine operation. The difference between the predicted HI and the actual HI is utilized to verify the effectiveness of the Ensemble-BRB model in prediction problems. This case focuses on the HI prediction of engine and studies the applicability of Ensemble-BRB model in prediction.

The engine dataset utilized in this section is obtained from the Commercial Modular Aero-propulsion System Simulation (C-MAPSS) large turbofan engine simulation model [50], [51], which collects 21 parameters including the Low Pressure Compressors (LPC), High Pressure Turbines (HPT), et al. for 100 engines with different initial attrition over their life cycle. The descriptions of these parameters are shown in Table A1 in *Appendix*. A set of parameters are recorded as a sample when an engine turns a round, and there are 20631 samples in the base dataset.

Notice that Parameters 1, 5, 6, 10, 16, 18, and 19 are constants in the dataset, which will not affect the health status of the engines [53]. Therefore, these parameters are ignored in this case, and the remaining 14 parameters are regarded as the antecedent attributes of the Ensemble-BRB model in the health status prediction. The health status is affected by these performance parameters, which have been studied in many literatures [52], [53]. The HI calculated by Eq. (6) is regarded as the consequent part to construct the Ensemble-BRB model.

(2) Parameter setting

As mentioned in Section III. A, there are three parameters in the modeling of Ensemble-BRB: the size of each training set, the attribute number of each training set and the number of weak BRBs. There is no general setting method for these parameters yet, which needs to be set by decision makers according to the practical applications.

In this case, 20% of the samples are randomly selected as the training set of the Ensemble-BRB model, the number of samples in the training set is recorded as *P*. The remaining 80% samples are regarded as the test set. Each weak BRB model independently repeats P-time random sampling with replacement on the training set to construct its own training set, following the process in Fig. 2.

Excessive attribute number (denoted as *m*) will cause the combinational explosion problem for the weak BRB models, and *m* is usually set to a small number in the practical applications. In order to find a suitable *m*, the following experiment is designed. Three weak BRB models with *m=2*, *m=3* and *m=4* are respectively constructed based on the same training set. The modeling accuracy is measured by the OOB error and test set error, which is measured by MSE through Eq. (13):

 (13)

To avoid contingency, the experiment is repeated 10 times and the results are shown in Table 4:

Table 4: OOB error, test set error and time cost of weak BRBs

|  |  |  |  |
| --- | --- | --- | --- |
| Index | OOB error | Test set error | Time cost (s) |
| m | 2 | 3 | 4 | 2 | 3 | 4 | 2 | 3 | 4 |
| Set1 | 0.0202 | 0.0154 | 0.0119 | 0.0236 | 0.0198 | 0.0122 | 460.92 | 692.15 | 1079.3 |
| Set2 | 0.0182 | 0.018 | 0.0115 | 0.0225 | 0.0259 | 0.0156 | 461.81 | 688.24 | 1080.3 |
| Set3 | 0.0194 | 0.0146 | 0.0113 | 0.0226 | 0.0188 | 0.016 | 461.58 | 696.07 | 1069.2 |
| Set4 | 0.0188 | 0.0162 | 0.0146 | 0.022 | 0.0228 | 0.0178 | 466.26 | 689.16 | 1069.6 |
| Set5 | 0.0211 | 0.021 | 0.0144 | 0.025 | 0.0242 | 0.017 | 464.39 | 682.22 | 1073.6 |
| Set6 | 0.0278 | 0.0139 | 0.0154 | 0.0338 | 0.0191 | 0.0193 | 465.62 | 691.77 | 1072.8 |
| Set7 | 0.0157 | 0.0144 | 0.019 | 0.0199 | 0.0191 | 0.0208 | 463.47 | 691.98 | 1083.1 |
| Set8 | 0.0205 | 0.0132 | 0.016 | 0.0253 | 0.0193 | 0.0204 | 462.68 | 687.57 | 1065.8 |
| Set9 | 0.0229 | 0.0203 | 0.0116 | 0.0268 | 0.0239 | 0.0205 | 470.34 | 693.19 | 1066.3 |
| Set10 | 0.0292 | 0.017 | 0.0251 | 0.0284 | 0.0211 | 0.0459 | 468.26 | 690.75 | 1065 |
| average | 0.02138 | 0.0164 | 0.01508 | 0.02499 | 0.0214 | 0.02055 | 464.533 | 690.31 | 1072.5 |

Although the mean error of the weak BRBs are slightly lower when *m=3* and *m=4*, the time costs are greatly increased. Considering the model error and time cost, we take *m=2* in this case.

The number of weak BRBs affect the modeling accuracy of the Ensemble-BRB model. When *L* is small, the more integrated weak BRBs, the higher accuracy of the Ensemble-BRB model. With the increase of the integrated weak BRBs, the accuracy increase rate of the Ensemble-BRB model will decreases and converges. At this time, integrating more of weak BRBs has little significance for improving the accuracy of the Ensemble-BRB model. Therefore, *L* only needs to be greater than the convergence threshold of the modeling accuracy of the Ensemble-BRB model, which is determined by the decision makers according to the practical application. In this case, we set *L=20*, which can guarantee the convergence of the modeling accuracy of the Ensemble-BRB model.

As mention above, set *m*=2 and L=20 to construct the Ensemble-BRB model in this case. Each weak BRB model is optimized by the DE algorithm detailed in Section 2.1, where the population size is set to 50 individuals and the iteration number is set to 200 generations. Fig. 5 shows the performance of the Ensemble-BRB model for HI prediction on the sample set.

(3) Numerical analysis

 Fig. 5: Predicted HI of Ensemble-BRB model

As shown in Fig.5 (a), the green line represents the actual value of HI, and the yellow line represents the predicted value of the Ensemble-BRB model, which indicates that the Ensemble-BRB model can accurately predict the HI of the engines. It is difficult for the Ensemble-BRB model to give accurate predictions when the samples are near the end or near the initial stage of engine life. The failure of engine is the result of the interaction of various attributes. However, each weak BRB only predicts HI through two features, which reduce the accuracy of weak BRBs and in turn affects the effective of the Ensemble-BRB model in the initial and final stages of engine life. Without loss of the generality, we take the first weak BRB as an example to compares the prediction capacity of weak BRB and Ensemble-BRB in Fig. 5 (b). The first weak BRB model has a large fluctuation range, and the Ensemble-BRB model prediction curve is relatively stable. This is because only two attributes are considered to predict the HI in the weak BRB model. An incorrect measurement of an attribute will exert great influence on the consequent, which leads to a large fluctuation of the weak BRB prediction result. The Ensemble-BRB contains multiple antecedent attributes. The incorrect measurement of an attribute does not have a great impact on the result, so its prediction result is relatively stable. Fig. 5(c) shows the MSE values of the Ensemble-BRB model and each weak BRB, their trends are similar to the trends shown in Fig. 4. In the initial stage of ensemble, the MSE decreases with the increase of integrated weak BRBs. When the integration of weak BRBs reaches a certain amount, the MSE of the Ensemble-BRB model tends to be stable.

(4) Comparative research

The original intention of the Ensemble-BRB model is to deal with the combinatorial explosion problem existing in the traditional BRB model. In order to verify the effectiveness of the Ensemble-BRB model in rule reduction, this paper compares the Ensemble-BRB model with two extended BRB models which are developed for solving the combinatorial explosion problem (i.e., PCA-BRB and intersection BRB). In order to ensure the fairness of the comparison, these three models share the same training set and are trained by DE algorithm. Under the assumption that each attribute has four reference values, Table 5 shows the BRB sizes of different BRB models and their MSE values on the test set.

Table. 5: Comparison with other BRB size reduction model

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | Traditional BRB | PCA-BRB | Disjunctive BRB  | Ensemble-BRB |
| MSE | -- | 0.0226 | 0.1613 | **0.0145** |
| No. of rules | 414 | 16 | **4** | 16 |

The traditional BRB model includes 414 belief rules, which is almost impossible to be determined in practice. So we fail to build the traditional BRB model in this case due to the excessive number of rules and its MSE value is not given in Table 5. The three models can effectively reduce the BRB size and avoid the combinatorial explosion problems. The disjunctive BRB model contains only 4 belief rules in this case, which is convenient in practical applications. But its MSE value is higher than the PCA-BRB model and the Ensemble-BRB model. This may be because that the health condition of the engine is decided by the interaction of the parameters, so that the attributes are suitable to be connected with “and” operator. The disjunctive assumption is not applicable in this case. The PCA-BRB model extracts two principal components from all attributes as antecedent attributes through the PCA method, which downsize the rules into . The PCA-BRB model ignores some information during the dimension reduction process, resulting in lower prediction accuracy than the Ensemble-BRB model. Compared with these models, the Ensemble-BRB model can guarantee the prediction accuracy as well as reduce the number of rules in this case.

To further verify the effectiveness of the proposed Ensemble-BRB model in prediction, the MSE is compared with five classical prediction model, named Support Vector Machines (SVM) [54], Linear Regression (LR) [55], k-Nearest Neighbor (KNN) [56], Decision Tree (DT) [57] and Random Forest (RF) [33]. In order to ensure the fairness of the comparison, the following models share the same training set. The MSEs measure the error between the predicted HI and the actual HI on the same test set. The results are shown in Table 6:

Table. 6: Comparison with SVM/LR/KNN/DT/RF

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model | SVM | LR | KNN | DT | RF | Ensemble-BRB |
| MSE | 0.0277 | 0.0192 | 0.0264 | 0.0257 | 0.0266 | **0.0145** |

All these prediction models have achieved relatively accurate prediction results. The MSE of the Ensemble-BRB model is only 0.0145, which is less than other prediction models. The MSE value illustrates that the Ensemble-BRB model has higher prediction accuracy than traditional prediction models such as SVM/LR/KNN/DT and classical integrated model like RF, which verifies that the Ensemble-BRB model can achieve satisfactory prediction results in this case.

5. Conclusion

In this study, an Ensemble-BRB model is proposed to deal with the combinatorial explosion problem in traditional BRB models. The Ensemble-BRB model consists of several weak BRBs through using the bagging framework. The weak BRBs with different antecedent attributes and training sets are further optimized by the DE algorithm. The antecedent attributes in each weak BRB are randomly selected from all attributes without replacement, while the samples in the training sets are also randomly selected from the base datasets with replacement. The voting and the weighted averaging are used as the combination methods for classification and prediction respectively.

Four benchmark classification problems, namely *Iris*, *Wine*, *Seed* and *Haberman*, were used to validate the efficiency of the proposed Ensemble-BRB model in classification, and a real case about the engine HI prediction proved the feasibility of the Ensemble-BRB model in prediction. The comparison with other classical models indicated that the Ensemble-BRB model has the ability to solve both classification and prediction problems with high accuracy, and the comparison with other BRB dimension reduction methods illustrated that it can effectively reduce the number of rules as well as keep a high modeling accuracy. In summary, the Ensemble-BRB model can effectively improve the applicability of the BRB methodology in practice

In future research, the relationships between modeling accuracy and the number of weak BRBs, modeling accuracy and the number of attributes in each weak BRB will be taken into consideration. Considering that the current time complexity of the Ensemble-BRB modeling is relatively high, the next step can focus on reducing the modeling complexity.

Reference

1. Koziel S, Leifsson L. Shape-Preserving Response Prediction for Surrogate Modeling and Engineering Design Optimization [J]. IEEE Transactions on Microwave Theory & Techniques, 2010, 58(11):2829-2837.
2. Kong G , Xu D L , Yang J B , et al. Belief rule-based inference for predicting trauma outcome[J]. Knowledge-Based Systems, 95(C): 35-44, 2016.
3. Zhi-Jian F, Yang-Geng F U, Jian-Hua C. Belief Rule Base Inference for Texture Image Classification [J]. Journal of Applied Sciences, 2017.
4. Tunchan Cura. Use of support vector machines with a parallel local search algorithm for data classification and feature selection [J], Expert systems with applications. 145 (2020) 113133.
5. Park J, Kwon I H, Kim S S, et al. Spline regression based feature extraction for semiconductor process fault detection using support vector machine[J]. Expert Systems with Applications, 2011, 38(5):5711-5718.
6. Lloyd B G R. Support Vector Machines for classification and regression[J]. Analyst, 2010, 135(2):230-267.
7. Chanak P, Banerjee I. Fuzzy rule-based faulty node classification and management scheme for large scale wireless sensor networks[J]. Expert Systems with Applications, 2016, 45(MAR.):307-321.
8. Pham B T , Tien Bui D , Prakash I , et al. Rotation forest fuzzy rule-based classifier ensemble for spatial prediction of landslides using GIS[J]. Natural Hazards, 2016, 83(1):1-31.
9. Khodabandelou G, Ebadzadeh M M. Fuzzy neural network with support vector-based learning for classification and regression[J]. Soft Computing, 2019, 23(2):1-16.
10. Rodríguez-Fdez, I, Mucientes M , Bugarín, A. FRULER: Fuzzy Rule Learning through Evolution for Regression[J]. Information Sciences, 2016, 354:1-18.
11. Ishibuchi H . Effect of rule weights in fuzzy rule-based classification systems[J]. Fuzzy Systems IEEE Transactions on, 2001, 9(4):506-515.
12. Liu, Jun, et al. Fuzzy rule-based evidential reasoning approach for safety analysis. International Journal of General Systems 33.2-3 (2004): 183-204.
13. Yang J B, Liu J, Wang J, et al. Belief rule-base inference mechanism using the evidential reasoning Approach-RIMER [J]. IEEE Transactions on Systems Man & Cybernetics Part A Systems & Humans, 2006, 36(2):266-285.
14. Jiao L, Denoeux T, Quan P. A Hybrid Belief Rule-Based Classification System Based on Uncertain Training Data and Expert Knowledge [J]. IEEE Transactions on Systems Man & Cybernetics Systems, 2016, PP (99):1-13.
15. Yang-Geng F U, Long-Hao Y, Ying-Jie W U. Evidential Reasoning Approach for Solving Complex Evaluation Models [J]. Pattern Recognition and Artificial Intelligence, 2014.
16. Dempster A P. A generalization of Bayesian inference [J]. Journal of the Royal Statistical Society, 1968, 30(2):205-247.
17. Hayes-Roth, Frederick. Rule-based systems [J]. Communications of the ACM, 1985, 28(9):921-932.
18. Liu J, Martinez L, Calzada A, et al. A novel belief rule base representation, generation and its inference methodology [J]. Knowledge-Based Systems, 2013, 53:129-141.
19. Yang L H, Wang Y M, Chang L L, et al. A Disjunctive Belief Rule-Based Expert System for Bridge Risk Assessment with Dynamic Parameter Optimization Model [J]. Computers & Industrial Engineering, 2017:S0360835217304448.
20. Wang Y M, Yang J B, Xu D L, et al. Consumer preference prediction by using a hybrid evidential reasoning and belief rule-based methodology[J]. Expert Systems with Applications, 2009, 36(4):8421-8430.
21. Yang J B, Liu J, Xu D L, et al. Optimization Models for Training Belief-Rule-Based Systems[J]. IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, 2007, 37(4):569-585.
22. Qian B, Wang Q Q, Hu R, et al. An effective soft computing technology based on belief-rule-base and particle swarm optimization for tipping paper permeability measurement [J]. Journal of Ambient Intelligence & Humanized Computing, 2017(15):1-10.
23. Chang Rui, Wang Hongwei, Yang Jianbo. An algorithm for training parameters in belief rule-bases based on the gradient and dichotomy methods [J]. Systems Engineering, 2007, 25(S): 287-291.
24. Chang L, Zhou Z J, You Y, et al. Belief rule based expert system for classification problems with new rule activation and weight calculation procedures [J]. Information Sciences, 2016, 336(C):75-91.
25. Chang L, Zhou Y, Jiang J, et al. Structure learning for belief rule base expert system: A comparative study [J]. Knowledge-Based Systems, 2013, 39(none):159---172.
26. Wang Y M , Yang L H , Fu Y G , et al. Dynamic rule adjustment approach for optimizing belief rule-base expert system [J]. Knowledge-Based Systems, 2016, 96(C):40-60.
27. Li B, Wang H, Yang J, et al. A belief-rule-based inference method for aggregate production planning under uncertainty [J]. International Journal of Production Research, 2013, 51(1):83-105.
28. Longhao Y, Yanggeng F U, Yingjie W U. Structure Learning Approach of Belief Rule Base for Best Decision Structure [J]. Journal of Frontiers of Computer Science & Technology, 2014.
29. Cordón, Oscar, Arnaud Quirin, and Luciano Sánchez. A first study on bagging fuzzy rule-based classification systems with multicriteria genetic selection of the component classifiers. 2008 3rd International Workshop on Genetic and Evolving Systems. IEEE, 2008.
30. Nguyen T T , Nguyen M P , Pham X C , et al. Heterogeneous Classifier Ensemble with Fuzzy Rule-based Meta Learner[J]. Information Sciences, 2017:S0020025517301548.
31. Weikun W U , Yanggeng F U , Qun S U , et al. GDA Based Ensemble Learning Methods for Parameter Training in Belief Rule Base [J]. Journal of Frontiers of Computer Science and Technology, 2016, 10(12):1651-1661.
32. Breiman L. Bagging predictors [J]. Machine Learning, 1996, 24(2):123-140.
33. Cutler A, Cutler D R, Stevens J R. Random forests [J]. Machine Learning, 2004, 45(1):157-176.
34. Oliveira E M D, Oliveira F L C. Forecasting mid-long term electric energy consumption through bagging ARIMA and exponential smoothing methods [J]. Energy, 2018, 144:776-788.
35. Truong C, Zhang M, Andreae P, et al. Bagging and Feature Selection for Classification with Incomplete Data [C]// Evostar 2017. 2017.
36. Bauer E, Kohavi R. An Empirical Comparison of Voting Classification Algorithms: Bagging, Boosting, and Variants [J]. Machine Learning, 1999, 36(1-2):105-139.
37. Dudoit S, Fridlyand J. Bagging to improve the accuracy of a clustering procedure [J]. Bioinformatics, 2003, 19(9):1090-9.
38. Zhou Z H. Ensemble Methods - Foundations and Algorithms [M]. Taylor & Francis, 2012.
39. UCI repository of machine learning database, http://www.ics.uci.edu/∼learn/MLRepository.html/.
40. Vichare N M, Pecht M G. Prognostics and health management of electronics [J]. IEEE Transactions on Components & Packaging Technologies, 2006, 29(1):222-229.
41. I. Rish. An empirical study of the naive Bayes classifier, in: Proceedings of the IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence, 3, Morgan Kaufmann, 2001, pp. 41–46 .
42. N. Friedman, D. Geiger, M. Goldszmidt. Bayesian network classifiers, Mach. Learn. 29 (2–3) (1997) 131–163.
43. Y. Freund, L. Mason. The alternating decision tree learning algorithm, in: Pro- ceedings of the ICML, 99, ACM, 1999, pp. 124–133.
44. T. Cover, P. Hart. Nearest neighbor pattern classification, IEEE Trans. Inform. Theory 13 (1) (1967) 21–27.
45. C. Zhang , Y. Hu , F.T.S. Chan , R. Sadiq , Y. Deng , A new method to determine basic probability assignment using core samples, Know.-Based Syst. 69 (1) (2014) 140–149 .
46. Xu X, Zheng J, Yang J B, et al. Data Classification using Evidence Reasoning Rule [J]. Knowledge-Based Systems, 2016, 116.
47. Sanchez M A , Castillo O , Castro J R , et al. Fuzzy granular gravitational clustering algorithm for multivariate data [J]. Information Sciences, 2014, 279:498-511.
48. An L T H, Minh L H, Tao P D. New and efficient DCA based algorithms for minimum sum-of-squares clustering [J]. Pattern Recognition, 2014, 47(1):388-401.
49. Wang Z, Shao Y H, Wu T R. A GA-based model selection for smooth twin parametric-margin support vector machine [J]. Pattern Recognition, 2013, 46(8):2267-2277.

1. <https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/#phm08_challenge>.
2. Anderson S. Damage propagation modeling for aircraft engine run-to-failure simulation [C]// International Conference on Prognostics & Health Management. 2008.
3. Malhotra P, Tv V, Ramakrishnan A, et al. Multi-Sensor Prognostics using an Unsupervised Health Index based on LSTM Encoder-Decoder [J]. 2016.
4. Sun J, Zuo H, Wang W, et al. Application of a state space modeling technique to system prognostics based on a health index for condition-based maintenance [J]. Mechanical Systems & Signal Processing, 2012, 28(2):585-596.
5. Saunders C, Stitson M O, Weston J, et al. Support vector machine [J]. Computer Science, 2002, 1(4):1-28.
6. Montgomery D C, Peck E A. Introduction to linear regression analysis [M]. 1982.
7. Phillips P C B, Moon H R. Linear Regression Limit Theory for Nonstationary Panel Data [J]. Econometrica, 1999, 67(5):1057-1111.
8. Fayyad U M, Irani K B. On the Handling of Continuous-Valued Attributes in Decision Tree Generation [J]. Machine Learning, 1992, 8(1):87-102.