



A novel PCA-DC-Bagging algorithm on yield stress prediction of RAFM steel

Sifan Long¹ · Ming Zhao¹ · Jieqiong Song¹

Received: 22 January 2019 / Accepted: 15 May 2019 / Published online: 20 September 2019
© Springer-Verlag GmbH Austria, part of Springer Nature 2019

Abstract

For most regression tasks, we often use an ensemble learning technology of Bagging algorithm. However, the traditional Bagging algorithm is susceptible to extreme values. This leads to high bias and high variance in the prediction process. Therefore, this paper proposes an improved Bagging algorithm based on a best decision Committee model and the idea of selecting the base learner, and we have presented the idea of using the decision-making committee to filter learner, train the decision-making committee by the base learner to classify the error on the test set. Using the optimal interval separation factor's mathematical model which is derived by the Lagrange multiplier method to classify the evaluation levels. The decision committee is trained according to the assigned evaluation level, and the learner is selected and assembled according to the decision result of the decision committee members. Meanwhile, our theoretical analysis shows that there are two different cases, which we can use maximum likelihood estimation and stochastic process theory to build mathematical models for analysis. The analysis results based on reduced activated ferritic/martensitic (RAFM) steel data sets show that the proposed algorithm can be applied to data sets with high dimension, high redundancy, high contradictory samples, sparse data sets, and then, we give the strict theoretical framework to guarantee the further development and promotion. This gives algorithm model.

Keywords Bagging algorithm · Decision Committee · Dimension · Optimization

Mathematics Subject Classification 68T04 Artificial intelligence

This research is supported by National Natural Science Foundation of China under Grant No. 61572526 and the China Institute of Atomic Energy.

✉ Ming Zhao
meanzhao@csu.edu.cn

Extended author information available on the last page of the article

1 Introduction

The reduced activated ferritic/martensitic (RAFM) steel is the main structural material recommended for future magnetically constrained nuclear fusion reactors. Because controllable nuclear fusion has its inherent safety and efficiency, coupled with sufficient energy sources, it become one of the materials used in fusion power stations. Europe, the United States, Japan, Russia and China have developed a variety of RAFM steel, forming a series of steel types, such as CLAM steel, F82H steel, etc. [1, 2].

RAFM steel uses low-irradiation sensual elements W, V and Ta instead of the commonly used high- irradiation living alloy materials Mo, Ni and Nb, which can effectively reduce the radiation-induced radioactivity, radiation swelling and thermal expansion coefficient of the material. It can also increase the high temperature thermal properties of the material as well as the high thermal conductivity. Because of the influence of many factors, it is difficult to make a predictive assessment from the physical mechanism. In recent years, with the rapid development of computer technology, the ability to model with computer-aided technology, and the establishment of reliable predictive models to study the performance of RAFM steel is an emerging research tool and technology.

Bagging algorithm has strong robustness, generalization and high variance reduction. It has a wide range of practical applications in dealing with multi-classification problems and regression problems [3]. For example, Binh used the Bagging algorithm based on SVM as a base learner to study the landslide problem [4]. Huang combined the Bagging algorithm of genetic algorithm to solve the fault diagnosis of transformer [5]. Yang and Jiang used the hybrid Sampling-Based clustering ensemble with global and local constitutions [6]. At the same time, their research on adaptive bi-weighting toward automatic initialization and model selection for HMM-Based hybrid meta-clustering ensembles [7]. The Bagging algorithm has many variant-derived algorithms, such as attribute resampling algorithms (Attribute Bagging). In the traditional Bagging algorithm, extra-package estimation is not used in bootstrap sampling [8], therefore, it loses the a priori information provided by the verification set. Through the evaluation of the decision-making committee, we can select the base learners with superior performance on the verification set, and then assemble the best base learners to get a hybrid prediction model with better prediction effect than the single learner. This is also the principle that ensemble learning can often show better performance, because the process of selecting base learners is similar to that of voting by committee, so we call this method of selecting model as decision-making committees. The main contributions of this paper are as follows:

1. This work analyzes the traditional Bagging algorithm which is affected by the extreme values of the base learner. We use the prior information of the verification set to assist the decision-making process, and put out the idea that the decision-making committee carefully screens the learner.
2. The concept of interval separation factor is proposed. The error evaluation criteria of basic learning unit are given, and then, we use the Lagrange multiplier optimization theory to find out the optimal interval separation factor and gives a strict mathematical proof.

3. In view of actual case, this paper gives some characteristics of the decision-making committee, and analyze the two cases in which the model exists from the theory. For case 1, it has been found that the maximum likelihood estimation method can be used to analyze the property of the model; for case 2, it is found that the property of the model can be analyzed theoretically by a stochastic process. Finally, combining the two cases, we present the ML-RP evaluation algorithm.
4. For a sparse, highly redundant, multi-repetitive sample data set with many outliers (not limited to the RAFM steel studied in this paper), we present a general algorithm model. As long as the preconditions are met, it can use the algorithm model for research and analysis, or only use the DC-Bagging algorithm of the decision committee.

2 Related work

In this section, we survey the related work on RAFM steel and ensemble learning. RAFM steel is a kind of material used in future fusion power plant, but its service conditions are very harsh, such as requiring it to be exposed to high temperature, high pressure, high irradiation and high corrosive experimental conditions for a long time. In addition, the material selection of fusion power plants is closely related to the safe run of nuclear power plants. Therefore, it is necessary to study the characteristics of fusion materials to ensure the promotion of large-scale engineering applications in the future.

At present, RAFM steel is mostly studied and analyzed from the physical mechanism. For example, Vijayanand et al. studied on the microstructure evolution of electron beam welds under creep loading [9], and microstructural evolution in creep tested electron beam welded Reduced Activation Ferritic Martensitic (RAFM) steel and 316LN stainless steel dissimilar weld joints has been studied at 823 K under different stress levels. In addition, Laha et al. studied the effects of tungsten and tantalum contents on impact, tensile, low cycle fatigue and creep properties of Reduced Activation Ferritic-Martensitic (RAFM) steel were studied to develop India-specific RAFM steel [10], and their research found that the RAFM steel having 1.4 wt% tungsten with 0.06 vol% tantalum was found to possess optimum combination of impact, tensile, low cycle fatigue and creep properties and was considered for Indian-specific RAFM steel. Mao et al. studied the correlation among microstructural parameter and dynamic strain aging (DSA) in influencing the mechanical properties of a reduced activation ferritic-martensitic (RAFM) steel [11], and then the contributions of these microstructural parameters on the tensile properties at elevated temperatures were studied with the modified Crussard–Jaoul (C–J) analysis based on the Swift equation. In addition to the above studies, there are few related studies on the application of machine learning in nuclear fusion materials, so the focus of our work is to carry out cross-cutting research in the above two fields.

In the field of machine learning, we take regression task as an example. There are many classical prediction and analysis algorithms, such as artificial neural network model, support vector machine model, decision tree model and random forest model. These models are all single learners (we call them “base learners”). When dealing with very complex experimental data, such as high dimensions, many contradictory samples

and many outliers, we discuss the defects exposed by a single learner in subsequent chapters. Therefore, in order to solve this problem, we use the technology of ensemble learning to overcome this obstacle [12]. After repeated experiments, we finally chose the Bagging algorithm as the blueprint, combined with the decision-making committee prediction algorithm to get the best results.

3 Preliminary study

In this section, we will introduce some basic research, including the data set we used, the PCA algorithm for dimensionality reduction, and the traditional bagging algorithm.

3.1 Database

The data set used in this paper is the result of irradiation experiments. For related information, refer to “Appendix”, which described the nonlinear relationship between yield strength and experimental conditions (e.g. irradiation temperature, irradiation dose and test temperature etc.) and element content (e.g. Cu, Fe and S etc.). Elemental content is given by the material company, the corresponding experimental conditions are measured in the laboratory, the associated attribute set involves 37 attributes. The corresponding statistics are given: maximum, minimum, average, variance and standard deviation. These statistics can reflect the distribution of data to a certain extent. It can be concluded that the biggest feature of data distribution is sparsity.

3.2 Principal component analysis

The experimental data used in this study have the following characteristics.

1. There are many associated attributes, reaching 37 dimensions.
2. Data distribution is sparse and it is difficult to integrate information.
3. There are many outliers which interfere with the constructed model.

Based on the above challenges, this study will use principal component analysis (PCA) technique to effectively alleviate the above difficulties [13]. It can search for m pieces of n -dimensional orthogonal vectors that best represent the original data, where $m \leq n$. Therefore, the original data is projected onto a small space, which achieves the purpose of data dimensionality reduction, data cleaning, and noise removal. After data dimensionality reduction, the unrelated, weakly correlated, and redundant attributes and dimensions are detected and deleted. High-dimensional attributes are successfully dimension-reduced to obtain new data sets with low noise, low-dimensional, high-information, and then we can retrain the model on the new data set.

3.3 Bagging algorithm

In machine learning, whether it is discriminant analysis or regression analysis, the commonly used algorithms are single learners, such as neural network model [14], LS-SVM [15, 16], Decision tree, etc. [17, 18]. However, it can use an ensemble learning strategy to obtain a strong learner with superior performance. A good learner requires

good predictive ability and diversity. In other words, in order to make more accurate predictive ability after the integration of multiple base learners, each base learner needs to have good performance and diversity. For example, for a data set, we train five base learners. We need to predict a data point that we given in advance is 120. Then, if five learners are good, the output of the learner is 118, 118, 118, 118, 118 (if we use a arithmetic average as the output after integration). The output of ensemble learning is 118. If our learners are more diverse, and then the output is 118, 121, 119, 122, 120, then the output of ensemble learning is 120. Of course, 120 is more accurate than 118. Based on the principle described above, Bryll proposed an improved Bagging algorithm [19], the voting model is selected by random feature subset, and then a strong learner is selected according to the voting result, thereby we assembled a predictive model of the strong learner.

The Bagging algorithm is one of the most representative algorithms in the parallel ensemble learning method. In the model training process, it is necessary to obtain a learner with superior performance and diversity. First, self-sampling or bootstrap sampling is used on the generated training data set. For a given data set D that contains M samples, in order to produce a discrepant dataset D' , m samples are randomly selected from the overall sample in each time, and the process is repeated m times. From this, it can be concluded that after m sampling is completed, a new subset containing the total number of samples m is obtained as D' . It can be seen that a part of the samples in D will appear multiple times in the subset D' . Another part of the sample will not appear, so the probability that the sample was taken in m samples is:

$$\left(1 - \frac{1}{m}\right)^m \quad (1)$$

Therefore, taking the limit, we have:

$$\lim_{m \rightarrow \infty} \left(1 - \frac{1}{m}\right)^m \rightarrow \frac{1}{e} \approx 0.368 \quad (2)$$

The (1) and (2) equation shows that about 63.2% of the data in training of the model, and 36.8% of the data does not participate in the waste of the data set. Therefore, the algorithm performs out-of-package application of data sets that are not involved in model training, which ensures maximum use of the data set, which is critical for areas with high data acquisition costs (e.g. the nuclear material field).

4 PCA-DC-Bagging algorithm

This section mainly introduces the Bagging-derived algorithm based on principal component analysis and decision-making committee, which is named PCA-DC-Bagging algorithm. In order to effectively avoid the high dimension and the influence of noise, redundancy, repetition, and contradiction sample data characteristics, so we need to find more robust algorithms to meet these challenges. Therefore, this paper proposes a PCA-DC-Bagging algorithm based on the PCA dimension reduction technology combined with the decision-making committee's Bagging algorithm.

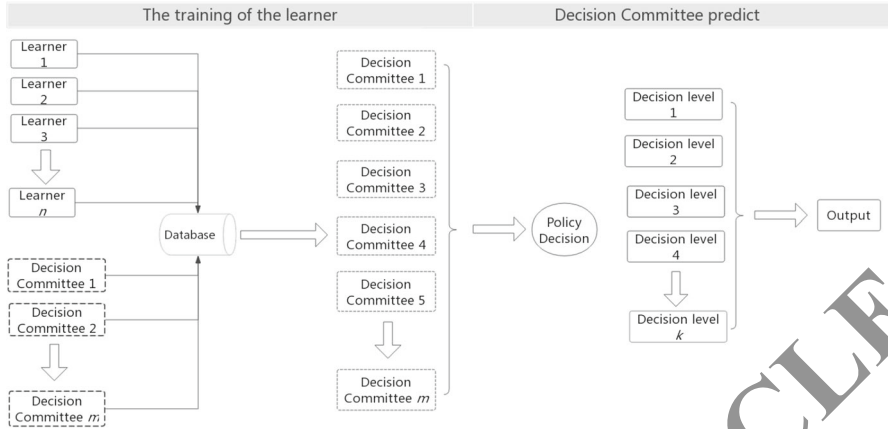


Fig. 1 The main schematic chart of the PCA-DC-Bagging algorithm

4.1 PCA-DC-Bagging algorithm principle

It can be seen from the analysis in the previous section that to solve the problem that the prediction result of the learner in the Bagging algorithm is easily affected by extreme values, how to select the learner in a targeted manner is the key to improve the performance of the algorithm. Therefore, this study gives an efficient filtering of learners based on the decision-making committee model. That is the PCA-DC-Bagging algorithm. The principle of the algorithm is that in the training of the learner, bootstrap sampling is used to filter the data set, then m pieces of base learners are trained on the filtered training set, and a decision committee is established on the verification set for error evaluation. The details of the algorithm can be referred to Fig. 1 and Algorithm 1.

In the process of training the base learner, the base learner is trained on the total data set D . The data set D is divided into three parts, a training data set D_{train} , a verification set $D_{verification}$, and a test set D_{test} . First, m data points are randomly selected as the test set, and then the remaining data sets are divided into training set D_{train} and test set D_{test} according to the bootstrap sampling method (the data set ratio is 63.2%: 36.8%) [20]. Combining the data set of this study, this paper takes the experimental data of RAFM steel as an example. At this time, data set D contains 1811 pieces of data, which divides data set D , and 100 pieces of data are used as test set D_{test} , 1711 data for bootstrap sampling, reference to the partition of bootstrap about data sets in "Appendix", training set D_{train} about 63.2% of the data for the training of the learner (about 1080 data points), and about 36.8% of the verification set $D_{verification}$ for model performance filtering. Figure 2 shows the process that the data set is divided.

In Fig. 2, m pieces of data is randomly selected as the test set, and then the data in the remaining training set is divided into the training set D_{train} by the bootstrap sampling method to train model. And the set $D_{verification}$ is used to filter models to select the best base learners.

It has been concluded from the previous analysis that in order to improve the predictive performance of the learner after the integrated algorithm, the proposed algorithm focuses on more accurate selection of the base learner than the traditional

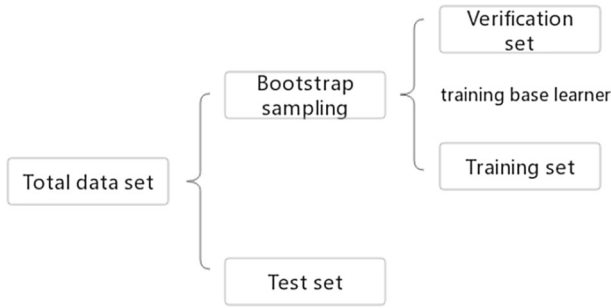


Fig. 2 Division scheme of data sets

Bagging algorithm which does not do any processing on the base learner. Therefore, the PCA-DC-Bagging algorithm is presented in the form of pseudo code, please refer to Algorithm 1. The schematic chart of the PCA-DC-Bagging algorithm is given Fig. 1. Some specific properties of the model are given below.

Algorithm 1 PCA-DC-Bagging Algorithm

Require: High-dimensional data sets, cumulative contribution rates after PCA dimensionality reduction, decision-making committee member models (using classic classification algorithms such as neural networks, decision trees, SVM, random forests).

Ensure: Given the number of learners m , standard deviation L .

Execute PCA algorithm; //Dimensional processing of high-dimensional data;

Random sampling method to extract test set data D_{test} ;

for $i = 1$ to m do

 Bootstrap sampling, divide data into training data set D_{train} , verification set $D_{verification}$;

 Train the corresponding base learner $\{learner_1, learner_2, learner_3, \dots, learner_m\}$ on the

training set D_{train} ;

 The model is verified on the verification set $D_{validation}$, and the mean square error is obtained.

 The model is divided and a new mark set $D_{result-validation}$ is obtained;

 Train the decision-making committee model on the data set $D_{result-validation}$ and get the decision-making member model of the decision-making committee as $DC_1, DC_2, DC_3, \dots, DC_m$;

end for

repeat

 // Forecast on the test set.

 The decision committee makes decisions on the data set and achieve the corresponding decision level $l_1, l_2, l_3, \dots, l_i$;

 Follow the ranking rules and select the top h higher level learners for prediction;

 Reassemble the learner $\{learner_1, learner_2, learner_3, \dots, learner_m\} (h < m)$ by a simple average strategy. Make predictions;

 //PCA-DC-Bagging algorithm ended.

until $j < v$

Definition 1 In the Bagging algorithm, the given base learner $l_1, l_2, l_3, \dots, l_m$. Divides the training set D_{train} and the verification set $D_{validation}$ by bootstrap sampling method. To train m pieces of base learners in the training set D_{train} , and define the mean squared error set $E = \{e_1, e_2, e_3, \dots, e_m\}$ of the base learner on the verification set $D_{validation}$ as the performance error for the m base learners, where $e_k = \sum_{i=1}^n (\hat{y}_i^k - y_i^k)^2$.

Considering the particularity of the bootstrap sampling method in the Bagging algorithm, training decision-making committees on validation sets can be used to evaluate the quality of learners, however, in practical applications, it often requires more abundant prior information to obtain a more accurate evaluation method. Our ultimate goal is to establish a quantitative decision-making committee model to assist selecting the best base learners. In order to increase the decision-making accuracy requirements of the decision-making committee and reduce the forecasting risk after decision-making mistakes, here gives a definition of the error level.

Definition 2 For a given closed interval $[a, b]$, where $a > 0, b \geq 0$. Without loss of generality, we give the separation factor $\xi = \{\xi_1, \xi_2, \dots, \xi_n\}$, and satisfy $\sum_{i=1}^n \xi_i \leq b - a, \xi_i \geq 0$. Then the separation factor divides the interval $[a, b]$ into n subintervals $[a, a + \xi_1], [a + \xi_1, a + \sum_{i=1}^2 \xi_i], \dots, [a + \sum_{i=1}^{n-1} \xi_i, b]$.

Definition 3 For the m pieces of given base learner $l_1, l_2, l_3, \dots, l_m$, the performance error set of the corresponding base learner in the bootstrap sampling process is $E = \{e_1, e_2, e_3, \dots, e_m\}$. The distribution interval of the given error distribution E is $[a, b]$, where $a = \min(E), b = \max(E)$. Given the interval separation factor $\xi = \{\xi_1, \xi_2, \xi_3, \dots, \xi_n\}$, Then the error set E is divided into t error levels, which may be recorded as $L = (l_1, l_2, l_3, \dots, l_t)$, so L is declared as the criterion for judging the error level of the decision committee.

The above sections introduced decision maker on the validation set and defined a quantitative model to evaluate the performance indicators of the learner. Train m learners on the training set D_{train} at the same time, make prediction errors in the verification set $D_{validation}$, and divide t levels into $L = (l_1, l_2, l_3, \dots, l_t)$, according to the criteria of the decision-making committee to judge the error level, on this basis, training the decision-making committee. The general decision-making committee can use some common classifier algorithms, such as BP neural network [21], SVM classification algorithm [22], decision tree [23], naive Bayes classifier [24], random forest algorithm [25], etc. By classifying the level $l_1, l_2, l_3, \dots, l_t$, on the verification set, the new verification set is also marked as the training data set of the model, in order to give the predictive ability to decision-making committee. Train the decision-making committee on the new error result set and obtain m pieces of decision committee prediction model $(DC_1, DC_2, DC_3, \dots, DC_m)$. If the decision-making committee does not play a predictive effect, in the worst case, it is equivalent to random guessing. If the prediction is better than random guessing, the prediction is made on the new data set. For each decision-making member in the decision-making committee to make predication on new data set, the correct probability of decision-making is $\frac{1}{t}$.

In theoretical analysis, it requires to optimize some important parameters. Taking the interval separation factor as an example, if the value is too large, it will increase the tolerance of the decision-making members leading to a decrease in the accuracy of the decision; if the value is too small, it is difficult to construct a decision-making committee that meets the accuracy requirements. The above analysis shows that the value of the interval separation factor has an important influence on the classification error level. Therefore, this article uses the technique of convex optimization theory to establish a quantitative model to optimize the interval separation factor.

Definition 4 For the members of the decision-making committee, given the certain data set, we believe that the interval separation factor $\xi = \{\xi_1, \xi_2, \xi_3, \dots, \xi_n\}$ has a nonlinear multivariate function relationship with the error results. And set it as the committee member loss function $f\{\xi_1, \xi_2, \xi_3, \dots, \xi_n\}$. The decision committee's overall loss function is defined as $\Gamma = \log_2(2 + |f_1| + |f_2| + |f_3| + \dots + |f_n|) - 1$.

In practical applications, the facing problems are complex. For different data sets, the loss function given by Definition 4 is different. So, it can build specific models in other ways, such as polynomial fitting models, least squares fitting models, etc. [26, 27], this paper doesn't do more discussion about this. This paper establishes the nonlinear function relationship between the interval separation factor and the decision committee error by Definition 4. The next step is to optimize the loss function by the convex optimization theory to find the corresponding value of the interval separation factor $\xi = \{\xi_1, \xi_2, \xi_3, \dots, \xi_n\}$ when the loss function $f\{\xi_1, \xi_2, \xi_3, \dots, \xi_n\}$ is minimum [28]. It uses the Lagrange multiplier method to get the results below [29].

Theorem 1 Assuming Loss function $f_j, g_i, h_i : \Psi^n \rightarrow \Psi (i = 1, 2, \dots, j = 1, 2, \dots, t)$ is continuous differentiable. The decision-making committee overall loss function satisfies the following formula:

$$\min \Gamma = \log_2(2 + |f_1| + |f_2| + |f_3| + \dots + |f_m|) - 1 \tag{3}$$

$$s.t. g_i(\xi) \geq 0, h_i(\xi) = 0 \tag{4}$$

Then, there is a Lagrange multiplier vector $\lambda^* \in \Psi^s, \mu^* \in \Psi^t$, which make the first-order optimality condition for the committee's overall loss function Γ , that is, the KKT condition is established:

$$\begin{cases} \nabla_{\xi} \Gamma(\xi^*, \lambda^*, \mu^*) = 0, \\ h_j(\xi^*) = 0, j \in E, \\ \lambda_i^* \geq 0, g_i(\xi^*) \geq 0, \\ \lambda_i^* g_i(\xi^*) = 0 \end{cases} \tag{5}$$

Proof Referring to Kolmogorov existence theorem [30], it can be seen that the interval is separated by the factor ξ , and assuming that the linearization feasible direction (LFD) is the same as the serialization feasible direction (SFD), that is, $SFD(\xi^*, D) = LFD(\xi^*, D)$. Then, for any $d \in LFD(\xi^*, D)$, there is $d^T \nabla f(\xi^*) \geq 0$, the Lagrange

multiplier vector $\lambda^* \geq 0, \mu_i^*, i \in I(\xi^i), j \in E$ must exist, so that the following formula holds $\nabla_{\xi} \Gamma(\xi^*, \lambda^*, \mu^*) \nabla f(\xi^*) - \sum_{i \in I(\xi^*)} \lambda_i^* \nabla g_i(\xi^*) - \sum_{j \in E} \mu_j^* \nabla h_j(\xi^*) = 0$, make $\lambda_i^* = 0, \forall i \in I \setminus I(\xi^*)$, the conclusion of Theorem 1 can be obtained.

When given the interval separation factor ξ Theorem 1 is actually a special case of the Lagrange multiplier method for solving the optimization problem with the smallest overall loss function of the decision-making committee. Through mathematical modeling, the problem of interval separation factor selection in decision-making is transformed into convex optimization problem. Then, using the Lagrange multiplier method to find the corresponding first-order optimal condition under the constraints of the minimum objective function (here, the overall loss function Γ), namely the KKT condition [29], and find the optimal interval separation factor ξ sequence. The above is the research method given from the theory.

5 Analysis of the model

The PCA-DC-Bagging algorithm model given in this study belongs to a hybrid model [31]. In order to evaluate the quality of the model and choose the appropriate combination strategy based on the performance of the base learner, this paper theoretically gives the mathematical model of the evaluation method, and finally gives the ML-RP evaluation algorithm of the evaluation model.

Definition 5 During the model training process, the correct probability of the i -th committee member of the decision-making committee $DC = \{dc_1, dc_2, dc_3, \dots, dc_m\}$ is $p_i = p\{l_i \neq \hat{l}_i\}$, l_i is the level decision of the i -th committee member, \hat{l}_i is the actual level of the base learner. At this time, the correct probability $p = (p_1, p_2, p_3, \dots, p_m)$ of the judges of the m committee members is called the probability of successful judgment by the decision committee.

The basic assumption is that the performance of the basic learning device is stable. It is a prerequisite for defining 5. In practical applications, it is not always satisfied with this basic assumption. Therefore, the quality of the final judgment will depend on the degree of violation of this basic assumption. Therefore, It required to establish a quantitative model based on the basic concept of the above definition and the final result. In theory, there are two kinds of cases.

Case 1 When the number of samples is large, the weak learner is sufficient for the sample learning, and under the condition of repeated experiments, the probability $p = (p_1, p_2, p_3, \dots, p_m)$ of the committee members is stable. Therefore, there is no significant difference in the results of repeated iterations.

Case 2 When the number of samples is scarce, the prediction error of the basic learning device is relatively large. Under repeated experiments, the probability P of the members of the Committee showed a significant fluctuation. Therefore, how to analyze the quality of models in repeated experiments is the key to maximize the prediction results.

Below we classify and discuss some characteristics of the model in two cases, and explore the inherent regularity characteristics of the algorithm from the theoret-

ical point of view. Therefore, the following two theorems are given for two different situations:

Theorem 2 *In the process of repeated experiments, if the prediction accuracy of the decision members is stable. Let the decision committee members judge the correct probability as the random variable D , and D obeys the uniform distribution on $[s, t]$, where s, t is unknown, $d_1, d_2, d_3, \dots, d_m$ are the observations of the random variable D . Then the maximum likelihood estimator of s, t is:*

$$\hat{s} = \min_{1 \leq i \leq m} D_i, \hat{t} = \max_{1 \leq i \leq m} D_i \tag{6}$$

Proof It is assumed that the probability that the decision-making committee members judge the correct is constant and is recorded as D . According to the results of Fig. 5, it is reasonable to assume that D obeys a uniform distribution on $[s, t]$. And set

$$d_1 = \min\{d_1, d_2, d_3, \dots, d_m\} \tag{7}$$

$$d_m = \max\{d_1, d_2, d_3, \dots, d_m\} \tag{8}$$

Then we have, the probability density of D is:

$$f(d; s, t) = \begin{cases} \frac{1}{t-s}, & s \leq d \leq t, \\ 0, & \text{other.} \end{cases} \tag{9}$$

So the likelihood function can be written as:

$$L(s, t) = \begin{cases} \prod_{i=1}^m \frac{1}{t-s}, & s \leq d_1, d_2, \dots, d_m \leq t, \\ 0, & \text{other.} \end{cases} \tag{10}$$

$$= \begin{cases} \frac{1}{(t-s)^m}, & s \leq d_1, d_m \leq t, \\ 0, & \text{other.} \end{cases} \tag{11}$$

For any given s that satisfies the condition of $s \leq d_1, t \geq d_m$, we have

$$L(s, t) = \frac{1}{(t-s)^m} \leq \frac{1}{(d_m - d_1)^m} \tag{12}$$

It can be seen from the above formula that the necessary and sufficient conditions for the likelihood function $L(s, t)$ to take the maximum value of $(d_m - d_1)^{-m}$ is: If and only if $s = d_1, t = d_m$. Therefore, for the decision committee to satisfy the case 1, the maximum likelihood estimate of s, t is

$$\hat{s} = d_1 = \min_{1 \leq i \leq m} d_i, \hat{t} = d_m = \max_{1 \leq i \leq m} d_i \tag{13}$$

Theorem 2 mainly shows that under the condition of satisfying case 1, the decision-making differences of decision-making committee members are mainly determined by the worst and best decision-making members. Therefore, during the establishment of the decision-making committee, the key to increase the decision-making precision of the decision-making committee members is to improve the accuracy of decision-making as a whole while weakening the difference in decision-making precision of decision-making members. The Theorem 2 gives some intrinsic characteristics that the PCA-DC-Bagging algorithm can satisfy, which can be used as the theoretical basis for the performance analysis of the prediction results. When the theoretical analysis satisfies case 2, the PCA-DC-Bagging algorithm satisfies different characteristics. The Theorem 3 and the practical application in this study are given as follows.

Theorem 3 *During repeated experiments, it is assumed that the prediction accuracy of decision members exhibits an unstable distribution with state T . The probability that the decision of the committee members is correct is the random variable D , and the probability space is denoted as (Ω, \wp, P) . Then, for each given $t \in T$, the decision variable predicts the correct probability of the random variable family $\{D(t, e), t \in T\}$ as a stochastic process on the probability space (Ω, \wp, P) .*

Proof Under the condition of repeated experiments, it is assumed that the probability that the decision member correctly predicts is the random variable D , and D exhibits an unstable distribution with the given number of times $t_1, t_2, t_3, \dots, t_m \in T$. Obviously, for arbitrarily arranged $\{i_2, i_3, \dots, i_m\}$ of the set $1, 2, 3, \dots, m$, and when $k \leq m$ satisfies the lower form

$$\left. \begin{aligned} F_{t_1, \dots, t_m}(d_1, \dots, d_m) &= F_{t_1, \dots, t_m}(d_{i_1}, \dots, d_{i_m}) \\ F_{t_1, \dots, t_k}(d_1, \dots, d_m) &= F_{t_1, \dots, t_m}(d_1, \dots, d_m, \infty, \dots) \end{aligned} \right\} \tag{14}$$

Where F is a family of finite dimensional distribution functions [32]. In the course of the training committee, the training sequence is unaffected. The finite dimensional distribution family F satisfies the above equation to satisfy the symmetry and compatibility. According to the Kolmogorov existence theorem, there must be a probability space (Ω, \wp, P) and a random process $\{D(t, e), t \in T\}$ defined in it, and its finite dimensional distribution function family is F .

Theorem 3 is mainly mathematically modeled by the specific application scenarios of this study combined with the relevant theories of stochastic processes. The decision-making behavior of the committee is described by stochastic process theory, so the

random behavior of the decision-making committee is transformed into a descriptive mathematical model under certain conditions. Then the following metrics are given under the framework of stochastic process theory. For the stochastic process $D_T = \{d(t), t \in T\}$, defined by the decision committee's correct prediction, the following method is given as the evaluation index.

1. The mean function of D_T is $m_D(t) = ED(t), t \in T$.
2. The covariance function of D_T is $B_D(s, t) = E\{[D(s) - m_D(s)][D(t) - m_D(t)]\}, s, m \in T$.
3. The variance function of D_T is $D_D(t) = B_D(t, t) = E[D(t) - m_D(t)]^2, t \in T$.
4. The correlation function of D_T is $R_D(s, t) = E[X(s)X(t)], s, t \in T$

The mean function $m_D(t)$ is the average value of the stochastic process $\{D(t), t \in T\}$ in the t state, so it can be used to describe the average value of the correct prediction probability of the decision committee in the t state. When the forecasting results does not meet the accuracy requirements, the mean function can be used to analyze the model to find a solution to improve the prediction performance. The variance function $D_D(s, t)$ describes the degree of deviation from the mean function $m_D(t)$ in the t state. When the predictive performance of decision-making members shows significant differences, the system should be analyzed by the variance function to find out the problems of the model itself. The covariance function $B_D(s, t)$ and the correlation function $R_D(s, t)$ are the linear correlations of the response in the stochastic process $\{D(t), t \in T\}$ in the s and t states. For example, when training a PCA-DC-Bagging algorithm model, take two states to investigate the system. The covariance function and correlation function can be used to analyze the state of the model at this time, and then evolve to the final state. Model training, if this process can be described theoretically, then a contradictory departure from the irreconcilable over-fitting and under-fitting is a new starting point [33].

The above analysis is the result of meeting the case 2. Because when the probability of a decision-making committee's decision is affected, it can no longer be assumed that the random variable D is approximately stationary, so case 1 doesn't apply any more. Why is there such a problem in practical application? The explanation given in this paper is that in the process of training decision-making committee, the decision-making level according to the prediction effect of the base learner is affected by the bootstrap sampling method of the bagging algorithm. Therefore, the assumptions made are reasonable and have a strong correlation with a given data set. Through the above analysis, the following algorithm can be used to describe two situations. We call to this as the ML-RP evaluation algorithm.

Algorithm 2 ML-RP Algorithm

Require: Sample size of S , committee $d_1, d_2, d_3, \dots, d_m$.

Ensure: Set the sample size is s_0 .

if $s < s_0$ **then**

 //Case 1.

 //Calculating the maximum likelihood estimator.

$\hat{s} \leftarrow \min_{1 \leq i \leq m} d_i$; //Maximum likelihood estimator of S .

$\hat{t} \leftarrow \max_{1 \leq i \leq m} d_i$; //Maximum likelihood estimator of T .

else

 //Case 2.

 //Mean function of D_t .

$m_D(t) \leftarrow ED(t)$;

 //Covariance function of D_t .

$B_D(s, t) \leftarrow E[\{D(s) - m_D(s)\}\{D(t) - m_D(t)\}]$;

 //Variance function of D_t .

$D_D(t) \leftarrow B_D(t, t) = E[D(t) - m_D(t)]$;

 //Related functions of D_t .

$R_D(s, t) \leftarrow E[X(s)X(t)]$;

end if

 Decision analysis, making decisions, ending;

The proposed ML-RP evaluation algorithm can be considered in both case 1 and case 2. As long as the corresponding parameters, such as the sample size, are given, it is possible to analyze which case is more suitable for the data set given by the algorithm, and then use the corresponding statistic for more accurate analysis, and finally approximate the true result.

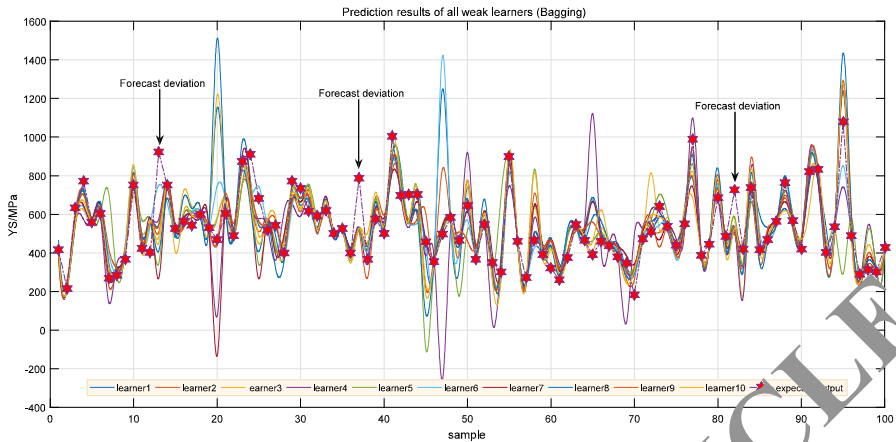


Fig. 3 The prediction effect of all base learners on the test set in the Bagging algorithm

6 Experimental analysis

The previous section theoretically gives the corresponding model evaluation criteria for two different cases. This section mainly conducts comparative analysis of the models by experiments on the RAFM steel data set. Compared with the traditional Bagging algorithm, the PCA-DC-Bagging algorithm has stronger anti-noise ability and more accurate prediction results, and the results of comparative analysis with the model show that. Different models have different application backgrounds, therefore, there will be a significant difference in the prediction effect.

6.1 Problems with the Bagging algorithm

For RAFM steel experimental data to construct a prediction model based on Bagging algorithm, the base learner uses a neural network model. And the network structure of 11-12-1 (optimized by grid search algorithm) is adopted. The traditional Bagging algorithm is used to train the learners, and then the test sets are respectively tested, all the base learner prediction result graphs are drawn as Fig. 3.

Obviously, for most base learners (10 learners trained here), there are only 3 prediction deviation points, so the learner can cover 97% of the experimental data points, which can satisfy most experimental data points and can effectively predict. Figure 3 also shows that the prediction coverage interval of most learners contains real data points; therefore, it gets an important message: as long as the appropriate learner can be selected, and we can use arithmetic averaging as the final output strategy [34], the Bagging algorithm uses an average strategy to fit all the base learners for predictive performance with great potential for improvement. Since the algorithm of arithmetic averaging cannot effectively avoid the extreme value, it is impossible for the learner to make an estimate. This bad situation will greatly deviate from the decision result.

6.2 Simulation

In the RAFM steel dataset, we trained 10 neural network prediction algorithms as the base learners. The member decision model in the decision-making committee uses a random forest classification algorithm to train the decision-making committee according to this algorithm.

Figure 4 represents the error of the learner on the cross-validation set. It can be seen that the prediction effect of a single learner is not good, but the subsequent results show that the prediction effect after integration is better than that of a single learner. We divide the level according to the size of the residual, and finally complete the training of the decision committee. The decision-making committee adopts the random forest classifier algorithm. In the classification level, referring to Fig. 4, the mean square error of the learner on the verification set is calculated by Theorem 1. The division result of the interval separation factor is divided into 6 prediction levels. At this time, the interval separation factor is $\xi = \{7, 20, 32, 18, 44, \infty\}$. Next, we solve the interval separation factor according to the mathematical formula given by the fourth section theory.

Therefore, the corresponding results are as follows:

- (I) *Level 1: 0–7
- (II) *Level 2: 8–27
- (III) *Level 3: 28–59
- (IV) *Level 4: 60–77
- (V) *Level 5: 78–121
- (VI) *Level 6: 122– ∞

According to this division, the decision-making committee compares the level of the test set with the actual situation. The probability of correct division is 42%, 38%, 45%, 36%, 37%, 38%, 47%, 41%, 41%, 40%. Compared with random guessing, the probability of random guessing is about 16% in the case of dividing 6 levels, and the decision-making committee proposed in this paper is 2.8 times of random guessing. The specific experimental results are described below.

Figure 5 is a neural network model, linear regression, traditional Bagging algorithm, and the PCA-L₂-Bagging algorithm given in this study is comparatively analyzed:

1. Use the same data set (this study used the RAFM data set) and divide it into a training set and a test set to train the compared model.
2. Test the model on the test set, compare the predicted result with the actual output actual result, and calculate the goodness of fit [35, 36]. The closer the goodness of fit is to 1, the better the regression effect would be; in the opposite, the worse effect will be.
3. The predicted output and the actual output result are compared to obtain a residual. Perform residual analysis [37, 38], including the mean of the residuals (generally subject to a normal distribution with a mean of 0) [39, 40], variance, and standard deviation. Theoretically, the closer the mean value is to 0, the smaller the variance and standard deviation is, and the better the prediction result is.

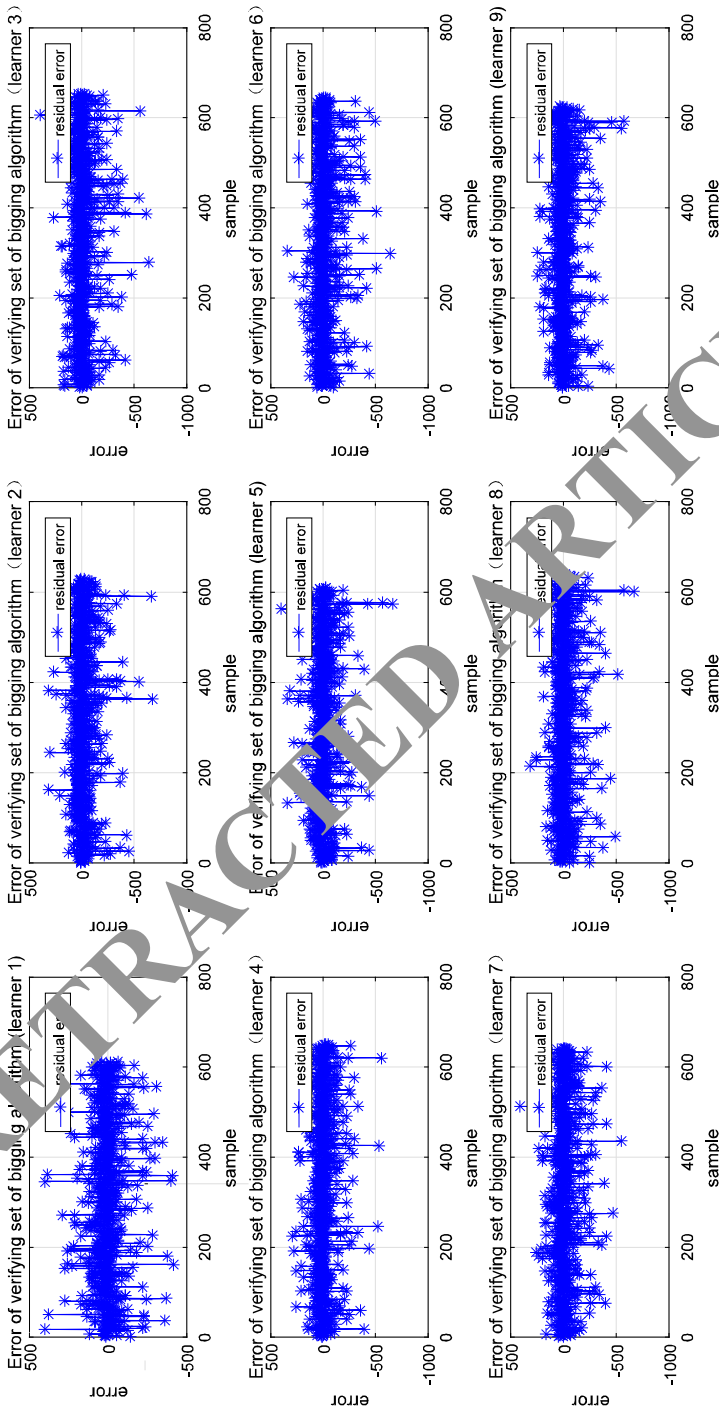
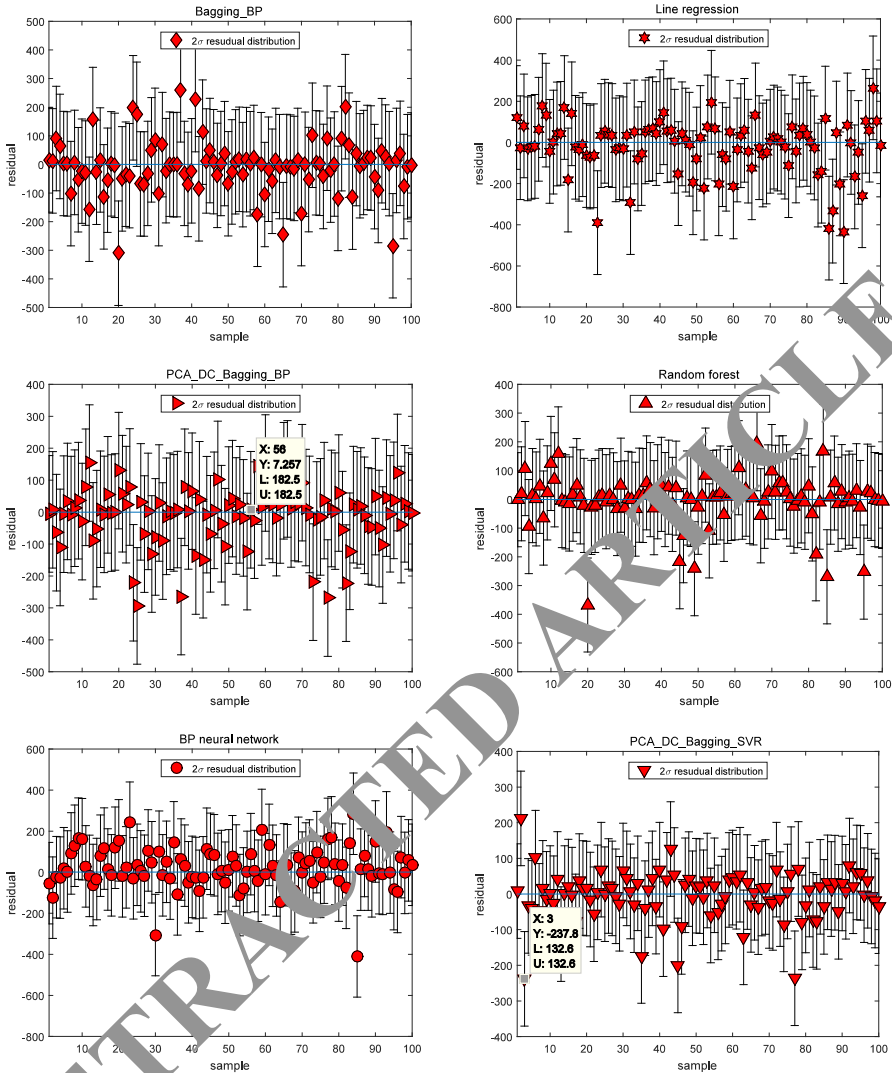


Fig. 4 The error results of the base learner on the verification set, the residuals are used to evaluate the performance of the learner



(a) Comparison of similar algorithms

(b) Comparison of different types of algorithms

Fig. 2 Comparison with BP neural network, generalized regression neural network, linear regression, random forest, BP neural network based framework algorithm, SVR-based framework algorithm

4. Calculate the mean square error by predicting the output and the actual output. The mean square error directly reflects the overall deviation of the prediction result on the test set [41, 42]. Similarly, the smaller the value is, the better the prediction is.

Relevant statistic information is given in “Appendix”. Among them are mean, variance, standard deviation, mean square error, and goodness of fit used to judge regression results. The test results of 100 test data set in Table 1 show that the best

Table 1 Comparison of the algorithm in this paper with other models

Error statistics	BP neural network	Random forest	Linear regression	Bagging-BP	PCA-DC-Bagging-BP	PCa-DC-Bagging-SVR
Mean value	20.34	- 2.35	- 18.32	- 11.02	- 9.37	- 4.08
Maximum deviation	- 410.4	- 367.6	- 332	- 310.2	- 287.2	- 237.8
Variance	9.84e+03	6.69E+03	1.59E+04	8.32e+03	7.86e+03	4.40E+03
Standard deviation	99.21	81.85	126.01	91.24	88.66	66.32
Mean square error	1.07e + 04	6.64E + 03	1.62E + 04	8.42e + 03	6.41e + 03	4.37E+03
Goodness of fit	0.71	0.80	0.60	0.76	0.79	0.87

RETRACTED ARTICLE

results for various statistics are: the average value is -2.35 , the maximum deviation is -237.8 , the variance is $4.40e+03$, the standard deviation is 66.32 , and the mean square error is $4.37e+03$, the goodness of fit is 0.87 , except that the best mean is obtained by the random forest algorithm, and the other best results are obtained by the PCA-DC-Bagging-SVR algorithm of SVR-based learner. This shows that the proposed algorithm is not only superior to the traditional Bagging algorithm, but also shows that the improved Bagging algorithm of the decision-making committee is superior to the traditional algorithm. At the same time, it can be seen that for a

single learner, the PCA-DC-Bagging-BP algorithm proposed in this paper has a significant improvement over the performance of the traditional BP neural network. Compared with random forests, the prediction performance of PCA-DC-Bagging algorithm based on SVR learner is higher than that of random forest algorithm, while the performance of PCA-DC-Bagging algorithm based on BP neural network learner is slightly lower than that of random forest algorithm. The explanation given here is: On the one hand, the random forest algorithm is also a variant of the Bagging algorithm, which increases the diversity of the learner by means of attribute scrambling, which makes the random forest algorithm often appear with superior performance in many learning tasks [43]. On the other hand, different algorithm performances may be different for different data sets. For example, the PCA-DC-Bagging algorithm based on SVR can get the best result for RAFM steel data set. Although the PCA algorithm has been used for dimensionality reduction, an important feature is that the target data set (here is RAFM steel) to be processed in this study is sparse [44]. For the SVR prediction process, it is advantageous to map the training data to the high-dimensional feature space to divide the hyperplane [45]. Therefore, the SVR-based learner is more effective than the BP neural network in prediction results. It also shows that the decision-making method of the decision-making committee and the Bagging algorithm using the arithmetic average combination strategy make the model very inclusive. Through theoretical derivation and the above experimental verification, the PCA-DC-Bagging framework algorithm proposed in this study is effective.

7 Conclusion

The prediction results of a few base learners will cause great deviation for the traditional bagging algorithm. Therefore, this paper proposes a discriminant analysis based on the decision committee model. The level of each learner is evaluated and divided, and the decision committee model is trained by verifying the error performance on the set. Decision committee members can use most of the discriminant classifiers, such as neural networks, decision trees, and naive Bayesian model. The algorithm presented in this paper is universal.

The test result shows that the PCA-DC-Bagging algorithm presented in this paper solves the prediction of this dataset where high redundancy, multiple repetitions, and sample data with many outliers. This algorithm not only solves the shortcomings of the traditional Bagging algorithm, but also gives a strict theoretical framework to ensure further development. The future development direction of the PCA-DC-Bagging algorithm. For the data set above the RAFM steel, decision-making committees are able to

effectively screening base learners, and the results are 2.8 times of random guesses. If there is an improvement in the decision-making accuracy of the committee, the effect of ensemble learning will be greatly improved.

Appendix: Statistic information of sample data

Refer to Table 2.

Table 2 Basic information of the input parameters

Index	Input variable	Minimum	Maximum	Average	Variance	Standard deviation
1	Processing parameters (%)	0	10	0.0828	0.8005	0.94
2	C (wt%)	0.087	0.2	0.0971	0.00010	0.013
3	Cr (wt%)	2.25	12	8.3269	1.54729	1.027
4	W (wt%)	0	3	1.843	0.605284	0.778
5	Mo (wt%)	0	1	0.15	0.131769	0.363
6	Ta (wt%)	0	0.54	0.0635	0.010404	0.102
7	V (wt%)	0	0.3	0.1822	0.002916	0.054
8	Si (wt%)	0	0.37	0.0546	0.002704	0.052
9	Mn (wt%)	0	0.3	0.1445	0.041616	0.204
10	N (wt%)	0	0.06	0.0025	0.006561	0.081
11	Al (wt%)	0	0.054	0.0008	0.001369	0.037
12	As (wt%)	0	0.005	0	0.0000009	0.0003
13	B (Vwt%)	0	0.0085	0.0007	0.00000169	0.0013
14	Bi (wt%)	0	0.005	0	0.0000009	0.0003
15	Ce (wt%)	0	0.036	0.0001	0.00000484	0.0022
16	Co (wt%)	0	0.01	0.0002	0.00000081	0.0009
17	Cu (wt%)	0	0.035	0.0006	0.00001024	0.0032
18	Ge (wt%)	0	1.2	0.0132	0.016384	0.128
19	Mg (wt%)	0	0.01	0	0.00000036	0.0006
20	Nb (wt%)	0	0.16	0.00164	0.000121	0.011
21	Ni (wt%)	0	2	0.0566	0.0961	0.31
22	O (wt%)	0	0.009	0.0002	0.00000121	0.0011
23	P wt%	0	0.007	0.0013	0.00000196	0.0014
24	Pb (wt%)	0	0.005	0	0.0000009	0.0003
25	S (wt%)	0	0.005	0.0012	0.00000121	0.0011
26	Sb (wt%)	0	0.003	0	0.0000004	0.0002
27	Se (wt%)	0	0.003	0	0.0000004	0.0002
28	Sn (wt%)	0	0.003	0	0.0000004	0.0002

Table 2 continued

Index	Input variable	Minimum	Maximum	Average	Variance	Standard deviation
29	Te (Wi%)	0	0.005	0	0.00000009	0.0003
30	Ti (wt%)	0	0.25	0.01	0.002116	0.046
31	Zn (wt%)	0	0.005	0	0.00000009	0.0003
32	Zr (wt%)	0	0.059	0.003	0.00001296	0.0036
33	Irradiation temperature (wt%)	273	925	401.0928	32148.49	179.3
34	Irradiation dose (dpa)	0	90	3.4287	100.8016	10.04
35	Irradiated He (dpa)	0	5000	35.697	129456.04	359.8
36	He_dpa (dpa)	0	6315.7895	56.1696	343119.5151	585.7643
37	Test temperature (k)	123	973	549.78	43,848.06	209.4

References

- Peng L, Ge H, Dai Y et al (2016) Microstructure and microhardness of CLAM steel irradiated up to 20.8 dpa in STIP-V. *J Nucl Mater* 468:255–259
- Kano S, Yang HL, Suzue R et al (2016) Precipitation of carbides in F82H steels and its impact on mechanical strength. *Nuclear Mater Energy* 9(C):331–337
- Li A, Zhao Y (2018) Application of improved genetic algorithm based on bagging ensemble blustering in assembly line balancing. *Machinery*
- Pham BT, Bui DT, Prakash I (2018) Bagging based support vector machines for spatial prediction of landslides. *Environ Earth Sci* 77(4):145
- Xinbo H, Wenjunzi LI, Tong S et al (2016) Application of Bagging-CART algorithm optimized by genetic algorithm in transformer fault diagnosis. *High Volt Eng* 42:1617–1623
- Yang Y, Jiang J (2016) Hybrid sampling-based clustering ensemble with global and local constitutions. *IEEE Trans Neural Netw Learn Syst* 27(5):952–965
- Yang Y, Jiang J (2019) Adaptive bi-weighting toward automatic initialization and model selection for HMM-based hybrid multi-clustering ensembles. *IEEE Transactions on Cybernetics* 99:1–12. <https://doi.org/10.1109/TCYB.2018.2809562>
- Gardner BJ, Gransberg DP, Rueda JA (2017) Stochastic conceptual cost estimating of highway projects to communication uncertainty using bootstrap sampling. *ASCE-ASME J Risk Uncertain Eng Syst Part A Civ Eng* 3(3):04017002
- Vijayanand MD, Vanaja J, Das CR et al (2018) An investigation of microstructural evolution in electron beam welded RAFM steel and 316LN SS dissimilar joint under creep loading conditions. *Mater Sci Eng* 742:432–441
- Lan K, Saroja S, Moitra A et al (2013) Development of India-specific RAFM steel through optimization of tungsten and tantalum contents for better combination of impact, tensile, low cycle fatigue and creep properties. *J Nucl Mater* 439(1–3):41–50
- Mao Chunliang, Liu Chenxi et al (2019) The correlation among microstructural parameter and dynamic strain aging (DSA) in influencing the mechanical properties of a reduced activated ferritic-martensitic (RAFM) steel. *Mater Sci Eng A* 40:90–98
- Zhang L, Shah SK, Kakadiaris IA (2017) Hierarchical multi-label classification using fully associative ensemble learning. *Pattern Recognit* 70:89–103
- Oh TH, Tai YW, Bazin JC et al (2016) Partial sum minimization of singular values in robust PCA: algorithm and applications. *IEEE Trans Pattern Anal Mach Intell* 38(4):744–758
- Gao M, Yin L, Ning J (2018) Artificial neural network model for ozone concentration estimation and Monte Carlo analysis. *Atmos Environ* 184:129–139
- Zhang G, Wang S, Wang Y et al (2017) LS-SVM approximate solution for affine nonlinear systems with partially unknown functions. *J Indus Manag Optim* 10(2):621–636

16. Baghdadi MHE, Darvish H, Rezaei H et al (2018) Applying LSSVM algorithm as a novel and accurate method for estimation of interfacial tension of brine and hydrocarbons. *Pet Sci Technol* 36(15):1–5
17. Meng Q, Ke G, Wang T et al (2016) A Communication-efficient parallel algorithm for decision tree. In: *Proceedings of the 30th international conference on neural information systems*. Curran Associates Inc, USA, pp 1279–1287
18. Kim K (2016) A hybrid classification algorithm by subspace partitioning through semi-supervised decision tree. Elsevier Science Inc, Amsterdam
19. Bryll R, Gutierrez-Osuna R, Quek F (2003) Attribute bagging: improving accuracy of classifier ensembles by using random feature subsets. *Pattern Recognit* 36(6):1291–1302
20. Guisan A, Thuiller W, Zimmermann NE (2017) Boosting and bagging approaches. *Habitat suitability and distribution models: with applications in R. Ecology, biodiversity and conservation*. Cambridge University Press, Cambridge, pp 202–216. <https://doi.org/10.1017/9781139028271.018>
21. Folkes SR, Lahav O, Maddox SJ (2018) An artificial neural network approach to the classification of galaxy spectra. *Mon Not R Astron Soc* 283(2):651–665
22. Lachaize M, Le Hégarat-Masclé S, Aldea E, Maitrot A, Reynaud R (2016) SVM classifier fusion using belief functions: application to hyperspectral data classification. In: *Vejnarová J, Kratochvíl V (eds) Belief functions: theory and applications. BELIEF 2016. Lecture notes in computer science*, vol 9861. Springer, Cham. https://doi.org/10.1007/978-3-319-45559-4_12
23. Yeo B, Grant D (2018) Predicting service industry performance using decision tree analysis. *Int J Inf Manag* 38(1):288–300
24. Marchiori E, Sebag M (2005) Bayesian learning with local support vector machines for cancer classification with gene expression data. In: *Rothlauf F et al (eds) Applications of evolutionary computing. EvoWorkshops 2005. Lecture notes in computer science*, vol 3449. Springer, Berlin. https://doi.org/10.1007/978-3-540-32003-6_8
25. Liping Z, Jiekang W, Feida T et al (2018) Oil-paper insulation evaluation method of transformer based on kernel principal component analysis and random forest algorithm. *Sichuan Electric Power Technol*
26. Ji-Shan LI, Liu QP, Qiao JJ et al (2018) Application of least square method to power grid voltage fitting waveform function. *Value Eng*
27. Li J, Cen Z, Li X (2018) Simulation of aspheric tolerance with polynomial fitting. In: *International conference on optical instruments and technology 2017: Optical systems and modern optoelectronic instruments*, p 14
28. Bertsekas D, Boplug C (2016) *Convex optimization algorithms*. Athena Scientific, Belmont
29. Li M (2018) Generalized Lagrange multiplier method and kkt conditions with application to distributed optimization. *IEEE Trans Circuits Syst II Express Briefs* 66(99):1
30. Bhat BVR, Parthasarathy K (1994) Kolmogorov's existence theorem for Markov processes in C^* algebras. *Proc Math Sci* 104(1):253–262
31. Wang D, Guo H, Luo H et al (2017) Multi-step-ahead electricity price forecasting using a hybrid model based on two-layer decomposition technique and BP neural network optimized by firefly algorithm. *Appl Energy* 190:40–47
32. Bogachev VI, Gnaniourov AF (2016) On weak convergence of finite-dimensional and infinite-dimensional distributions of random processes. *Natl Res Univ High Sch Econ* 21:1–11
33. Andrews J (2018) Addressing overfitting and underfitting in Gaussian model-based clustering. *Comput Stat Data Anal* 127:160–171
34. Li B, Luo J, Du J et al (2017) Improved algorithm of C4.5 decision tree on the arithmetic average optimal selection classification attribute. In: *IEEE international conference on bioinformatics & biomedicine*
35. Glen AG, Leemis LM, Barr DR (2017) Order statistics in goodness-of-fit testing. *IEEE Trans Reliab* 50(2):209–213
36. Liu Q, Lee JD, Jordan M (2016) A kernelized stein discrepancy for goodness-of-fit tests. In: *International conference on machine learning*. www.jmlr.org, pp 276–284
37. Pescim RR, Ortega EMM, Cordeiro GM et al (2017) A new log-location regression model: estimation, influence diagnostics and residual analysis. *J Appl Stat* 44(2):233–252
38. Lu C, Zhou Z, Zhu Q et al (2017) Using residual analysis in electromagnetic induction data interpretation to improve the prediction of soil properties. *CATENA* 149:176–184
39. Azzalini A, Capitanio A (2010) Statistical applications of the multivariate skew normal distribution. *J Roy Stat Soc* 61(3):579–602

40. Picinbono B (2018) Second-order complex random vectors and normal distributions. *IEEE Trans Signal Process* 44(10):2637–2640
41. Ghanem SAM (2016) Mutual information and minimum mean-square error in multiuser Gaussian channels. *IEEE* 10:18–21
42. Brassington G (2017) Mean absolute error and root mean square error: which is the better metric for assessing model performance? In: EGU general assembly conference. EGU General Assembly Conference Abstracts
43. Sylvester EVA, Bentzen P, Bradbury IR et al (2018) Applications of random forest feature selection for fine-scale genetic population assignment. *Evol Appl* 11(2):153–165
44. Bach F, Jenatton R, Mairal J et al (2012) Optimization with sparsity-inducing penalties. *Found Trends Mach Learn* 4(1):1–106
45. Aburomman AA, Reaz MBI (2016) A novel SVM-kNN-PSO ensemble method for intrusion detection system. *Appl Soft Comput* 38(C):360–372

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Sifan Long¹  · Ming Zhao¹ · Jieqiong Song¹

Sifan Long
1539895062@qq.com

Jieqiong Song
joansung@csu.edu.cn

¹ School of Software, Central South University, Changsha, China