doi:10.3772/j.issn.1006-6748.2023.02.011

Prediction of users online purchase behavior based on selective ensemble learning①

TAN Hui(谭 惠), DUAN Yong②
(School of In formation Science Engineering, Shenyang University of Technology, Shenyang 110870, P. R. China)

Abstract

A probabilistic multi-dimensional selective ensemble learning method and its application in the prediction of users' online purchase behavior are studied in this work. Firstly, the classifier is integrated based on the dimension of predicted probability, and the pruning algorithm based on greedy forward search is obtained by combining the two indicators of accuracy and complementarity. Then the pruning algorithm is integrated into the Stacking ensemble method to establish a user online shopping behavior prediction model based on the probabilistic multi-dimensional selective ensemble method. Finally, the research method is compared with the prediction results of individual learners in ensemble learning and the Stacking ensemble method without pruning. The experimental results show that the proposed method can reduce the scale of integration, improve the prediction accuracy of the model, and predict the user's online purchase behavior.

Key words: users' online purchase behavior, Stacking, selective ensemble, ensemble pruning, feature engineering

0 Introduction

Under the new situation of online shopping, predicting consumers' online purchasing behavior has attracted more and more scholars' attention. By analyzing massive data sets, it is of great significance for e-commerce platforms to use powerful machine learning algorithms to predict users' purchase intentions. Compared with the traditional recommendation algorithm based on the relationship between users and products[1-3], user network behavior prediction also needs to consider the user's overall behavior data, which is more complicated in implementation. At present, the processing of user behavior data is still in the exploratory stage. A key step in studying the behaviors of users when they purchase online is to mine the information in their behaviors from scattered data sets through feature engineering.

At the same time, due to the complexity of users' online purchase behavior, it is easy to fall into overfitting by using only a single model for prediction. For this reason, many scholars have proposed a method of combining multiple models for prediction through ensemble learning to solve this problem. Ref. [4] integrated decision tree and logistic regression method to build a prediction model of users' online purchase behavior. Refs. [5-6] integrated logistic regression and support vector machine methods to predict the consumption behavior of users. Ref. [7] combined long short-term memory (LSTM) and random forest method and found that the fusion method has better accuracy and recall rate than the single learner constructed by the random forest method. Ref. [8] built models based on convolutional neural networks-LSTM (CNN-LSTM), and at the same time balanced the samples through the 'segment downsampling' method, and obtained a better F1 value.

The ensemble learning method improves the prediction accuracy and generalization ability of the model by integrating a large number of base learners[9], but with the increase of base learners, the requirements for computational efficiency and storage capacity are also higher[10]. In order to solve the above problems, Ref. [11] proposed the concept of 'selective integration', which selects a subset of learners for integration through a certain method, and reduces the computing time and storage space under the premise of maintaining the computing accuracy. In addition, extensive experiments show that selective ensemble methods can also improve generalization ability[12-13]. Over the past decade, a variety of effective selective integration methods have emerged one after an-

① Supported by the Scientific Research Foundation of Liaoning Provincial Department of Education (No. LJKZ0139).
② To whom correspondence should be addressed. E-mail: duanyong0607@163.com.
Received on Aug. 23, 2022

other [14-16]. Among them, how to construct screening rules to obtain the best subset of learners is a key step in selective integration.

Based on this, this paper builds a prediction model through a probabilistic multi-dimensional selective integration method for the prediction of users' online purchase behavior. Firstly, based on the pruning problem, research is carried out in three aspects: the construction of individual learner subsets, the integration of the output results of the learner subsets, and the construction of the evaluation rules of the learner subsets. Then, the greedy forward pruning method based on probability multi-dimensionality is combined with the traditional stacking method to obtain a selective integration method based on probability multi-dimensionality, and a model is established based on the prediction problem of users' online purchase behavior.

The rest of this article is organized as follows. Section 1 introduces a greedy forward pruning method based on probabilistic multi-dimensionality. Section 2 introduces a selective ensemble method based on probabilistic multi-dimension. Section 3 presents the experimental results of this method. Finally, it concludes in Section 4.

1 Greedy forward pruning method based on the dimension of predicted probability

Dataset of this paper is Alibaba's E-commerce data from Tianchi Big Data platform. The dataset includes user basic information table, user behavior log table, user click information table, and advertising basic information table. It describes user attribute information, product details, user product interaction information, and so on. The user behavior log table and user click information table record more than 700 million behavior information that users interact with different commodities within 22 d. Combined with the user behavior trajectory in this dataset, the prediction of the user's future behavior is completed.

However, due to the mass of commodity and user information, the complexity and randomness of user interaction, and the instability of network log information records, the whole dataset has the problems of huge data sample size, missing key information, and insufficient key features.

First of all, user behavior data is extracted proportionally from each sub-data block according to the number of user occurrences, and the data is merged. A user behavior dataset is obtained for data analysis. Secondly, the data density of the invalid matrix in the dataset is analysed to understand the missing distribution in the dataset. And for features with high correlation, the random forest method is used to complete the filling. Finally, new derived features are constructed through feature transformation and fusion, so as to improve the information content of dataset and the accuracy of machine learning prediction. Based on the dimensions of business and research objectives, XGBoost method is used to calculate the importance scores of continuous features and discrete features, and the threshold is set to complete feature screening.

The operation steps of selective integration can be summarized as follows: after training multiple individual learners through different methods, select individual learners according to certain rules to form multiple individual learner subsets, and integrate the output results of the individual learner subsets, obtain the prediction result of the subset, and then judge the prediction effect of the individual learner subset according to certain judgment rules, and finally obtain the individual learner subset with the best prediction effect. The above process is repeated continuously until the ensemble scale is reached, the pruning is completed, and the subset of individual learners is used as the meta-learner of the second layer ensemble.

In this process, how to form individual learner subsets, how to integrate the output results of individual learner subsets, and how to select the optimal individual learner subsets are the main research issues of this paper.

1.1 Probability based on voting

Hard voting is one of the most widely used methods for integrating the outputs of subsets of individual learners. For each sample, the prediction results of all individual learners are obtained, and then the category with the most occurrences is selected by voting, which is the prediction result of the sample. Although this method can get the classification results simply and quickly, it only votes for the prediction results, but it is easy to ignore that the prediction probabilities of different learners for different categories are not the same, thus ignoring the details. Therefore, in this paper, the dimension of voting is adjusted from the category to the probability sum of each category.

Let D be a distribution on $\chi \times \{-1, +1\}$ and the validation set $V = \{(\chi_i, y_i)\}_{h=1}^{m}$ be a subset on distribution D. Given a set $H = \{h_i(x)\}_{h=1}^{n}$ containing n individual learners, where each learner $h_i: \chi \rightarrow \{-1, +1\}$ maps the feature space χ to the class label set $\{-1, +1\}$, the voting rule defines a decision function by taking the category with the largest sum of the pre-

dicted probabilities of different classifiers $h_i(x)$.

$$f(x, H) = \underset{j}{\operatorname{argmax}} \sum_{i=1}^{n} P_{(x, y)\sim V}[h_i(x) = j], \quad j \in \{-1, +1\} \tag{1}$$

where, $f(x, H)$ is the class label output by the sample after integrating the training results in the individual learner set $H = \{h_i(x)\}_{h=1}^{n}$. By summing and voting the probabilities of each category of the sample under different classifiers, find the category with the largest sample probability and the largest category, which is the predicted category of the sample.

1.2 Multi-dimensional evaluation criteria

In selective integration, evaluation criteria are a critical step in method design. By constructing reasonable evaluation criteria and accurately evaluating the available actions in each search step, a better pruned subset can be obtained.

The accuracy of the model has always been an important indicator for judging the performance of the learner. Compared with indicators such as accuracy and precision, the area under curve (AUC) indicator itself has nothing to do with the absolute value of the model prediction score, and only focuses on the sorting effect, so it is especially suitable for sorting business. At the same time, the AUC calculation method takes into account the learner's ability to classify positive and negative examples, and can still make a reasonable evaluation of the classifier in the case of unbalanced samples. The data samples in this paper are unbalanced samples, so the AUC index is considered to measure the accuracy of the model in the evaluation standard.

The number of positive samples of individual learner set $H = \{h_i(x)\}_{h=1}^{n}$ in validation set V is

$$M(x, H) = \sum_{(x, y)\in V} II[f(x, H) = y, y = 1] \tag{2}$$

where $II[z]$ is the indicator function, if z is true, then it is 1, otherwise it is 0.

Similarly, the number of negative samples of individual learner set $H = \{h_i(x)\}_{h=1}^{n}$ in validation set V is

$$N(x, H) = \sum_{(x, y)\in V} II[f(x, H) \neq y, y = 1] \tag{3}$$

Let P be a subset on the validation set V, and the points in P satisfy $II[f(x, H) = y, y = 1] = 1$, the AUC index of the individual learner set H acting on the validation set V is

$$A(x, H) = \frac{\sum_{i\in P} (rank_i - \frac{1}{2}M(x, H) \times (M(x, H) + 1))}{M(x, H) \times N(x, H)} \tag{4}$$

among them, $rank_i$ is the sequence number of the probability score of each point in $P = \{(\chi_i, y_i)\}_{h=1}^{l}$ in ascending order.

Furthermore, integration diversity is an important issue in the selective integration process. For performance to improve after combining, there must be differences between individual learners. This paper enhances ensemble diversity by measuring the similarity between learners and selecting individual learners that are complementary to the current ensemble.

Then on the validation set $V = \{(\chi_i, y_i)\}_{h=1}^{m}$, the diversity of individual learner h_t and individual learner set $H_t = H_t / \{h_t\}$ is

$$C(x, H) = \frac{1}{m} \sum_{i=1}^{m} II[h_t(x) f(x, H_t) < 0] \tag{5}$$

An ensemble of individual learners should be both precise and diverse. Compared with a single evaluation index, the combination of the two can obviously improve the selective integration effect better. Therefore, this paper combines the AUC index and the diversity index to give the final evaluation function:

$$S(x, H) = \alpha A(x, H) + \beta C(x, H) \tag{6}$$

among then, $\alpha, \beta \in (0, 1)$.

1.3 Selective ensemble based on probabilistic multi-dimension

Given a set of trained classifiers, it is difficult to select the subset with the best generalization performance for two main reasons: first, the generalization performance of the subset is not easy to estimate; second, finding the optimal subset is a computational combinatorial search problems with exponential complexity, so it is not feasible to compute exact solutions with exhaustive search, and approximate search is required.

In the past decade, many methods have been proposed to overcome this problem[14], and the main search methods can be roughly divided into two categories. The first class of methods uses a global search to directly select the optimal or near-optimal subset of classifiers, such as genetic algorithm, semi-definite programming[17], clustering[18-19], sparseness-induced prior[20] or l_1 norm constraint[21], etc. In practical applications, although those methods can achieve better performance, their computational cost is usually high. The second class of methods is a greedy local search of all possible ensemble subsets[12, 22]. According to the search direction, this group of methods can be further divided into greedy forward search methods that start from an empty set and iteratively add classifiers that optimize a certain condition, and greedy backward

search methods that start from full integration and iteratively eliminate classifiers. It has been shown that greedy local search methods can achieve performance and robustness compared with global search methods, but at a much lower computational cost[12, 23]. Furthermore, Partalas[22] found similar performance in both directions based on extensive experiments, but the forward generated ensemble size was smaller.

Therefore, this paper adopts the greedy forward search algorithm, based on the dimension of probability voting, combined with the two indicators of accuracy and complementarity, to selectively integrate the individual learner set $H = \{h_i(x)\}_{i=1}^{n}$. The algorithm is shown in Algorithm 1.

Algorithm 1 Pruning method based on probabilistic multi-dimension

Input: the set of classifiers to be pruned $H = \{h_i(x)\}_{i=1}^{n}$, validation set $V = \{(x_i, y_i)\}_{i=1}^{m}$, Specify the number of classifier prunings N, parameter α, $\beta \in (0, 1)$

Output: the set of pruned classifiers H^*

1: initialization $H^* \leftarrow \varphi$
2: $(x) \leftarrow \max_{h \in H} S(x, h)$, $(x, y) \in V$
3: $H^* \leftarrow \{h(x)\}$, $H \leftarrow H/\{h(x)\}$
4: **repeat**:
5: **for each** $h'(x)$ **do in** H
6: $H' \leftarrow H^* \cup h'(x)$
7: based Eq. (6), calculate $S_{h'} \leftarrow S(x, H')$
8: **end for**
9: list $L \leftarrow$ Sort the individual learners $h'(x)$ in H according to their corresponding $S_{h'}$ values in descending order
10: $H(x) \leftarrow$ Choose the first individual learner $h'(x)$ in L
11: $H^* \leftarrow \{h(x)\}$, and $H \leftarrow H/\{h(x)\}$
12: The number of learners in until H^* is equal to N

2 Prediction model of users' online purchase behavior based on selective integration

This section describes a selective ensemble method based on probabilistic multi-dimensionality, which combines the stacking ensemble method, adopts the greedy forward search algorithm, based on the dimension of probability voting, and combines the two indicators of accuracy and complementarity. By selective integration, a user online shopping behavior prediction model based on this method is finally built.

First of all, it is necessary to train a large number of base learners as weak learners for ensemble learning. However, since the weak learners are both good and bad, in order to avoid under-fitting, it is necessary to remove weak learners with poor performance. For fitting, it is also necessary to remove weak learners that perform too well. Through the above steps, a set of weak learners can be obtained.

The weak learner set is trained as the weak learner of the first layer of Stacking ensemble learning. For each weak learner, a four-fold cross-validation method is used for training. The training set is divided into four parts, one part is selected as the validation set each time, and the remaining three parts are used as the training set. The weak learner trains a model based on the training set, and then the model makes predictions on the validation set. But note that the model does not predict the class label, but the probability data corresponding to each class label of the sample.

Each weak learner can repeat the above training four times, and splice the four predictions, so as to obtain a prediction data based on the training set. At the same time, the model obtained by each training of the weak learner will also make a prediction on the test set, and the prediction data obtained by the weak learner based on the test set can be obtained by summing the prediction sets obtained by the four predictions and taking the average value.

Four-fold cross-validation training is performed on each weak learner, and each weak learner produces a prediction data based on the training set and prediction data based on the test set. According to the prediction probability results based on the training set corresponding to each weak learner, pruning is performed.

First determine the pruning threshold, that is, specify the number of models after pruning, and then select a weak learner based on the prediction probability output of the training set as the initial set, and then use the remaining weak learners' predictions based on the training set output. The probability results are added to the set, and the evaluation scores are based on accuracy and complementarity. Finally, the weak learner with the best score is added to the initial set. For the remaining weak learners, the above steps are repeated continuously, and weak learners are selected and added to the set until the number of weak learners in the set reaches the threshold. At this time, the weak learners in the set are the weak learners obtained after pruning.

Then, the predicted category probability results based on the training set and the test set corresponding to the selected weak learner set are converted into the predicted category results, so as to obtain the first layer output of Stacking ensemble learning. Taking the predicted category results based on the training set as input, the second-layer training of Stacking ensemble learning is performed to obtain the second-layer model. The second-layer model is then used to predict the first-layer prediction category results based on the test

set, which can be compared with the real category results to obtain a model score.

3 Experimental results and analysis

In this section, the selective ensemble method based on probability multi-dimensionality is compared with the individual learner method and Stacking ensemble method included in the ensemble through experiments, and the selective ensemble method based on probability multi-dimensionality is evaluated to verify the effectiveness of its theoretical results.

3.1 A comparative experiment between the selective ensemble method based on probabilistic multi-dimensionality and the individual learner method

In the ensemble learning process, this paper integrates heterogeneous learners. For a total of 31 learners generated by GassianNB, LogisticRegression, support vector classification (SVC), DecisionTree under different parameters, the traditional Stacking integration method is pruned based on the selective integration method of probability multi-dimensional, and finally 20 single learners are is obtained. Prediction model of user online purchase behavioris obtained by integration.

For the five indicators of precision, recall, f1_score, roc_auc, and accuracy, the model obtained by the selective integration method is compared with the maximum and average scores of the models generated by GassianNB, LogisticRegression, SVC, and DecisionTree.

As listed in Table 1, av_GB, av_LR, av_SVC, and av_DT respectively represent the average values of the models generated by the above four methods under the corresponding indicators, and max_GB, max_LR, max_SVC, and max_DT respectively represent the maximum values of models generated by the above four methods under the corresponding indicators. PMEP represents the score of the model generated by the probability multi-dimensional selective integration method under the corresponding index.

In the table, compare the index scores of each learner with the scores corresponding to PMEP, and mark the graph ●(○) after the effect is lower (higher) than the index score of PMEP.

Observing the table, it can be found that although the models generated based on GassianNB and Logistic Regression perform well in precision indicators, they do not perform well in other indicators because the problem of user purchase behavior prediction is an imbalanced sample. In addition, although the scores of the models generated by SVC and DecisionTree are slightly higher than those generated by the probabilistic multi-dimensional selective integration method, they still perform poorly on the overall indicators and are not stable. Therefore, it can be found that the selective integration method based on probabilistic multi-dimensionality performs better and is more stable than that of predicting user purchasing behavior with unbalanced samples.

Table 1 Performance comparison between PMEP and individual learners

	precision	recall	f1_score	roc_auc	accuracy
av_GB	0.928 14●	0.508 235●	0.656 811●	0.726 899●	0.691 72●
av_ LR	0.941 575●	0.761 939●	0.841 897●	0.847 941●	0.834 105●
av_SVC	0.924 613●	0.829 633○	0.873 652●	0.866 852●	0.860 864●
av_DT	0.858 428●	0.823 608○	0.838 531●	0.814 208●	0.815 72●
max_GB	0.928 14●	0.508 235●	0.656 811●	0.726 899●	0.691 72●
max_LR	0.957 28●	0.789 746●	0.852 615●	0.851 446●	0.841 52●
max_SVC	0.951 974●	0.884 226○	0.900 389○	0.886 865●	0.886 44○
max_DT	0.954 51●	0.908 07○	0.867 295●	0.857 252●	0.849 72●
average	0.886 476●	0.803 119●	0.839 079●	0.825 939●	0.822 267●
max	0.957 28●	0.908 07○	0.900 389○	0.886 865●	0.886 44○
PMEP	0.975 694	0.821 584	0.892 031	0.896 634	0.884 56

Experimental results show that the model obtained by the selective ensemble method for heterogeneous learners is far more effective than the model obtained by a single learner.

3.2 Comparative experiment between selective integration method based on probability multi-dimension and Stacking integration method

Compared with the Stacking ensemble method that

integrates all classifiers, the selective ensemble method can not only reduce the ensemble size, but even improve the generalization performance. In order to verify the effectiveness of the selective integration method based on probabilistic multi-dimensionality in the prediction of users' online purchase behavior, this paper combines several evaluation indicators commonly used in classification problems, and records the selective integration method based on probability multi-dimension and Stacking integration method. The comparison of the following five indicators on the prediction of user purchase behavior is given.

The proposed method is compared with Stacking method and Catboost method[24]. The experiment results are shown in Fig. 1 It can be seen that the selective integration method based on probabilistic multi-dimensionality has improved the scores of recall, f1_score, roc_auc, and accuracy. Among them, the recall index is better than Stacking method, which means that the method improves the prediction accuracy of users who may have purchase behavior. At the same time, the accuracy index is higher than Stacking method, indicating that this method improves the overall prediction accuracy of the model. Although there is a slight drop in the precision indicator, a small number of users who will not make purchases may be regarded as users with purchase intentions. But this will not have a negative impact on merchants, it will help merchants attract potential customers. In addition, the improvement of the method on the f1_score index proves that the method takes into account the precision index and the recall index of the classification model at the same time, so that the two can reach the highest level at the same time and achieve a balance, which is more stable than the previous method. Finally, the roc_auc index of the proposed method is the best, which indicates that the method has improved ability to make reasonable evaluations in the case of unbalanced samples, and is more suitable for the problem of user purchase behavior prediction with unbalanced samples.

Furthermore, the experiment results illustrate that all indexes of the proposed method are also better than Catboost method, which verifies the effectiveness of the research work.

Catboost is a gradient boosting decison tree (GBDT) framework based on symmetric decision tree, which has fewer parameters, supports category variables and has high accuracy. However, one of the contributions of this paper is to select and integrate different types of base classifiers through probability multi-dimensionality, so as to reflect the diversity, complementarity and effectiveness of ensemble learning method and to get better performance and efficiency.

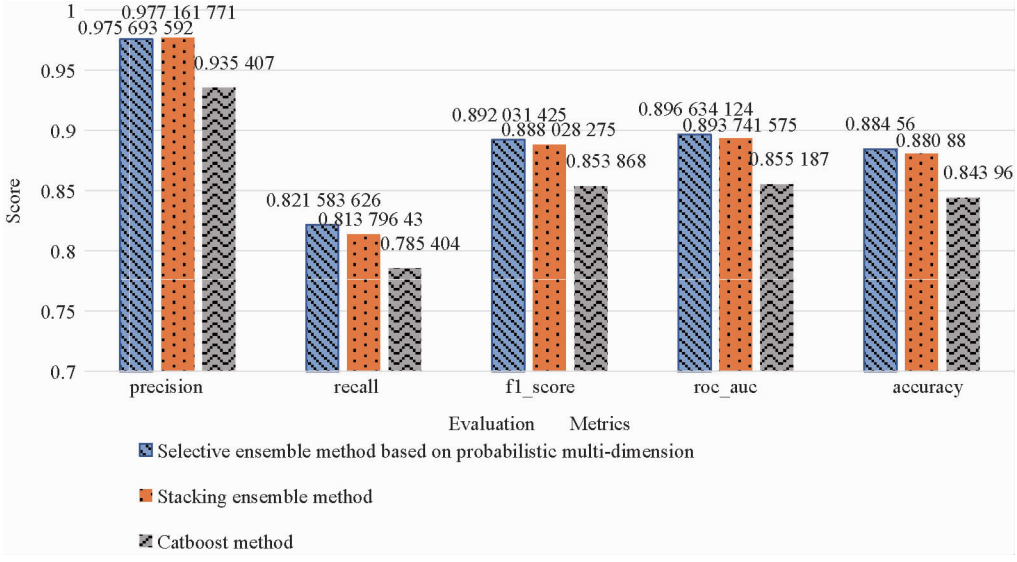

Fig. 1 Comparison of the scores of each indicator of different methods

4 Conclusion

Based on the prediction problem of users' online shopping behavior, this paper constructs a user's online shopping behavior information dataset through feature engineering methods such as data preprocessing, feature construction and screening. At the same time, in order to improve the prediction accuracy of the model and reduce the ensemble scale, the traditional Stacking ensemble method is pruned. Researches are carried out on how to form individual learner subsets, how

to integrate the output results of individual learner subsets, and how to determine the prediction performance of individual learner subsets. The learner subset is continuously constructed based on the greedy forward search method, the output of the learner subset is integrated in the dimension of category probability, and the prediction effect of the learner subset is evaluated based on the accuracy and diversity. Experiments show that, for the prediction of users' online purchase behavior, the selective integration method based on probability multi-dimension is better than the single algorithm without integration and the Stacking integration method without pruning, in the comprehensive performance of each evaluation index, which achieves better prediction results at a smaller scale.

References

[1] YU H, LI J H. Algorithm to solve the cold-start problem in new item recommendations[J]. Journal of Software, 2015, 26(6):1395-1408.

[2] WANG Q, YU J J. Diversity recommendation based on product co-purchase network[J]. Journal of Systems & Management, 2020, 29(1):61-72.

[3] SUN G F, WU L, LIU Q, et al. Recommendations based on collaborative filtering by exploiting sequential behaviors [J]. Journal of Software, 2013, 24(11):2721-2733.

[4] ZHANG P Y, WANG D X, JIAO Y F, et al. Predicting mobile purchase decisions based on user browsing logs [J]. Data Analysis and Knowledge Discovery, 2018, 2 (1):51-63.

[5] ZHU X, LIU X M, CHEN S G, et al. Research on network purchase behavior prediction based on machine learning fusion algorithm[J]. Statistics & Information Forum, 2017, 32(12):94-100.

[6] LIU X M. Prediction research of online purchasing behavior based on feature selection and model ensemble[D]. Beijing:Beijing Jiaotong University, 2017:5-8.

[7] LI X Y, SHAO F J. Purchase behavior forecasting model combining LSTM and random forest[J]. Journal of Qingdao University, 2018, 33(2):17-20.

[8] HU X L, ZHANG H B, DONG J C, et al. Prediction model of user buying behavior based on CNN-LSTM[J]. Computer Applications and Software, 2020, 37(6):59-64.

[9] FU H Y. Research on user purchase behavior prediction based on heterogeneous integration algorithm[D]. Jinan: Shandong University, 2020:10-15.

[10] MARTINEZ W G. Ensemble pruning via quadratic margin maximization[J]. IEEE Access, 2021, 9:48931-48951.

[11] ZHOU Z H, WU J, TANG W. Ensembling neural networks:many could be better than all[J]. Artificial Intelligence, 2002, 137(1-2):239-263.

[12] MARTINEZ-MUNOZ G, HERNANDEZ-LOBATO D, SUAREZ A. An analysis of ensemble pruning techniques based on ordered aggregation[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2009, 31(2): 245-259.

[13] ZHANG J, DAI Q, YAO C. DEP-TSPmeta: a multiple criteria dynamic ensemble pruning technique ad-hoc for time series prediction[J]. International Journal of Machine Learning and Cybernetics, 2021, 12(8):2213-2236.

[14] TSOUMAKAS G, PARTALAS I, VLAHAVAS I. An ensemble pruning primer[J]. Studies in Computational Intelligence, 2009, 245:1-13.

[15] GUO H, LIU H, LI R, et al. Margin & diversity based ordering ensemble pruning[J]. Neurocomputing, 2018, 275:237-246.

[16] NI Z W, ZHANG C, NI L P. Haze forecast method of selective ensemble based on glowworm swarm optimization algorithm [J]. Pattern Recognition and Artificial Intelligence, 2016, 29(2):143-153.

[17] ZHANG T, ZHANG L. Critical multipliers in semidefinite programming[J]. Asia-Pacific Journal of Operational Research, 2020, 37(4):1-20.

[18] TSAI C F, LIN W C, HU Y H, et al. Under-sampling class imbalanced datasets by combining clustering analysis and instance selection[J]. Information Sciences, 2019, 477:47-54.

[19] HAO Z H, ZHANG X J, XIE J C, et al. Building climate zones of major marine islands in China defined using two-stage zoning method and clustering analysis[J]. Frontiers of Architectural Research, 2021, 1:134-147.

[20] CHEN H, TINO P, YAO X. Predictive ensemble pruning by expectation propagation [J]. IEEE Transactions on Knowledge and Data Engineering, 2019, 21(7):999-1013.

[21] LIU Q, LIU F. Selective cascade of residual extra trees [J]. Computer Science, 2020, 1(6):1-11.

[22] PARTALAS I, TSOUMAKAS G, VLAHAVAS I. An ensemble uncertainty aware measure for directed hill climbing ensemble pruning[J]. Machine Learning, 2010, 81 (3):257-282.

[23] YE R, DAI Q. A novel greedy randomized dynamic ensemble selection algorithm[J]. Neural Processing Letters, 2018, 47(2):565-599.

[24] DOU X T. Online purchase behavior prediction and analysis using ensemble learning[C]//IEEE 5th International Conference on Cloud Computing and Big Data Analytics. Chengdu:IEEE, 2020:532-536.

TAN Hui, born in 1996. She is a graduate student in School of Information Science and Engineering of Shenyang University of Technology. She received her B. S. degree from Shenyang University of Technology in 2018. Her research interests include machine learning and intelligent software.