Eng Int Syst (2018) 4: 191–197

Engineering
Intelligent Systems

Research to the Universal Ensemble Learning Approaches Based on Adaboost Algorithm

Li Xiaolong*, **Chen Juanfen**†, **Liao Shiqin**‡

Jiangxi Institute of Fashion Technology Nanchang, Jiangxi, China

AdaBoost which is one of the most outstanding Boosting algorithms belonging to machine learning too, has a steady theoretical basis and has made great progress to the problem solving. AdaBoost can boost a weak learning algorithm with an accuracy slightly better than random guessing into an arbitrarily accurate strong learning algorithm, bringing about a new method and a new design idea to the design of universal learning approaches.This paper first introduces how AdaBoost, just a conjecture when proposed, was proved right, and how this proof led to the origin of AdaBoost algorithm, reviewing the development history of ensemble learning,and focuses on the three strategies of diversity generation, model training and model combination in ensemble learning, and then describes the relevant application scenarios of ensemble learning at the current stage. Finally, the future research direction of ensemble learning approaches is analyzed and discovered.

Keywords: Ensemble learning, AdaBoost algorithm, Boosting, Machine learning

1. INTRODUCTION

Boosting which was also called enhancing learning or promotion method was an important ensemble learning technology.Boosting could boost a weak learning algorithm with an accuracy slightly better than random guessing into an arbitrarily accurate strong learning algorithm, bringing about a new way and a novel idea to the design of learning algorithm. As a Meta-algorithm framework, Boosting nearly could apply to all popular machine learning algorithm to enhance the predicting accuracy of Meta-algorithm further. It was used in all wakes of life and had a great influence. However, Adaboost was the most successful representative among all algorithm, and was evaluated one of ten data mining[1]. During the past ten years of Adaboost, many famous scholars who were interested in the field of machine spent much time in the theory research of algorithm, which all put a solid foundation for the successful application of Adaboost algorithm. The success of Adaboost was due to the factors as follows: first, it was an effective learning algorithm; second, it promoted the development of Boosting, successfully the change from the original guessing into the economic algorithm.; third, the algorithm adopted some skills, such as breaking the original sample distribution, which brought out an important enlightenment to other statistics learning algorithm; finally, the relevant theory research drove the development of ensemble learning.

In recent decades, due to the efficient solving practical problem of ensemble learning which attracted much attention in the field of machine learning. At first, ensemble learn aimed

*Corresponding Author. E-mail: 576816683@qq.com
†E-mail: 464441913@qq.com
‡E-mail: 353613980@qq.com

to improve the accuracy of automatic decision. Nowadays, the method could solve all kinds of machine learning problems. The paper took AdaBoost as an example, introducing the theory origin of Boosting and dating back existing ensemble learning approaches, hoping to provide construction theory and ensemble learning approaches of ensemble learning system and for learners. In fact, decision making through ensemble learning ideas has existed since the beginning of civilization. For example, in a democratic society, citizens chose officials or made laws through the voting. For individuals, we could consult many doctors before major medical operations. These examples show that people need to weigh and combine various opinions to make the final decision. Actually, The original purpose of researchers using ensemble learning is similar to the reason why people use these mechanisms in their daily lives[2]Three basic reasons for the success of the ensemble approaches are explained mathematically: statistics, computation and representativeness. In addition, the effectiveness of ensemble learning can also be analyzed by means of bias variance decomposition.[3]

In 1979, the idea of ensemble learning was put forward for the first time by Dasarathy and Sheela.[4] In 1990, Anensemble model based on neural network was presented by Hansen and Salamon[5] The ensemble model has lower variance and better generalization ability. In the same year, the weak classifier can be combined into a strong classifier was proved through the method of Boosting by Schapire,[6] which made ensemble learning become an important research field of machine learning. Since then,ensemble learning research had developed rapidly, and many new ideas and models have emerged. In 2001, Random forest algorithm was put forward by Breiman which has been hailed as one of the best algorithm.[7] With the development of The Times, more and more ensemble learning algorithms have been proposed, and great breakthroughs have been made in many fields.

2. MAIN STRATEGY OF THE ENSEMBLE LEARNING

The main differences between ensemble learning algorithms lie in the following three aspects: The training data provided to individual learners are different; The process of producing the individual learner is different and combinations of learning outcomes are different. The paper will introduce the diversity generation method of ensemble learning, base learner training method and base learner combination strategy.

2.1 Diversity

Enhancing diversity is one of the keys to improve the generalization ability of ensemble learning. Many literatures have made a detailed theoretical analysis of the existing diversity measurement methods. The importance of diversity in ensemble learning is also discussed. In ensemble learning, diversity is mainly enhanced from three aspects: data, parameters and model structure.

Data sample diversity. There are three main ways to generate data diversity:[8] input sample perturbation, input attribute perturbation and output perturbation.

Input sample perturbation: generate multiple data subsets of different types according to the original data, and then use different data subsets to train the individual learner.The commonly methods which are used as follows: resampling method, data subset with fixed sample size can be obtained by sampling with put back, sequence sampling method, and sampling according to the previous round of learning results.

Input attribute perturbation: Several attribute subsets are extracted from the initial attribute set, and then the base learner is trained based on each attribute subset. This method can not only generate individuals with large differences, but also save computing time greatly because of the reduction of the number of attributes. The random subspace algorithm and Random forest algorithm belong to the algorithm. In this method, the original feature set is divided into several disjoint feature subsets to train individual learners respectively, and the final model is obtained by integrating multiple learners. Some studies show that feature set decomposition has good performance in learning tasks with high dimensional features.[9]

Output perturbation: The output representation can be manipulated to enhance diversity and the class standard of training samples can be changed. Flip method and output modulation method were used as to the output perturbation. The former randomly change the marking of some training samples; and the later fter the classification output is converted into regression output, the individual learner is constructed.

Algorithm parameter diversity. Algorithm parameter diversity refers to the use of different parameter sets to produce different individual learners. Even if each individual learner USES the same training set, the output of the individual learner will change with the change of the parameters due to the different parameters used. In order to improve the learning effect, the evaluation function of each learner is extended as a penalty item to enhance the diversity. Among them, the most common punishment method is negative correlation learning.[10] In ensemble learning, negative learning ideas use different learning devices to represent the different subspace of the problem, and when you're training the learning machine, you use the correlation penalty in the error function to increase the diversity of the learning machine.

Structural diversity. Structural diversity is mainly caused by the internal structure or external structure of the individual learner. In an ensemble learning system, if individual learners are generated by the same algorithm training, it is called homogeneous integration. On the contrary, if an ensemble system contains different types of individual learners, it is called heterogeneous integration.[11]

Basic learner training method. With the continuous development of the field of ensemble learning, researchers are constantly proposing new ensemble learning algorithms. But these algorithms are mostly adapted by some classical algorithms such as Bagging, Boosting, and Stacking These classical algorithms have good results and are widely used in various fields.

Bagging. Bagging algorithm is one of the earliest ensemble learning algorithms. It is simple in structure but superior in performance. The algorithm generates new training subsets by randomly changing the distribution of training sets, and then USES different training subsets to train individual learners, and finally ensembles them into a whole. In this algorithm, due to the use of self-sampling method to generate a new training subset, some instances will be sampled multiple times, while others will be ignored. Therefore, for a specific subspace, the individual learner will have a high classification accuracy, while for those parts that are ignored, the individual learner is difficult to correctly classify. However, the final prediction results are generated by multiple individual learners voting, so the better the effect of individual learners and the greater the difference between them, the better the effect of the integration algorithm will be. The Bagging algorithm is very effective for unstable learning algorithms due to the fact that the unstable learning algorithm is sensitive to the training set, so that the training set generates a small change, which leads to a large change in its prediction result.

Bagging algorithm is suitable for solving problems with small training sets, but its effect will be reduced for problems with large training sets. Therefore, Breiman has designed Pasting Small Votes algorithm based on Bagging, which can effectively deal with machine learning problems with large amount of data.

Figure 1 is the whole process of Bagging algorithm, which can train individual learners in parallel, so the algorithm has a high operation efficiency.

2.2 The Boosting

Boosting algorithm is a weak learning into the strong learning iteration method, it is by increasing the number of iterations, create a near-perfect performance strong learning. Among them, the weak learner refers to the learner whose classification effect is only slightly better than the random guess effect, that is, the classification accuracy is slightly higher than 50%. In practice, it is easier to acquire a weak learner than a strong one. Due to This, for Boosting series research significance of the algorithm. In addition to has a good practical performance, Boosting algorithm also has a strong theoretical foundation and the characteristics of the algorithm.

For supervised learning, each classifier generated by the algorithm after the first classifier learns from samples that were not correctly classified in the previous one. Therefore, this algorithm can effectively reduce the deviation of the model, but with the progress of training, the accuracy of the overall model in the training set keeps improving, leading to a larger variance. However, random sampling of features can reduce the correlation between classification models and thus reduce the variance of the whole model. When the primary classifier cannot be trusted to classify a given object.

This, for Boosting series research significance of the algorithm. In addition to has a good practical performance, Boosting algorithm also has a strong theoretical foundation and the characteristics of the algorithm. Because of the low confidence in the result, the data is transmitted to the auxiliary classifier and classifiers are added sequentially. A brief introduction to the process of Boosting algorithm was as follows: just like Fig. 2 shows, Boosting algorithm repeatedly run a weak learning to deal with different distribution of training data, then according to the order will produce each time the weak learning is combined into a composite good learning.

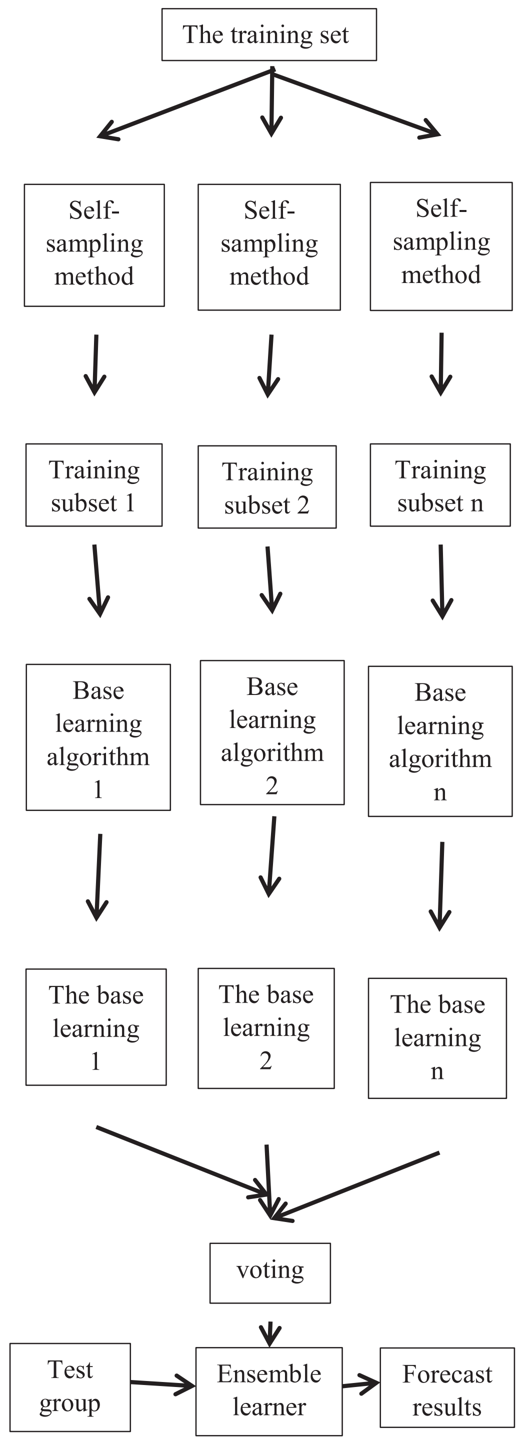

Figure 1 Procedure of Bagging algorithm.

2.3 The Stacking

Stacking is also called Stacked Generalization. [12]Training a model used to combine all the individual learners, that is, first train multiple different individual learners, and then use the output of these individual learners as input to train a model, so as to get a final output. Next, let me to briefly introduce the Stacking algorithm process. Just as like Fig. 3,

Multiple training subsets are obtained through the resampling method on the entire training data set, and then a series

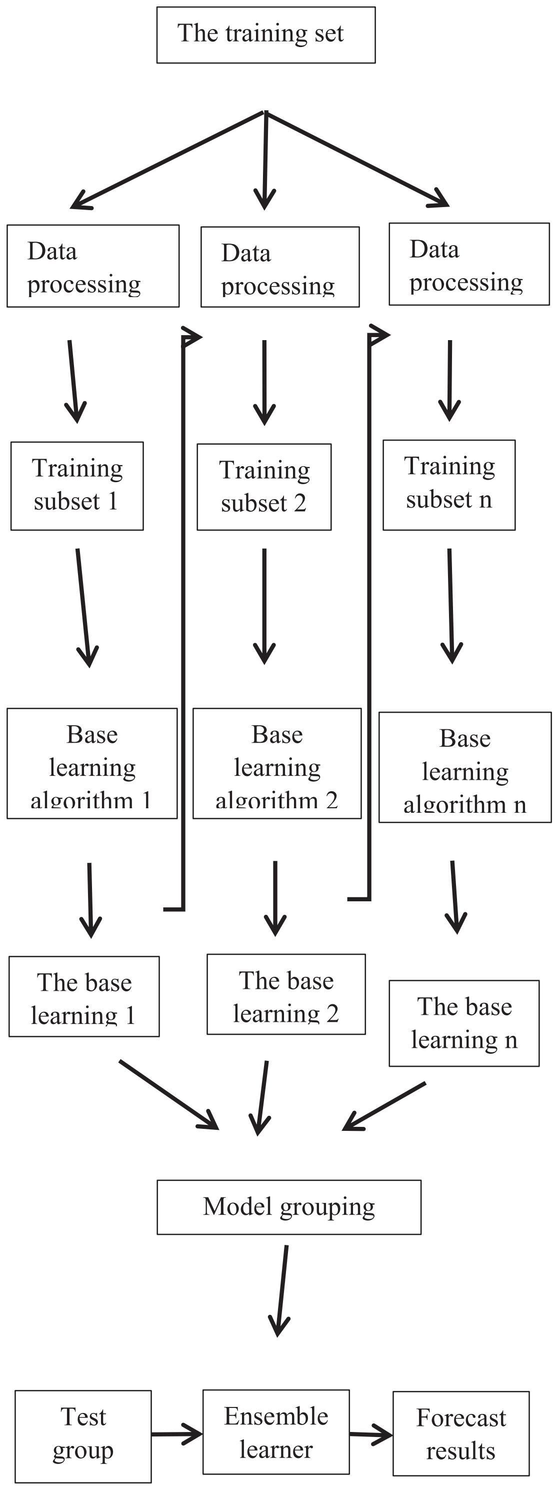

Figure 2 Procedure of Boosting algorithm.

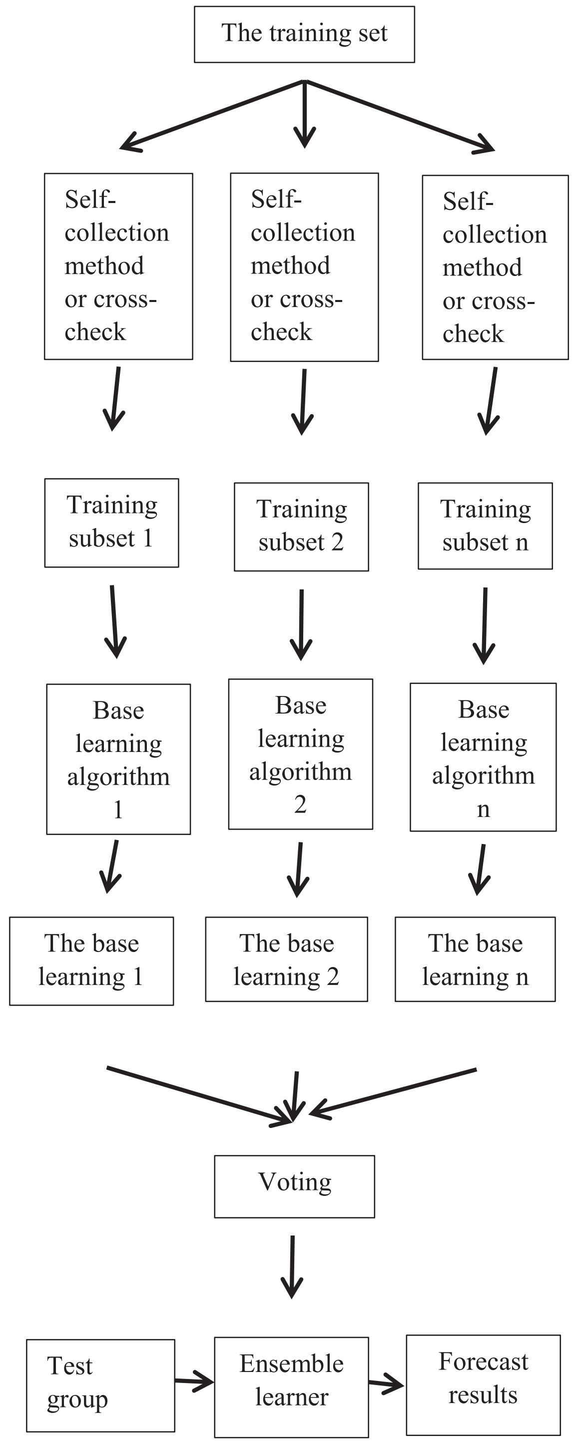

Figure 3 Procedure of Stacking agorithm.

of classification models are trained using these newly generated training sets, which are called Tier1, and then the output of Tier1 is combined to train the meta classifier of Tier2. In addition to the resampling method, cross-validation is often used in training Tier1 classifier, that is, first, the training set is divided into N equal parts, then each individual learner in Tier1 conducts training according to the first n-1 training set, and finally tests on the NTH training set.

The advantages and disadvantages of the three algorithms are analyzed. Bagging, Boosting, and Stacking ensemble learning algorithm, distinctive machine learning problems respectively from different angles, here I analyze and compare three kinds of algorithms to their advantages and disadvantages summarized as follows:

Bagging by resampling method from the original training focus back on the sampling to get multiple training subsets, due to various training subsets are independent of each other, reduces the variance of the base classifier, improved the generalization error, and resampling method can effectively reduce the original training concentrated random fluctuations caused by the error, makes the unstable learning instruments have better learning effect. Because the algorithm in the base learning equal weight, so choice will directly affect the integration of learning results, instability of learning can not only provide a good learning effect, and can generate diversity, depending on the training set so Bagging combined with instability of learning algorithms usually can produce a powerful learning model, and has a good ability to resist noise, and various learning can generate parallel, improve operation efficiency.

Boosting uses the training set of the same in each round training, but the training focus on each sample meeting adjustments according to the previous study results, the new learning in view of the existing study wrong samples to study. This method can significantly improve the learning effect of weak learners, but it is easy to be affected by noise to produce overfitting phenomenon, and each base learner can only be generated sequentially, and the training efficiency is relatively poor.

Stacking, The primary learner is used to generate a new training set to train the secondary learner, but if the primary learner's training set is directly used to generate the secondary training set, there is a great risk of over-fitting. Therefore, cross-validation method is usually used to generate the secondary training set. In this algorithm, the data type of secondary training set and the selection of secondary learner are two key factors. The use of multiple powerful and different primary learners and the use of class-label probability instead of predictive class-label as the attribute of secondary learners will yield better results, and the selection of simple models by secondary learners will reduce the risk of over-fitting.

3. COMBINATION STRATEGY OF BASER LEARNER

3.1 Experience data group

The final step in building an ensemble learning system is to select a combination strategy for individual learners. The ensemble learning system can decide whether to use an individual learner based on its performance.[13]

Voting, simple voting methods are generally divided into the following two types : majority voting and plurality voting. The former was as follws,When the number of votes for a category exceeds more than half of the number of individual learners, the category mark will be output as the prediction result. If there is no class mark with more than half of the votes, the prediction will be rejected.. The later was as follows, The class standard with the largest number of votes obtained is the output of the predicted result, and it is not necessary to consider whether the number of votes obtained exceeds half of the number of individual learners. If multiple class standards all get the highest number of votes, one of them will be randomly selected for output.

In the learning task, if the ensemble learning system must be required to provide prediction results, the absolute majority voting becomes the relative majority voting method, so these two methods are generally referred to as the majority voting method. Because there is a difference between the training method of the learner and the algorithm adopted, the learning ability of the individual learner in the ensemble learning system is also different. If the learning ability of the individual learner can be taken into account in the voting process, the performance of the whole system can be further improved..

3.2 Weighted voting

Firstly, the error of the individual learner is estimated, and then the weight size is inversely proportional to the error size. Finally, the weighted voting results of T individual learners can be expressed by formula as follows:

$$H(x) = \sum_{i=1}^{T} w_i h_i(x) \tag{1}$$

Here W_i means the weight of the i individual learner, W_i and all beyond 0.

3.3 Average method

Average method which contains simple averaging and weighted averaging is a common method to combine continuous numerical output. Simple averaging could be showed by the formula as follows:

$$H(x) = \frac{1}{T} \sum_{i=1}^{T} h_i(x) \tag{2}$$

Here i stands for the individual learners.

The weighted average method combines the simple average method and the weighted voting method. What is different from the weighted voting method is that the weights in the weighted average method are not used for class standards, but directly applied to the actual continuous output values. The weight of the weighted average method can be obtained as part of the training during the generation of the ensemble system, such as the weight generation in the AdaBoost algorithm, or can be obtained through separate training. The weighted average method can be expressed as follows:

$$H(x) = \sum_{i=1}^{T} w_i l_i(x) \tag{3}$$

4. CONCLUSION

The ensemble learning algorithm imitates the behavior of people who seek multiple opinions to assist them in making important decisions. In the late 1970s, researchers in subjects like pattern recognition, statistics, and machine learning began studying ensemble learning methods. With the continuous growth of research enthusiasm and the deepening of ensemble learning research, a variety of ensemble learning methods have been proposed and widely applied in various fields.

Ensemble learning is a machine learning paradigm used to enhance performance results[14]]. Ensemble learning provides solutions to various machine learning problems by combining multiple learners, and its model can solve many problems that cannot be solved by a single model. Since most ensemble learning algorithms have no restrictions on the types of basic learners and are applicable to many mature machine learning frameworks, ensemble learning is also called "algorithmic algorithm without algorithm". Ensembles are known as a mixture of experts to reduce over fitting and errors from all combined base learners and have proved their performance in many real-world applications [15].

There are still many shortcomings and limitations of existing ensemble learning algorithms. For example, if Bagging algorithm is to achieve a good integration effect, the base learner needs to have both efficient learning ability and high data sensitivity. Boosting algorithm in training data with noise easily produced fitting problem. The main drawbacks of extreme learning machine are that it has the random initialization and its prediction precision is very sensitive to the noise and the number of hidden layer nodes, which will lead to a poor robustness.[16] Therefore, ensemble learning needs to be further studied in many aspects, and follow-up research can be carried out from the following aspects: Firstly,ensemble learning structure optimization. The internal structure and external structure of the ensemble learning system are studied to further improve the performance of the ensemble learning system. Second, Integration learning model selection. Select models in an ensemble learning system to remove redundancy and models that have negatively impact towards the results. Finally, the fusion of ensemble learning models. To the unsupervised algorithm, its output results are relatively complex, the model fusion strategy which was suitable for the supervised algorithm could not be used towards unsupervised algorithm.

REFERENCES

1. Cortes C Vapnik V. Support-Vector Networks. Machine Learning 1995, 20(3) 273–297
2. DIETTERICH T G. Ensemble methods in machine learning [J] . Proc International Workshop on Multiple Classifier Systems, 2000, 1857(1) : 1-15
3. KOHAVI R, WOLPERT D. Bias plus variance decomposition for zero–one loss functions [C] . Thirteenth International Conference on International Conference on Machine Learning, 1996: 275-283.
4. Zhou.Z H, Wu. J X, Tang W, Ensemble learning Neural Networks: Mnay Could Be Better Than All. Artificial Intelligence, 2002, 137(1/2):239-263
5. DASARATHY B V, SHEELA B V. A composite classifier system design: Concepts and methodology [C] . Proceedings of the IEEE, 1979, 67(5) : 708-713
6. HANSEN L K, SALAMON P. Neural network ensembles [C] . IEEE Computer Society, 2008: 993-1 001.
7. FREUND Y, SCHAPIRE R E. A decision–theoretic generalization of on – line learning and an application to boosting [J] . Journal of Computer and System Sciences, 1997, 55(1) : 23-37.
8. FREUND Y, SCHAPIRER E. Experiments with a new boosting algorithm [C] . International Conference on Machine Learning, 1996: 148-156.
9. ZHANG C, ZHANG C. HC/Technik/ Sonstiges. Ensemble Machine Learning [M] . US: Springer, 2009.
10. Dietterich T G. Ensemble Methods in Machine Learning / / Proc of the 1st International Workshop on Multiple Classifier Systems. Cagliari Italy 2000 1–15
11. Mason L, Baxter J, Bartlett P L, Frean M. Functional gradient techniques for combining hypotheses. Advances in Large Margin Classifiers. Cambridge, MA: MIT Press, 2000. 221-247
12. YANG Y, CHEN K. An ensemble of competitive learning networks with different representations for temporal data clustering [C] . International Joint Conference onNeural Networks, 2006: 3 120-3 127
13. CAO Y, MIAO Q G, LIU J C, et al. Advance andprospects of adaBoost algorithm [J] . Acta Automatica Sinica, 2013, 39(6) : 745-758
14. WOLPE R T D H. Stacked generalization [M] . US: Springer, 1992: 241-259.
15. KUMARI G T P. A study of Bagging and Boosting approaches to develop meta – classifier [J] . Engineering Science and Technology, 2012, 2(5) : 850-855
16. T. Lofstr ¨ om, On effectively creating ensembles of classifiers: Studies on creation strategies, diversity and predicting with confidence, PhD dissertation, Dept.Comput. Syst. Sci., Stockholm Univers.
17. N. Perez-Diaz, D. Ruano-Ordas, F. Fdez-Riverola,and J. R. Mendez, Boosting accuracy of classical machine learning antispam classifiers in real scenarios by applying rough set theory, Sci. Program., doi:
18. Yunjie WEI, Shaolong SUN, Kin Keung LAI, Ghulam ABBAS, Journal of Systems Science and Information Aug., 2018, Vol. 6, No. 4, pp. 289–301

LI Xiaolong, he graduated from Jiangxi Normal University in 2011 received his mater's degree, now working in Jiangxi Institute of Fashion Technology, lecture, his research interests include education theory, higher education management and education information.

CHEN Juanfen, graduated from Zhejiang Sci-Tech University. She works as a professor in Jiangxi Institute of Fashion Technology. Her current research includes higher education management and curriculum construction.

LIAO Shiqin, was born on 12-DEC-1987, lecture, is currenting working in Jiangxi Institute of Fashion technology, received her master's degree from Jiangnan University in 2012. Her research interest includes textile engineer technology and talent cultivation.