Research on SVM Ensemble and Its Application to Remote Sensing Classification

Hengnian Qi[1,2] **Meili Huang**[1,3]

[1] School of Information Engineering, Zhejiang Forestry University, Lin'an 311300, P. R. China
[2] School of Mechanical Engineering, South China University of Technology, Guangzhou 510641, P. R. China
[3] School of Computer Engineering and Science, Shanghai University, Shanghai 200072, P. R. China

Abstract

The paper analyzes the key concepts, theories and methods of machine learning ensemble, and reviews the related studies on support vector machine (SVM) ensemble. The experiments on the remote sensing classification show that SVM ensemble is more accurate than single SVM. To obtain an effective SVM ensemble, we propose a selective SVM ensemble approach based on fuzzy clustering and discuss the issues on it.

Keywords: Support vector machine, Ensemble, Remote sensing classification, Fuzzy clustering

1. Introduction

Support vector machine ensemble is a classification method which includes two procedures: Firstly, it gives the sub-forecasts of a new sample using sub-SVM classifiers that can be obtained by various ways, and then, combines these sub-forecasts to decide the final class[1].

Theoretically, compared with the single SVM, the SVM ensemble method has better classification accuracy. Supposing there are n sub-classifiers and a sample to be classified, it is obvious that the SVM ensemble has the same classification ability if all sub-classifiers are equal, else if they are different and uncorrelated then the SVM ensemble outperforms the individuals. Actually, the theory is also suit to any other learning mechanism (strictly, weak learning mechanism). Since ensemble learning method behaves remarkably well, it has been a hot topic in academic circles in recent years.

The paper is organized as following: in section 2, the key concepts, theories and methods of machine learning ensemble are analyzed. In section3, the related studies on SVM ensemble are reviewed and experiments on the remote sensing classification data sets are presented. In section4, a selective SVM ensemble approach based on fuzzy clustering is proposed and issues on SVM ensemble are discussed.

2. Ensemble of classifier

Given a set S of training examples, a learning algorithm outputs a classifier which is an hypothesis about the true function f can make classification-decisions from given examples. We denote classifiers by $h_1,h_2,\cdots,h_L$.

An ensemble of classifiers is a set of classifiers whose individual decisions are combined in some way (typically by voting) to classify new examples. Presently, one of the most active research areas in supervised learning is method for constructing good ensemble of classifiers. The main discovery is that ensembles are often much more accurate than the individual classifiers that make them up. In 1990, Hasen & Salamon pointed out a necessary and sufficient condition for an ensemble of classifiers to be more accurate than any of its individual members is if the classifiers are accurate and different. Accurate, here, means that the classifier has a better error rate than random guessing, and it requires the error rates of individual classifiers below 0.5. Two classifiers are diverse if their error rates are uncorrelated, or they get different classification results on samples which may be because of their different learning mechanisms, training samples, or other factors.

By giving three fundamental reasons, Thomas G. Dietterich points out that it is possible to construct good ensembles[3]:

- Statistical：A learning algorithm can be viewed as searching a space H of hypotheses to identify the best hypothesis in the space. If the amount of training data available is too small compared to the size of the hypothesis space, the learning algorithm would find many different hypotheses in H that all give the same accuracy on the training data and no one is the best hypothesis. However, a good hypothesis about the true function can be received by constructing an ensemble out of all of these classifiers.

- Computational：Many learning algorithms work by performing some form of local search that may get stuck in local optima. For example, neural network algorithms employ gradient descent to minimize an error function over the training data. Unlike the first situation, suppose there is enough training data, it may still be very difficult computationally for the learning algorithm to find the best hypothesis, indeed, it is NP-hard. So, an ensemble constructed by running the local search from many different starting points may provide a better approximation to the true unknown function than any of the individual classifiers.
- Representational：In most applications of machine learning, the true function f cannot be represented by any of the hypotheses in H. By forming weighted sums of hypotheses drawn from H, it may be possible to expand the space of representable functions.

These three fundamental issues are the three most important ways in which existing learning algorithms fails, but ensemble methods have the promise of reducing these shortcomings. Because of this, research on ensemble learning mechanism is prospective. Many of methods for constructing ensembles have been developed[3], including: Bayesian Voting, Manipulating the Training Examples, Manipulating the Input Features, Manipulating the Output Targets, Injecting Randomness and so on. To construct a good ensemble, it needs the sub-classifiers differentiate greatly. Being the main thought of manipulating the training examples, the differentiation can be realized by using different training samples on the same learning method. There are some familiar methods:

- Bagging

Bagging was first put forward by Leo Breiman in his technical report titled "Bagging Predictors" [4] in 1994. this methods main idea is: given a weak learning algorithm and a training set $(\boldsymbol{x}_1, y_1), \cdots, (\boldsymbol{x}_n, y_n)$, each time randomly select m training samples from the training set and put back to it after training, then got a forecast function list $h_1, h_2, h_3, \cdots, h_T$, and the final forecast function H can received by equal weight revoting method to classification problems and average value of votes to value type problems.

- AdaBoost

Same as Bagging, each sub-classifier of AdaBoost was received by using different training samples, which difference is every subset of training samples was decided by the performance of its preceding subset. AdaBoost is the representative algorithm of Boosting, following is its Pseudo-code：

Input: $((\boldsymbol{x}_1, y_1), \cdots, (\boldsymbol{x}_m, y_m))$ presents m numbers of labeled samples, where $\boldsymbol{x}_i \in X$, $y_i \in Y = \{1, 0\}$; D is the distribution of samples; weak learning algorithm WeakLearning; trials T.

Initialization: $D_1(i) = 1/m$,

for each $t = 1, 2, \cdots, T$ do following steps：

Step 1 utilizing distribution D_t to train weak learner;

Step 2 obtain weak hypothesis $h_t : X \to R$;

Step 3 select $\alpha_t \in R$;

Step 4 modify：

$$D_{t+1}(i) = \frac{D_t(i)\exp(-\alpha_t y_i h_t(x_i))}{Z_t}.$$

Where Z_t is normalization factor, D_{t+1} is distribution function;

Output: $H(x) = sign(\sum_{t=1}^{T} \alpha_t h_t(x))$.

- Random

It's the simplest method of manipulating the training examples which constructs the training sets by dividing the training data into disjoint subsets randomly and trains one sub-classifier from each subset.

3. Ensemble of SVM

In recent years, a lot of researchers pay much attention to SVM ensemble. Examples are as follows, SVM ensemble based on Improved Bagging, which improved the classification rate compared to Bagging and Boosting , was put forward by Tammy Williams in 1999[5]; Rong Yan & Yan Liu applied the SVM ensemble method to scene classification and compared some of the methods in 2003[6]; Chuhong Hoi & Michael R. Lyu provided an algorithm called group-based relevance feedback with SVM Ensembles in 2004 and successfully applied it t o CBIR[7]; At the same year, Ling Wei & Wenxiu Zhang proposed SVM ensemble based on majority voting mechanism, then did simulate classification experiment on data set Hepatitis and Ionosphere of UCI benchmark database, it found out the error rate lowered 10% averagely compared to single classifier [1] .

As we can see, approaches of neural network ensemble are useful to study SVM ensemble approaches.

3.1. Neural network ensemble

Neural network, which has been experienced several dozens years of development since it was presented, has obtained great success thanks to its widely application. Researches show that only a Single layer feed-forward network is needed to approach to any complicated function, however, the network architecture is a NP problem for the absence of the prior knowledge in practice. As a result, it influenced the network generalization ability directly.

Neural network ensemble, originates from Hansen & Salamon’s work in 1990 [2], shows to be a simple and realizable approach to improve the generalization ability of a neural network system.

In general, a neural network ensemble is constructed by training a number of component neural networks and then combining the component predictions by weighted averaging or simple averaging. Some researchers recommend simple averaging instead of weighted averaging since it is easy to suffer over-fitting. However, the differences among the neural networks which obtained by simple averaging were guaranteed by the training sets, network architecture, differences or randomness among learning algorithms. Therefore, it is natural that there are very similar neural networks. But ensemble of them maybe couldn’t lower the generalization error. Considering this issue, Zhihua Zhou etc.[8]-[9] indicated and proved that ensemble of many of the available neural networks(select those who satisfies certain conditions) has better generalization ability than that of all of those networks, i.e. many could be better than all. Based on it, Zhihua Zhou presented GASEN, a selective ensemble approach based on genetic algorithm.

The theory of ensemble technology is established on the foundation that the individual learners are different. GASEN is a kind of ensemble algorithm by utilizing genetic algorithm to select individuals. So, we can design an algorithm to replace genetic algorithm to eliminate redundancy individuals.

Clustering is a process of organizing objects into groups whose members are similar in some way. It can broaden the object differences by using a single object to represent the group. So Guozheng Li etc. [10] proposed CLUSEN (clustering algorithm based selective ensemble) by applying artificial neural network to multi-class classification tasks. Experiments on UCI database show that compare to GASEN, CLUSEN has same classification accuracy but its efficiency is 25 to 100 times higher than that of GASEN.

3.2. Remote sensing classification based on SVM ensemble

Whether based on intuitively speaking or strictly inferring, we have the full confidence to believe that SVM ensemble is a good choice to improve the classification rate of the single SVM classifier. So, in this research, we do experiments that applied the SVM ensemble to the remote sense classification in which field mostly using classification methods based on pixel. The TM remote sensing image of Yuyao city of Zhejiang province using in our experiment, which contains an area of 408×356 pixels(see Fig.1), is taken by Lansat-7. Representing a pixel by a six dimension vector for it is the dimension commonly used in the ground mulch classifications, and the six dimensions are 1 to 5 and 7 of the eight wave bands. The space resolution is 30m.

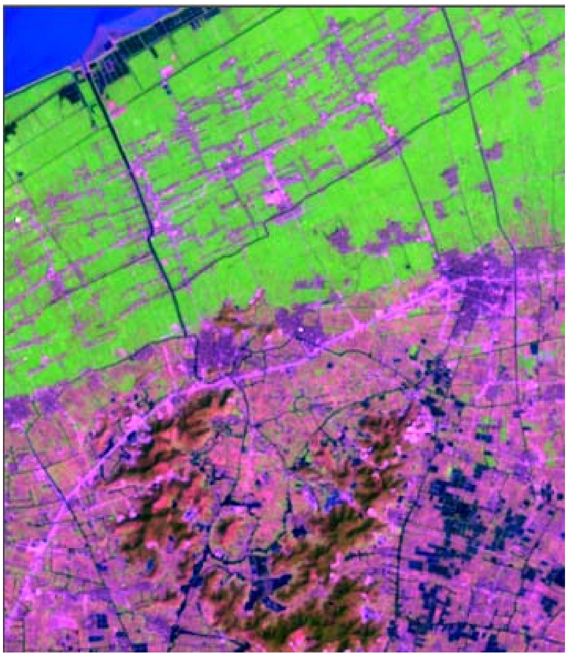

Fig.1: The TM subset image（Band5,4,3）in the experiment

Since SVM is a supervised classification approach, it needs to make sure the number of classes and the labeled training samples. Two experiments were done. One is training and classifying with 1-a-1 multi-class classifier. The other is the ensemble of the first one’s results by Bagging, in which used voting strategy and each group have eleven objects. Results of the experiments show in Table 1. In the experiments, we classify the test area’s mulches into 27 classes by referring to the corresponding 1:50000 scale map, 200 labeled samples (100 for training and others for testing) are chosen from each class. The third and fifth polynomial kernel functions are used respectively.

We can see from Table1 that classification rates of all SVM ensembles are above 90%, which are higher about 2~3% averagely compare to SVM. The classification accuracy not broadened much may be because of three reasons. First, the random noises are accompanied in the process of taken remote sensing images and image has the characteristic of “same object but different spectrum” or “different object but same spectrum”. Second, the ensembles may not the most optimum ones since it is done on the subjectively chosen 11 SVM classifiers. Third, The multi-class classifier is based on the 1-a-1 combination of bi-class classifiers, in which the predict results are obtained by voting of 27×26/2, i.e. 351 of BSVM classifiers[11], so it guaranteed the relative stability and in certain meaning can be seen as an ensemble.

kernel function	C	SVM classification rate (%)	SVM ensemble classification rate (%)
second order polynomial	10	88.63	90.61
third order polynomial	10	89.78	91.83
third order polynomial	5	89.52	92.72
fifth order polynomial	5	88.63	90.86

Table 1 Results of TM classification experiment based on SVM

4. Selective SVM ensemble based on fuzzy clustering algorithm

Based on GASEN and CLUSEN, we provide a selective SVM ensemble approach using fuzzy clustering algorithm which is realized in three steps: firstly, train a number of SVMs from the training set with Bootstrap, then, obtain the selected SVMs using clustering algorithm to select objects near to each clustering center, finally, construct the ensemble with the selected SVMs through voting. The group number of clustering is that having highest testing precision. The algorithm can be described as following:

Input： training set S,testing set T,parameters of SVM, such as kernel function, penalty coefficient, etc；

Initialization：set the number of training samples n＝N,i＝1,j＝1；

Step1 if i<=N then

Establish sub-set Si with Bootstrap algorithm;

Train SVMi;

Let i=i+1;

Else

Goto Step 2；

Step2 if j<=N then

Execute the FCM algorithm to cluster the SVMs into j groups based on the differences of the classification results by the SVMs;

Choose one SVM which is nearest to the group center from each group to obtain j SVMs for constructing the ensemble;

Construct the ensemblej by voting;

Calculate the classification rate Rj of the ensemblej;

Let j＝j＋1；

Else

End

Output：Choose the ensemble with biggest Rj as the selected ensemble, j denotes the number of SVMs.

5. Conclusions

Worth to mention, the number of SVMs in our algorithm can be adjusted according to the complexity of the task or the scale of training sets.

Generally, ensemble means to integrate homogenous sub-classifiers and in our algorithm it isn’t an exception. However, the clustering rule in our algorithm is the differences of the sub-classifiers shown by the classification results obtained from the same testing set. In fact, it can support ensemble of heterogeneous sub-classifiers. If the sub-classifiers are trained with different learning models, to a certain extent, the process of constructing ensemble can be seen as an information fusion process.

As we know, classification experiments in the paper based on SVM ensemble is just a probe and it needs to be further systematically experimented. The work we mentioned is just a start of our study on SVM ensemble, so the validity of the selective SVM ensemble based on FCM needs to be further experimented. Meanwhile, the possibility of heterogeneous sub-classifier ensemble and information fusion also needs to be validated by experiments.

References

[1] L. Wei and W.X. Zhang, Classification Based on SVM Ensemble,*Computer Engineering*,30: 1-2,2004.

[2] L.Hansen and P. Salamon, Neural network ensembles. *IEEE Trans. Pattern Analysis and Machine Intell.*, pp. 993-1001, 1990.

[3] T. G. Dietterich, Ensemble Methods in Machine Learning. In J. Kittler and F. Roli (Ed.) *First International Workshop on Multiple Classifier Systems, Lecture Notes in Computer Science* . New York: Springer Verlag, pp. 1-15, 2000

[4] L. Breiman, Bagging Predictors. *Technical Report 421*, University of California, Berkeley, Sep, 1994.

[5] D. Grossman and T. Williams, Machine Learning Ensembles: An Empirical Study and Novel Approach. University of Washington, Seattle, WA 98195 CSE 573 -- Artificial Intelligence -- Autumn 1999 TERM PROJECT #2. Available from http://www.danielgrossman.org/projects/573projects /learning/.

[6] R. Yan, Y.Liu, R. Jin, et al, On Predicting Rare Classes with SVM Ensembles in Scene Classification, *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'03)*, Apr. 2003.

[7] C.H. Hoi and M.R. Lyu., Group-based Relevance Feedback with Support Vector Machine Ensembles, *Proceedings of the 17th International Conference on Pattern Recognition (ICPR 2004)*, Vol 3, pp. 874- 877, Aug. 2004.

[8] Z.H. Zhou, J. Wu, and W. Tang, Ensembling neural networks: many could be better than all, *Artificial Intelligence*, pp. 239-263, 2002.

[9] J.X. Wu, Z. H. Zhou, X. H. Shen, et al,A Selective Constructing Approach to Neural Network Ensemble,*Journal of Computer Research & Development*,37: 1039-1044, 2000.

[10] G.Z. Li and J. Yang, A. S. Kong, et al,Clustering algorithm based selective ensemble, *Journal of Fudan University(Special Issue on CCML2004)*, 43:689-691, 2004.

[11] H.N. Qi, J.G. Yang, Y. W. Zhong and C. Deng,Multi-class SVM Based Remote Sensing Image Classification and Its Semi-supervised Improvement Scheme, *Proceedings of 2004 International Conference on Machine Learning and Cybernetics*, 5: 3146-3151, 2004.