

Subspace Selective Ensemble Algorithm Based on Feature Clustering

Hui Tao^{1,2}

1. School of Information and Electrical Engineering, China University of Mining and Technology, Xuzhou, China
 2. School of Electrical Engineering and Automation, Henan Polytechnic University, Jiaozuo, China
- Email: taohui@hpu.edu.cn

Xiao-ping Ma¹, and Mei-ying Qiao^{1,2}

1. School of Information and Electrical Engineering, China University of Mining and Technology, Xuzhou, China
 2. School of Electrical Engineering and Automation, Henan Polytechnic University, Jiaozuo, China
- Email: Xpma@cumt.edu.cn, and qiaomy@foxamil.com

Abstract—A feature-clustering-based subspace selective ensemble learning algorithm was proposed to improve ensemble classifier performance, allowing for high dimensional data sets. First, features were clustered on *weighted average linkage* method and reduced subspaces were generated by extracting an attribute from each feature cluster. Then the feature reduced subsets served as inputs of individual GA-SVMs which had high accuracy to ensure individuals with significant diversities. Some individuals with both diverse and accurate were selected to construct ensemble system. Finally, In Matlab 2010a environment, the algorithm was simulated on 4 datasets. The *kappa-error* diagrams demonstrated that individual classifiers were both accurate and diverse, and the results showed the classification accuracy increase significantly.

Index Terms—ensemble learning; subspaces ensemble; feature clustering; selective ensemble; SVM; GA

I. INTRODUCTION

Combining recognition outputs from multiple individual classifiers, known as ensemble learning algorithm, can improve classification accuracy and enhance robustness of classification system, and have become one of research focuses in pattern recognition and knowledge mining fields [1-3]. Theoretical and empirical results suggest that ensemble classification system gives optimal improvements in accuracy if the base classifiers are both diverse and accurate [4-6].

Sample perturbation is the most popular technique to construct diverse individual classifiers and there are two methods of sample perturbation. One is to train classifiers with distinct sample subsets of the same feature space and the other is to construct classifiers in different feature subspaces [7]. Bagging and Boosting are the most popular algorithms of ensemble learning using distinct sample subsets. In bagging algorithm, the training sets randomly sampled with replacement are used to train classifiers and different predictors are obtained, which can be used to classify new data. Here, the sample sets are independently extracted and the base predictor can be constructed simultaneously [8]. Boosting, on the other

hand, must sequentially train the base classifier by adaptively changing the distribution of the training set based on the accuracy of the previously created classifiers and weight to selection of training examples based on classifier performance [7]. However, in Bagging and Boosting method, on high dimensional data sets, over-fitting phenomenon happens and the calculation process is complex. Combining classifiers in different feature subspaces, named ensemble feature subspaces (EFS), is a favorite with many researchers [9] because it can lessens the impact of the “curse of dimensionality” and speeds up the training process on high dimensional data sets. Ho presented a novel method for constructing decision forests with random subspace techniques [10]. Bryll proposed attribute bagging (AB) to improve the accuracy and stability of classifier ensembles, which establishes an appropriate attribute subset size and then randomly selects subsets of features [11]. Akay presented a SVM combined with feature selection for breast cancer diagnosis [12].

Furthermore, researches showed that a set of attribute subset with good predicting power would significantly improve the efficiency and accuracies of EFS. Hu and Shi raised an ensemble rough subspace (EROS), which rough set based attribute reduction was introduced to generate a set of reducts to train a base classifier [13,14]. Attribute clustering approach can reduce and select attributes and therefore obtain attribute reduction subsets on high dimensional data sets [15]. So, in our paper, we propose a novel feature subset selection algorithm based on feature clustering, which the base classifier is trained according the reduced feature subsets that are different from one another through feature clustering and selecting.

In general, the category number of features can't be predetermined, so we first use hierarchical clustering methods to cluster feature. Then, the feature clusters will be determined according clustering tree diagram, and the reduced feature subsets are generated by extracting an attribute from each feature cluster. Moreover, SVM is using as the individual classifier to ensure accurate of base classifier. SVM can gain expected classification accuracy only when its parameters (kernel parameter c

and penalty coefficient g) are properly adjusted in classifying, so GA is used to optimize individual SVM parameters. In addition, the ensemble performance will be enhanced when parts of individual models are integrated instead of all ones [16], so selective ensemble learning is adopted that parts of all are combined on the basis of their evaluation results.

II. SUBSPACE SELECTIVE ENSEMBLE ALGORITHM BASED ON FEATURE CLUSTERING

A. Basic Idea

Subspace ensemble algorithm based on feature clustering is a multiple classifier system, in which base classifiers are trained with feature-clustering-based reduced subspace, allowing for high dimensional data sets. The algorithm flow is shown in Figure 1. There is some work to do for constructing such an ensemble system. First, an algorithm for feature clustering is required. Second, multiple reduced feature subsets should be generate through extracting from feature space. Third, Base classifiers and its train algorithm with reduced feature subsets should be selection. Finally, for constructing an efficient and powerful ensemble system, a strategy of evaluating and selecting base classifiers is indispensable.

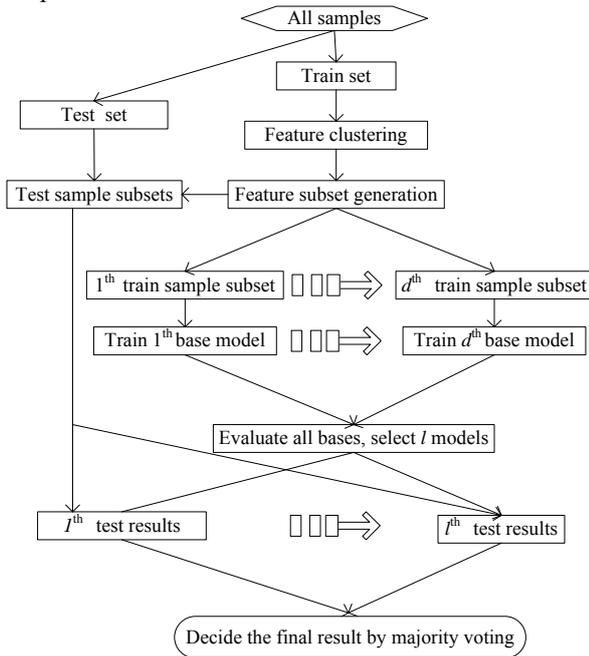


Figure 1~ Algorithm flow of subspace ensemble based feature clustering.

B. Feature Clustering

There are three common feature clustering methods including *hierarchical* clustering, *K-means* clustering, and *fuzzy c-means* clustering [17]. And in the last two methods, category amount must be determined in advance. Normally, the features category number can't be predetermined, so we use *hierarchical* clustering methods to cluster feature. The common heuristics for *hierarchical*

clustering work *bottom-up*, starting with a separate category for each feature, and then progressively merging the two “closest” categories until only one category remains.

The distance between two features

Before feature clustering, the distance between features must be defined. And in the paper, we use *correlation coefficient* define the distance between two features. Suppose

$$x_i = (x_{i1}, x_{i2}, \dots, x_{in}) \quad (1)$$

Where x_i is a feature, and n is the number of the samples.

Then, the *correlation coefficient* between x_i and x_j is defined as ρ_{ij} (2).

$$\rho_{ij}(2) = \frac{\sum_{k=1}^n (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j)}{\sqrt{\left[\sum_{k=1}^n (x_{ki} - \bar{x}_i)^2 \right] \left[\sum_{k=1}^n (x_{kj} - \bar{x}_j)^2 \right]}} \quad (2)$$

Where \bar{x}_i is the mean of x_i , \bar{x}_j is the mean of x_j . And

$$\bar{x}_i = \frac{1}{n} \sum_{k=1}^n x_{ki} \quad (3)$$

$$\bar{x}_j = \frac{1}{n} \sum_{k=1}^n x_{kj} \quad (4)$$

The distance between x_i and x_j is defined as d_{ij} .

$$d_{ij} = 1 - |\rho_{ij}| \quad (5)$$

Weighted average linkage clustering

There are different hierarchical schemes distinguished by their notion of closeness [17, 18]. In *single-linkage* methods, the distance between two clusters is the distance between their closest pair of feathers. In *complete-linkage* methods, it is the distance between their farthest pair of feathers. In *average-linkage* clustering, it is the distance between their means. For *average-linkage* clustering which takes full advantage of the information of all feathers is preferable method, we use *weighted average linkage* method, a variants for *average-linkage* to cluster feature.

The distance between cluster G_K and G_L is defined as

$$D_{KL} \cdot D_{KL}^2 = \frac{1}{n_K + n_L} \sum_{x_i \in G_K, x_j \in G_L} d_{ij}^2 \quad (6)$$

Where n_K is the feature size of G_K , and n_L is the size of G_L .

The recursion formula of distance between the new cluster G_M and one cluster G_J is defined as D_{MJ} .

$$D_{MJ}^2 = (1 - \beta) \left[\frac{n_K}{n_M} D_{KJ}^2 + \frac{n_L}{n_M} D_{LJ}^2 \right] + \beta D_{KL}^2 \quad (7)$$

Where n_M is the feature size of G_M , $\beta < 1$ and in this paper β is 0.25.

C. Feature Subset Generation

In this paper, the goal of feature cluster is seeking suitable reduced feature subspaces which consist of features successively extracting from each cluster. We expect that in each cluster the features should be redundant, but for they all describe certain a system characteristics, they are able to distinguish them from the features of other subspaces. If the cluster number is too little, the extracted features of each base classification cannot describe the system well. While the clusters number is too big, the extracted features cannot be independent. In above two cases the classification performance of base classifiers is poor. Thus, the clusters number determination (attribute subset sizes) is a key work. In [11], the best classification accuracy is achieved for attribute subset sizes between 1/3 and 1/2 of the total number of attributes on a hand-pose recognition dataset, but the best clusters number requires fully research on different datasets.

Clustering tree diagram is gotten through hierarchical clustering feature, and in the diagram the distances between various features are clear. We can subjective give a distance threshold λ_k between clusters according to clustering tree diagram to cluster features, but it can't ensure a good classification effect. The set of λ_k will be discuss in simulation experiments.

Once one feature is randomly extracted from each cluster, then a feature subspace is generated which consists of M extracted features. When we repeat k times, k feature subsets are obtain. Because the features are randomly selected, the feature subspaces are different from themselves. Though, comparing to attribute bagging methods, in our methods the feature subspaces diversity is less, the accuracy of base classifier is much higher simultaneously and the number of required base classifier is also less.

D. Base SVM Training

Support Vector Machine (SVM) proposed by Vapnik and his co-workers and has been successfully applied to real world [19-22]. Unlike traditional classifiers, SVM can overcome "curse of dimensionality" and have high classification performance with many input features. So, SVM is using as the individual classifier to ensure accurate of base classifier.

SVM can gain expected classification accuracy only when its parameters (kernel parameter c and penalty coefficient g) are properly adjusted in classifying, and though *grid* search can succeed in the optional (c and g) selection through *cross validation*, sometimes it is very time-consuming in a wide search range. However, genetic algorithm (GA), as a heuristic algorithm that can find the global optimal solution even not traversing all mesh point, is a practical and efficient optimization technology with strong robustness [23,24]. So GA is used to optimize individual SVM parameters.

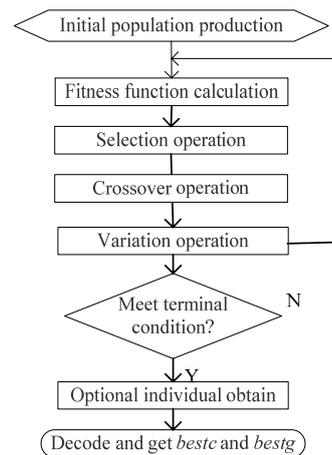


Figure 2- Algorithm flow of GA to optimize base SVM.

In genetic algorithm, SVM parameters are coded by binary system that $c \in (0,100]$ and $g \in [0,1000]$, fitness function is designed as average classification accuracy on K -fold cross validation that after the training data are divided equally into K groups we take each subset data as validation set, other $K-1$ subset as train set, then K models are established, fitness function is the average of classification accuracy of K models validation set.

$$f(X) = \frac{1}{K} \sum_{k=1}^K f_1(k) \tag{8}$$

Where $f_1(k)$ is the classification accuracy of the k th model.

In the paper, other parameters of GA are set as follows. The max evolution population is 100, population size is 20, K is 5, the crossover probability is 0.8 and the Variation probability is 0.1

E. Selective Ensemble Learning

Though there is some diversity among base SVMs with different feather subsets and identifying accuracy of individual SVMs is greatly increased by GA optimizing, there is no guarantee that all individuals are both diverse and accurate.

There are many measure methods of diversity in the ensemble [25-27], and *kappa* statistic is use to measure the diversity between pairings of base classifiers on the test data. The *kappa* statistic is defined as follows [7]. Suppose there are P classes, and let N be an $N * N$ square array such that N_{ij} contains the number of test examples assigned to class i by the first classifier and into class j by the second classifier, then define Θ_1 as an estimate of the probability that the pairing of classifiers agree and Θ_2 as a measure of the probability that the two classifiers agree by chance.

$$\Theta_1 = \frac{\sum_{i=1}^P N_{ii}}{m} \tag{9}$$

$$\Theta_2 = \sum_{i=1}^p \left(\sum_{j=i}^p \frac{N_{ij}}{m} * \sum_{j=i}^p \frac{N_{ji}}{m} \right) \quad (10)$$

Where m is the total number of test examples. $\sum_{j=i}^p \frac{N_{ij}}{m}$ is the fraction of examples that the first classifier assigns to class i , and $\sum_{j=i}^p \frac{N_{ji}}{m}$ is the fraction of examples that the second classifier assigns to class i .

With above definitions, the kappa statistic is calculated as

$$k = \frac{\Theta_1 - \Theta_2}{1 - \Theta_2} \quad (11)$$

$k = 0$ when the agreement of the two classifiers equals that expected by chance, and $k = 1$ when the two classifiers agree on every example. Negative values occur when agreement is less than expected by chance—that is, there is systematic disagreement between the classifiers.

According to [16], many could be better than all in ensemble learning, so selective learning is adopted that some individuals which are both diverse and accurate was selected to construct ensemble system. Base SVM diversity and accuracy are calculated on validation dataset and selective learning procedure is shown as follows.

- Compute all kappa value between pairings of base classifier and average identifying accuracy of base classifiers with validation samples.
- Give a threshold λ_k , remove one SVM with lower accuracy from each pairs of base classifier with kappa value above λ_k .
- Sort individual models from great accuracy to the little.
- Select the first l individuals to generate ensemble model.

III. EXPERIMENTAL SIMULATIONS

A. Data

Four most commonly binary class data sets with high dimension features from the University of California at Irvine (UCI) Machine Learning Repository are used. The detail information of experimental data set is shown in Table I[28].

TABLE I.
DETAIL INFORMATION OF EXPERIMENTAL DATA SET

Data Set	Instances size	Attributes size	Train set size	Validation set size	Test set size
Sonar	208	60	70	68	70
Ionosphere	351	34	151	100	100
Wdbc	569	30	250	159	160
SPECTF	267	44	80	87	100

B. Simulation and Analyse

The attribute clustering tree diagrams using weighted average linkage clustering are shown in Figure 3-6.

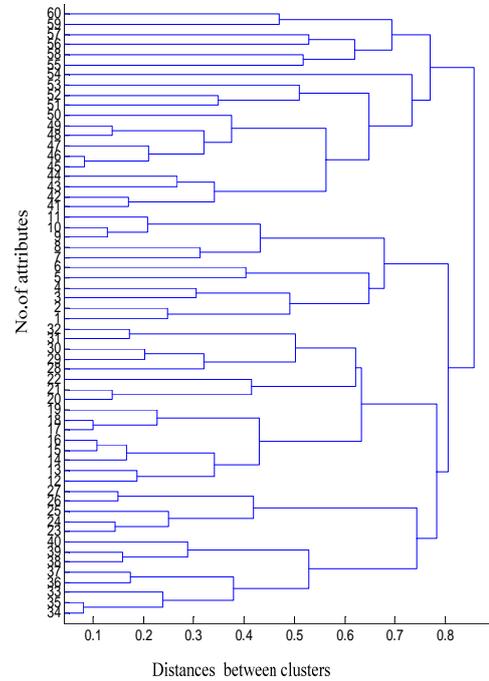


Figure 3. Attribute clustering tree diagram on **Sonar** data set

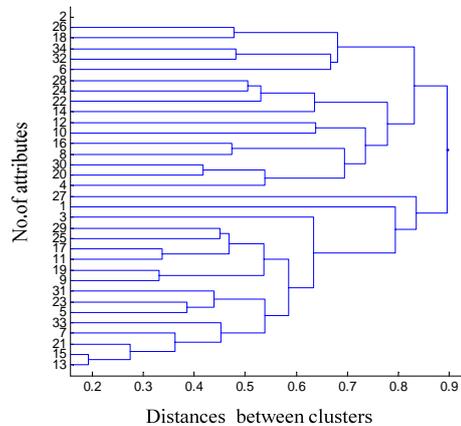


Figure 4. Attribute clustering tree diagram on **Ionosphere** data set

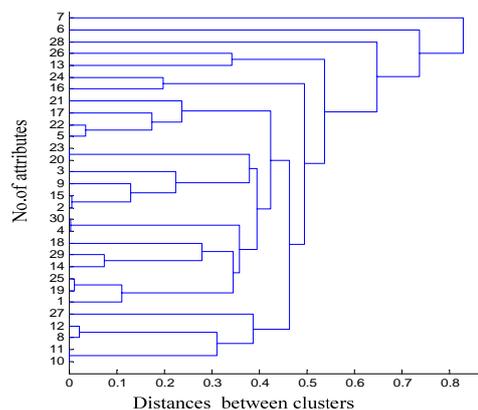


Figure 5. Attribute clustering tree diagram on **Wdbc** data set

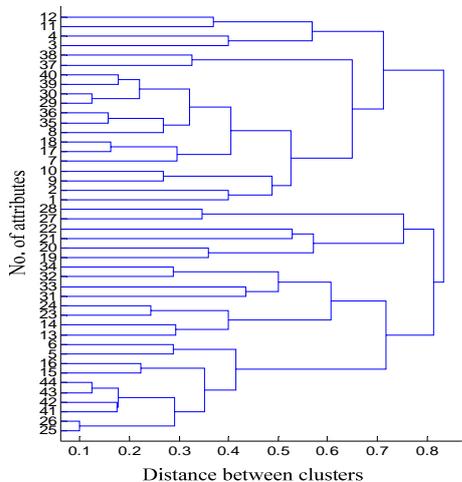


Figure 6. Attribute clustering tree diagram on SPECTF data set

It is clear the distances between clusters are different because the correlation of features is diverse on different data set. There are some redundant features on Wdbc data set, while on German credit data set the distances between feature clusters are far and there are strong independence among features. The cases of Sonar and Ionosphere data sets are in the median.

First all samples in validation set and test set taken as test samples, all 50 base SVMs combined, average classification accuracies of 10 calculations on different data set are shown in Figure 7 where dotted lines show the average of base SVMs and solid lines show ultimate ensemble accuracy of classification system. On Wdbc data set, because the features are redundant, the average classification accuracy of SVMs with parts of features is relatively high, but ensemble accuracy is significant higher than SVMs accuracy. While on Sonar data set, though SVMs average accuracy is only a bit more than 50%, ensemble classification system is capable of very high precision. Similarly, the ensemble accuracy is very high on SPECTF and Ionosphere data set. In all, whether the base classifier accuracy is high or not, the ensemble system can achieves very high accuracy.

Figure 7 also illustrates that average accuracy of base classification-SVM is relatively stable with the change of feature subspace size unless the size is too small. This is because SVM can overcome the impact of “curse of dimension”, namely, SVM predict precision will not descend even though the dimension of SVM input attribute is very big and the attributes are redundant. However, superabundant input attributes must add learning complexity of model, and when other classification such as Neural Network with redundant inputs used, overfitting is happen, then the predict precision will decline. In addition, Figure 7 suggests that ensemble accuracy will no longer be improved significantly with the increase of feature subspace size when the size above N , where N is a threshold value of feature subspace size, for instance, N is 24 on Sonar data set, and N is 11 on Wdbc data set. In fact, when

increasing the number of base SVMs with small amount of feature size, ensemble accuracy can, to some extent, enhance. Instead, ensemble accuracy will fall when feature subspace size is too big. To sum up, under the same ensemble accuracy, the less input attributes of base classification, the better of ensemble system.

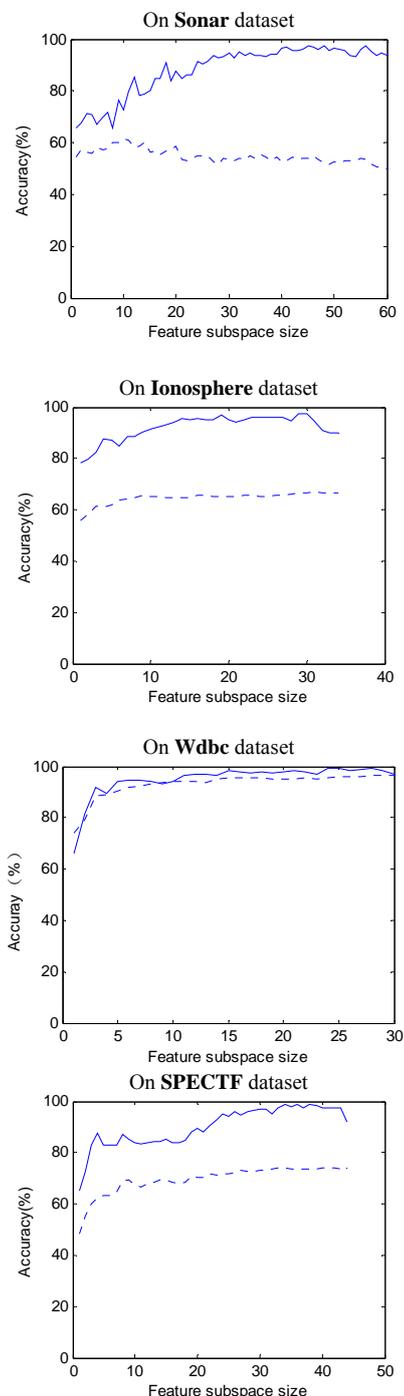


Figure 7. Accuracy with different feature subspace size when 50 SVMs all combined on different datasets

Combing Figure 3-6 and Figure 7, we research the correspondence between the high ensemble accuracy and the least feature subspace size (clusters number of attributes) and find a regulation that when let threshold

$\lambda_k = 0.35 \sim 0.5$ then the feature subspace size is bigger than 1/3 of the total of all features, combining classification has high accuracy.

In our algorithm, we use SVM, a stable classification, as individual to construct a collection system for SVM has high generalization capability and can overcome “curse of dimension”, experiments showed the ensemble classification performance is very strong. When we use unstable classification such as neural network or decision tree as individual, the ensemble effect is very good in the same way when feature subspace size (feature cluster) suitable, and comparing with other algorithms, the improvement is much significant. In addition, the ensemble effect becomes better with the increase of individual number and the smaller feature subspace size, the more individual number is required.

The above simulation results are obtained under the condition of all base SVMs combined. Next we discuss the effect of selective ensemble learning. There are two methods, one is selecting the SVMs with higher accuracy but diversity does not taken into account, and the other is selecting the SVMs with both higher accuracy and larger diversity. Though experiment we find that using the first selecting methods, the ensemble accuracy is slight higher than that of all SVMs ensemble when part of SVMs with lower precision are excluded. In the second method, one SVM with lower accuracy from each pairs with lower kappa value is removed then the first l individuals with high accuracy are selected to generate ensemble model, the ensemble performance is improved significantly.

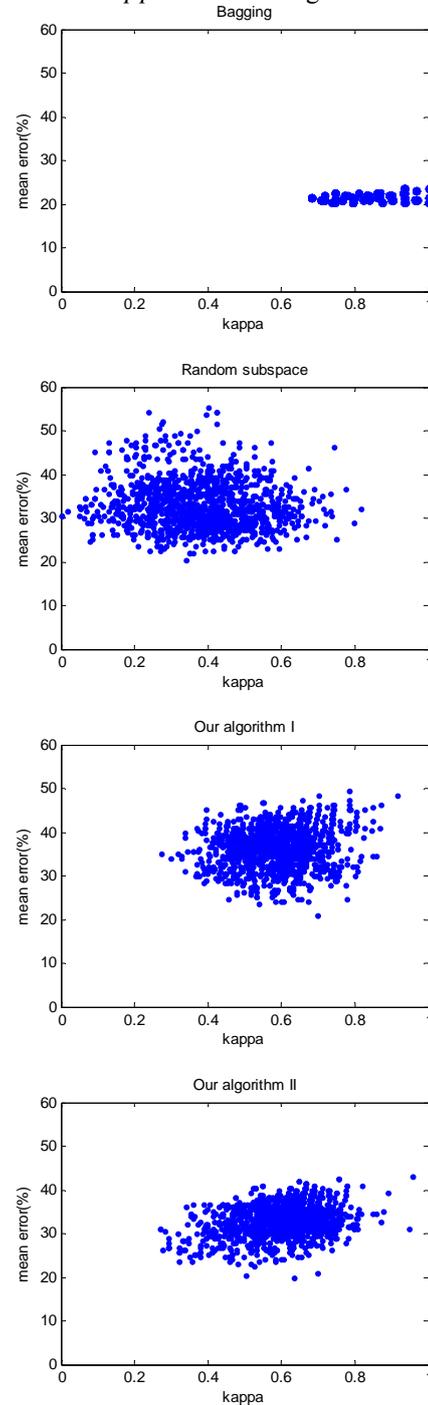
TABLE II.
CLASSIFICATION ACCURACY (%) OF DIFFERENT ENSEMBLE ALGORITHM

Dataset	Sonar	Ionosphere	Wdbc	SPECTF	
Single SVM	71.59	90.09	94.67	76.46	
Bagging	76.14	92.07	95.07	84.06	
Random subspace	85.65	93.53	96.32	87.65	
Our algorithm	I	93.25	95.80	96.53	91.53
	II	93.56	96.40	96.64	91.44
	III	96.72	98.80	99.32	94.53

The classification results with different algorithm are shown in TABLE II. our algorithm I combines all SVM, II combines 50% SVMs of all with higher accuracy, III removes 30% SVMs with lower kappa value and 20% SVMs with lower accuracy. In Bagging, iteration times is 50. In random subspace and our algorithms, the number of SVM is 50, and feature size is: 24 on Sonar, 14 on Ionosphere, 11 on Wdbc, 19 on SPECTF, respectively. TABLE II demonstrate that on Ionospher and Wdbc datasets that have high precision using single SVM and other ensemble algorithm, the accuracy is improved to a certain extent using our algorithms. Furthermore, on Sonar and SPECTF datasets which have not high accuracy using other algorithms, the classification accuracy is a substantial increase.

One furtherway to gain insight into the behavior of these ensemble methods is to construct *Kappa-error*

diagrams. For each pair of SVM, we measure their diversity by *Kappa* value, and measure accuracy as the average of prediction error. These diagrams help visualize the accuracy and diversity of the individual classifiers constructed by the ensemble method. Figure 8 shows the *kappa-error* diagrams of Bagging, Subspace and our algorithms on SPECTF dataset, and on other datasets the *kappa-error* diagrams are similar.



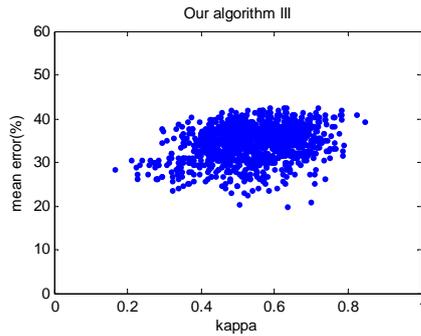


Figure 8. *kappa-error* diagrams using different algorithm on SPECTF dataset

The *kappa-error* diagrams demonstrate that *Bagging* method has highest accuracy but lowest diversity, *Random subspace* method has highest diversity but lowest accuracy, *Our algorithm I* is higher accuracy but lower diversity than *Random subspace method*, *Our algorithm II* improved accuracy, *Our algorithm III* improved both accuracy and relative diversity based on *algorithm I* and the accuracy is increase and relative diversity is acceptable, comparing to *Random subspace*.

IV CONCLUSIONS

A novel feature subspace ensemble learning algorithm based on feature clustering was proposed to improve ensemble classifier performance, allowing for high dimensional data sets. Features were clustered on weighted average linkage method and feature subsets consisting of the attributes extracted from each category at random served as the inputs of individual SVM to ensure that there were significant diversities among individual learning machine. The correspondence between the high ensemble accuracy and the least feature subspace was found that when let threshold $\lambda_k = 0.35 \sim 0.5$ and then feature subspace size was bigger than 1/3 of the total of features. GA-SVM was used to increase individual classification performances. Selective ensemble learning is adopted that both diversity and accuracy were considered using *Kappa* statistic to demonstrate the relative diversity of pair of individuals. The *kappa-error* diagrams demonstrate individual classifiers were both accurate and diverse, and the results showed the classification accuracy increase significantly on the datasets which had high or not high accuracy using other algorithms.

ACKNOWLEDGMENT

The authors are grateful to the National Natural Science Council of China and the Natural Council of Jiangsu Province for their financial supports (No. 60974126 and BK2009094).

REFERENCES

[1] L. Hansen and P. Salamon, "Neural Network Ensembles," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, pp. 993–1001, 1990.

[2] T.K. Ho, J.J. Hull, and S.N. Srihari, "Decision combination in Multiple Classifier Systems," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, pp. 66-75, Jan. 1994.

[3] H. Altınçay, "Decision trees using model ensemble-based nodes," *Pattern Recognition*, vol. 40, pp.3540-3551, 2007.

[4] L. Yu, S. Y. Wang, and K. K. Lai, "A novel nonlinear ensemble forecasting model incorporating GLAR and ANN for foreign exchange rates," *Computers and Operations Research*, vol.32, pp.2523–2541, 2005.

[5] D. A. Stone, M. R. Allen, F. Selden, "The Detection and Attribution of Climate Change Using an Ensemble of Opportunity," *American Meteorological Society*, vol.1, pp.504-516, 2007.

[6] L. Yu, S. Y. Wang, K. K. Lai, "Credit risk assessment with a multistage neural network ensemble learning approach," *Expert Systems with Applications*, vol.34, pp. 1434–1444, 2008.

[7] T.G. Dietterich, "An experimental comparison of three methods for constructing ensembles of decision trees: bagging, boosting, and randomization," *Machine Learning*, vol. 40, pp. 139-157, 2000.

[8] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, pp. 123–140, 1996.

[9] S. Gunter, H. Bunke, "Feature selection algorithms for the generation of multiple classifier systems and their application to handwritten word recognition," *Pattern Recognition Letter*, vol.25, pp.1323-1336, 2004.

[10] T.K. Ho, "The random subspace method for constructing decision forests," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, pp. 832-844, Aug.1998.

[11] R. Bryllia, O. R. Gutierrez, F. Queka, "Attribute bagging: improving accuracy of classier ensembles by using random feature subsets," *Pattern Recognition*, vol. 6, pp.1291-1302, 2003.

[12] M. E. Akay, "Support vector machines combined with feature selection for breast cancer diagnosis," *Expert Systems with Applications*, vol. 36, pp.3240-3247, 2009.

[13] Q. Hu, H. D. Yu, Z. X. Xie, et al, "EROS-ensemble rough subspaces," *Pattern Recognition*, vol. 12, pp. 3728-3739, 2007.

[14] L. Shi, X. M. Ma, L. Xi, et al, "Rough set and ensemble learning based semi-supervised algorithm for text classification," *Expert Systems with Applications*, vol. 38, pp. 6300–6306, 2011.

[15] T. P. Hong, P. C. Wang, C. Po, et at, "An effective attribute clustering approach for feature selection and replacement," *Cybernetics and Systems*, vol. 40, pp. 657-669, Nov. 2009.

[16] Z. H. Zhou, J. X. Wu, W. Tang, "Ensembling neural networks: many could be better than all," *Artificial Intelligence*, vol.137, pp.239-263, 2002.

[17] M. Sileshi, B. Gamback, "Evaluating clustering algorithms: cluster quality and feature selection in content-based image clustering," *2009 World Congress on Computer Science and Information Engineering, CSIE*, pp. 435-41, 2009.

[18] W. X. Zha, M. Li, Y. P. Fu, "Empirical analysis of highway traffic accident in average linkage clustering method," *Journal of East China Jiaotong University*, vol. 6, pp.113-118, 2010.

[19] H. Tao, X. P. Ma, and M.Y. Qiao, "Support vector machine selective ensemble learning on feature clustering and genetic algorithm," unpublished.

[20] V. N. Vapnik, *The nature of Statistical Learning Theory*, X. G. Zhang, trans, Beijing: Tsinghua University Press, 2000.

[21] O. Chapelle, V. N. Vapnik, O. Bousquet, et al, "Choosing kernel parameters for support vector machines," *Machine Learning*, vol. 46, pp.131–160, 2001.

- [22] H. C. Kim, S. Pang, et al, "Constructing Support vector machine ensemble," *Pattern Recognition*, vol.12, pp. 2757-2767, 2003.
- [23] K. Q. Wang, S. C. Yang, T. H. Dai, X.B. Bai, "Method of optimizing parameter of least squares support verctor machine by genetic algorithm," *Computer Application and Software*, vol.7, pp. 109-111, 2009.
- [24] Faruto, Y. Li, "LIBSVM – farutoUltimate Version - a toolbox with implements for support vector machines based on libsvm," *Software available at <http://www.ilovematlab.cn>*, 2009.
- [25] G. Brow, J. Wyatt, R. Harris, et al, "Diversity creation methods: A survey and categorization," *Journal of Information Fusion*, vol.1, pp. 1-28, 2005.
- [26] M. Aksela, J. Laaksonen, "Using diversity of errors for selecting members of a committee classifier," *Pattern Recognition*, vol. 39, pp. 608-623, 2006.
- [27] Y. Kim, "Toward a successful CRM: variable selection, sampling, and ensemble," *Decision Support Systems*, vol.41, pp. 542-553, 2006.
- [28] A. Frank, A. Asuncion, UCI Machine Learning Repository. Irvine, CA: *University of California, School of Information and Computer Science*, 2010.



Hui Tao was born in Xinyang, China in 1979. She received her B.S. of Automation in 2001 and M.S. degrees of control theory and engineering in 2007 all from School of Electrical Engineering and Automation, Henan Polytechnic University. She is currently a Ph.D. candidate of control theory and engineering in School of Information

and Electrical Engineering at China University of Mining and Technology in Xuzhou, China.

She was working as a lecturer in School of Electrical Engineering and Automation, Henan Polytechnic University between July 2001 and July 2010. She is currently working as a researcher in the Automation Laboratory at China University of Mining and Technology and a lecturer at Henan Polytechnic University. The paper is extended of the thesis "Support Vector Machine Selective Ensemble Learning on Feature Clustering and Genetic Algorithm" in 2011 International Conference on Electrical Information and Mechatronics (ICEIM 2011). Her main areas of interest include SVM classification, classifier ensembles, pattern recognition and coal mine safety.

Xiao-ping Ma received his Ph.D. of control theory and engineering in 2001, He is currently working as a professor and Ph. D. supervisor in control theory and engineering. His main areas of interest include control theory and applications, network control, and computer technology and application.

Mei-ying Qiao is a Ph.D. candidate of control theory and engineering in School of Information and Electrical Engineering at China University of Mining and Technology in Xuzhou, China. Her main areas of interest include SVM classification and regression, pattern recognition, chaos time series prediction and coal mine safety.