

Selected Features for Classifying Environmental Audio Data with Random Forest

Yan Zhang* and Dan-jv LV

School of Computer and Information Southwest Forestry University Kunming, China

Abstract: Environmental audio classification has been the focus of research in the area of speech recognition. It is difficult to find an optimal classifier and select the optimal features from various features extracted from environmental audio data. Random forest is a powerful machine learning classifier compared to other conventional pattern recognition techniques. In this paper, the performance of the Random Forest classifier and the selected features model for environmental audio classification is explored. The comparison and analysis of classification results, obtained by employing the Bagging, AdaBoost, and Random Forest for environmental audio data, are given. The experiments carry out the assessment and selection of importance of variables by means of GINI Index. The results show that the Random Forest method can effectively improve the performance of environmental audio data classification even under the fewer training examples. The classification model, built from the selected features, obtains better performance in both accuracy and efficiency than that of all features for environmental audio data.

Keywords: Ensemble learning, Environmental audio, Random forests, Variable importance.

1. INTRODUCTION

Audio classification is a basis for further audio retrieval and analysis, as well as an access to extract audio structure and content. The environmental audio classification is increasingly becoming the hot spot for researchers [1]. The selection of classification model has been the focus in the speech recognition and classification fields. The minimum distance classifier, neural network, support vector machines, decision tree, and hidden Markov Model are traditional techniques for audio classification [2]. It is difficult to find the optimal classifier with good generalization and to improve the performance of single classifier.

In the actual application, the large training examples of environmental audio data are too expensive or tedious to acquire. When the number of the labeled examples decreases in the supervised classification, the performance will get worse. How to use a small labeled data to improve the learning performance becomes the key problem, which researchers in the pattern recognition and machine learning are focusing on.

So far, research on the theory and algorithms of ensemble learning has always been a hot spot in the field of machine learning, which combines various algorithms and exploit complementary with different classifiers to boost the classification precision. It was listed as four major research directions in the field of machine learning by Dietterich, an authoritative scholar in the international machine learning field.

Ensemble learning technologies are dedicated to using a combination of different basic classifiers to generate multiple classifications of the same data or to the target different subsets of the data. These algorithms such as random forest, bagging and boosting have received increasing interest because they are more accurate and robust to noise than single classifier [3]. Based on the basic premise, a set of classifiers do perform better classification than an individual classifier. Random forest, one of ensemble learning techniques, is increasingly being applied in some classification issues [4]. The paper mainly focuses on the application of random forest into the environmental audio classification, and the behavior and performance of random forest are explored.

In addition, there are many features involved in the environmental audio data classification. It is an issue to solve that how to select effective features to improve the efficiency in classification without decreasing the performance.

2. STUDY DATA AND FEATURE EXTRACTION

The study data are acquired from network and field recording, with 8k sampling rate, 16 bits and mono-track. The environmental audio data consists of five classes of sound, namely, sound of bird, frog, wind, rain and thunder. The sound length amounts to almost 10 minutes. The silence and noise are removed in the pre-processing.

The feature extraction is executed based on the bit-stream through the G.723.1 data encoding on the Matlab 7.1 platform including two kinds of features. One is the CELP features and the other is MFCC features. CELP is characterized by coefficient of short tube cascade channel model, while MFCC is based on human auditory representation speech signal characteristics by means of frequency conversion.

*Address correspondence to this author at the School of Computer and Information Southwest Forestry University Kunming, 650224, China; Tel: +86 871 63863018; Fax: +86 871 63863018; E-mail: zydyr@163.com

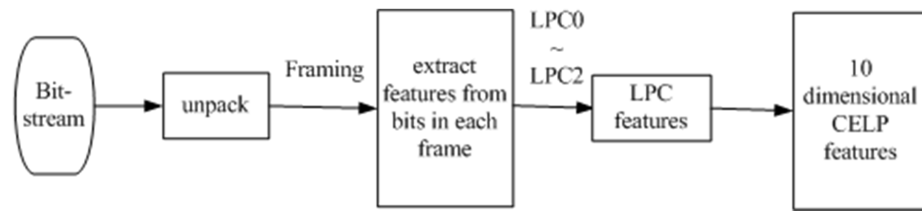


Fig. (1). Extraction of CELP environmental audio features.

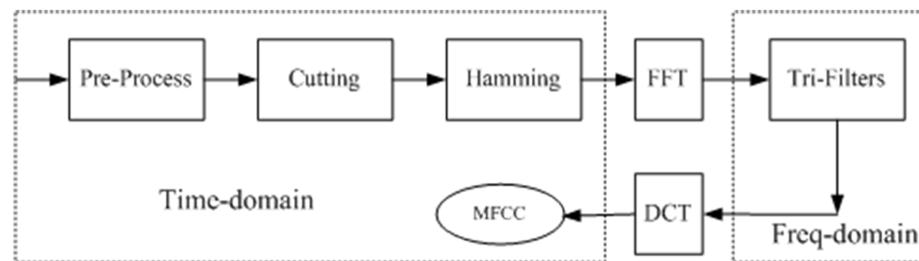


Fig. (2). Process of extracting MFCC features.

CELP features are mainly from LPC, Linear Prediction Coefficient, analyzing the sound mechanism from the original source. Through the short tube of channel cascade model research, the system transfer function is in line with the pile in the form of digital filter, so the signal of t time can be used several times before the combination of the signal to estimate. By estimating the actual sound samples values and linear prediction to achieve minimum mean square error between the sampling value LMS (Least Mean Square), the linear prediction coefficients can be obtained, that is LPC. LPC features are extracted at each bit-frame after the unpacked bit-stream. 10 order coefficients of LPC are obtained at each bit-frame, from 0 ~ 23bit (LPC0 ~ LPC2), which consists of 10 dimensions of LPC features. The process of extracting CELP features is shown in Fig. (1).

The other kind of feature is named MFCC (Mel-scale Frequency Cepstral Coefficients). The human ear is sensitive to the frequency of different levels. That is, it has the strong recognition ability for the low frequency of the voice signal while weak for the high frequency signal. The bit-streams of the data are transferred into the wave format. With the compression features of MFCC, the features extracting process is as follows (See Fig. (2)).

(1) To transfer the signals of environmental audio data from time domain to frequency domain with FFT (Fast Fourier Transformation),

(2) Convolution of the logarithmic energy spectrum of signals in accordance with Mel scale distribution in triangular filter,

(3) The vector composed of each filter output is transferred with DCT (Discrete Cosine Transform), taking the top 13 coefficients.

3. RANDOM FOREST

The random forest classifier consists of a combination of tree classifiers where each classifier is generated using a

random vector sampled independently from the training set of input vectors, and each tree casts a unit for the most popular class in which to place a given input vector [5]. The technique for generating a random forest is generally a combination of the bagging and random subspace methods. Bagging is a technique to improve classification accuracy and avoid over-fitting. Given a training set of size N , bagging method generates a number of new training sets, each of size n (where $n < N$) by randomly drawing samples with replacement from the original training set. Also, suppose the data contains M attributes. For each node of the tree, m ($m < M$) attributes are randomly chosen to provide the basis for calculating the best split at that node. Once the random forest is formed, each sample is classified into the class taking the most popular votes from all the tree predictors in the forest.

3.1. Construction of Random Forest

Random Forest (RF) [5] is an extension of Bagging, where the major difference with Bagging is the incorporation of randomized feature selection. During the construction of a component decision tree, at each step of split selection, RF first randomly selects a subset of features, and then carries out the conventional split selection procedure within the selected feature subset. Fig. (3) illustrates the construction of RF.

3.2. Features of Random Forest

The random forest technique holds several advantages as follows [5].

(1) Its accuracy is as good as AdaBoost and sometimes better,

(2) It is relatively robust to outliers and noise,

(3) It is faster than bagging or boosting, and it shows useful internal estimates of error and the importance of the variable.

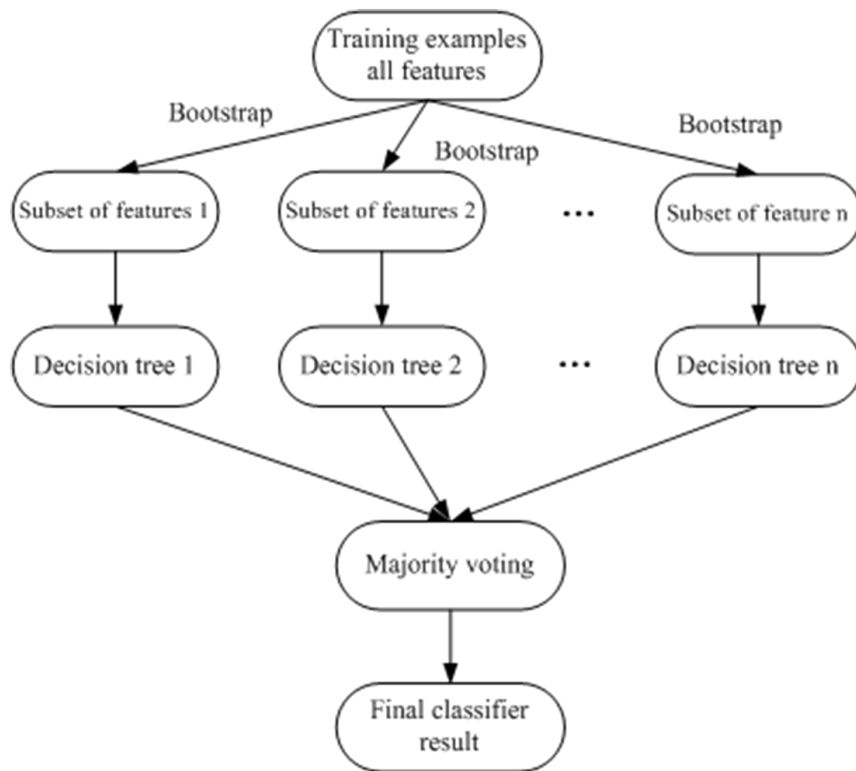


Fig.(3). Construction of Random Forest.

One of the remarkable benefits of random forest is to calculate useful information about errors, variable importance, and data outliers. That information can be used to evaluate the performance of the model and make changes to the training data if necessary.

In the forest building process, by sampling with replacement for each tree, random forest only uses roughly two-thirds of the training data to build the random forest model, and the remaining one-third of the training data (the out-of-bag samples) can be used to estimate the error of the predictions. The OOB error greatly simplifies the accuracy assessment portion of each analysis. The OOB error of Random Forest is given in [6, 7].

Decision tree building requires choosing a suitable attribute selection measure to maximize dissimilarity between classes. There are many approximations for selecting attributes which can be used for induction in decision trees. Some most frequent ones are gain-ratio [8] and GINI index [9]. The RF usually adopts the GINI index as a measure for the best split selection, which measures the impurity of a given element with respect to the rest of the classes [10]. The OOB sample is used to estimate the prediction error and then to evaluate variable importance.

An assessment of the relative importance of the different features or variables during the classification process is also provided by the random forest algorithm. By identifying variables that contribute information to the analysis, it is important to know how each predictive variable influences the classification model to be able to select the best variables, where data dimensionality is very high. To assess the importance of each feature, the random forest switches one

of the input random variables while keeping the rest constant, and it measures the decrease in accuracy which has taken place by means of the OOB error estimation and decrease of GINI Index.

4. ANALYSIS OF THE EXPERIMENTAL RESULTS

4.1. Data of Experiments

The frame of the environmental audio data is the basic unit for data statistics and classification in the experiments. In order to avoid large amount of data overflowing in training, the experimental data is randomly selected from the total number of frames in the sampling according to one-third of the total data in each category. The whole features of the dataset are including CELP and MFCC, summed up 23 attributes. 10 dimensions of features are involved in CELP, while other 13 variables are composed of MFCC.

It is 10 times sampling with 75% as the training sample, and 25% as the testing sample. Five classes of audio signal frame situation and classification of samples are shown in Table 1.

4.2. Performance of the Random Forest Classifier

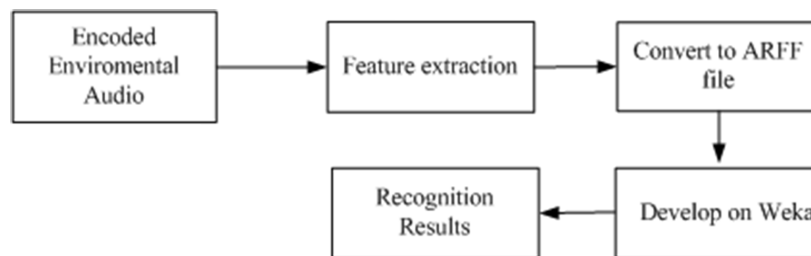
The experiment carries out based on the platform of development Weka [11]. At first, the environmental audio feature data is converted into an ARFF format file through code of Matlab. Then the ARFF format file can be obtained and classified with the module in development Weka. Finally, the results are saved in type of ARFF file. The flow chart of environmental audio classification is illustrated in Fig. (4).

Table 1. Information of dataset in the experiments.

Class	Total Frames	75%--Training	25%--Test
Bird	751	563	188
Wind	3102	2326	776
Rain	3000	2250	750
Frog	1472	1104	368
Thunder	808	606	202
Total	9133	6849	2284

Table 2. Classification performance of ensemble classifier under different percent training sample.

Training Samples	Bagging		AdaBoost		Random Forest	
	Kapp	err	Kapp	err	Kapp	err
10%	0.88	8.77%	0.8912	8.36%	0.9100	6.59%
20%	0.9121	6.45%	0.9284	5.26%	0.9424	4.22%
40%	0.9409	4.34%	0.9475	3.86%	0.9574	3.13%
60%	0.9488	3.78%	0.956	3.23%	0.9656	2.54%
80%	0.9565	3.33%	0.9619	2.80%	0.9688	2.29%

**Fig. (4).** Flow chart of environmental audio classification.

Bagging, AdaBoost and Random forest are involved in the ensemble strategies. In the experiment, the data are selected at one-third rate of each class from original samples.

The Bagging and Boosting algorithms are popular ensemble strategies in the application. Bagging uses bootstrap sampling to generate different data samples, while all the data samples have large overlap. So it should be used with unstable learners such as neural network and decision trees. That is, the more unstable the learner, the larger the performance improvement. If the base classifiers are stable, the improved performance of the Bagging is not obvious. The Bagging [12] is not sensitive to noise data. AdaBoost [13] is a popular multiple classifiers algorithm that improves the simple boosting algorithm via an iterative process.

In the ensemble classifier, the base classifier algorithm uses the decision tree J48. The rates of the training examples are 10%,20%,40%, 60%, and 80%, respectively. The results

of Bagging, AdaBoost and random forest (combined with 100 decision trees) are illustrated in Table 2. *Kapp* stands for the kappa statistic. With the increase of training examples rate, the kappa statistic is on rise while the error rate is on the decline in three classification methods. The random forest is stable and obtains better results than others.

In addition, it shows that random forest can achieve super classification performance with small training examples in most cases. For example, the error rate of random forest with 20% training samples is commensurate to the Bagging and AdaBoost methods with 40%. The result of AdaBoost is superior to Bagging. Fig. (5) shows the performance comparison of three ensemble strategies.

4.3. Selection of Features with Random Forests

Many variable selection (or feature selection) procedures are based on model estimation to evaluate and compare a family of models with ranked variable importance.

Table 3. Classification performance of RF classifier with different kind features.

Training Samples	MFCC		CELP		MFCC+CELP	
	Kapp	err	Kapp	err	Kapp	err
10%	0.9098	6.58%	0.8551	10.52%	0.9100	6.59%
20%	0.9334	4.88%	0.8643	9.89%	0.9424	4.22%
40%	0.9499	3.68%	0.8851	8.39%	0.9574	3.13%
60%	0.9576	3.12%	0.8914	7.92%	0.9656	2.54%
80%	0.9603	2.92%	0.8935	7.77%	0.9688	2.29%

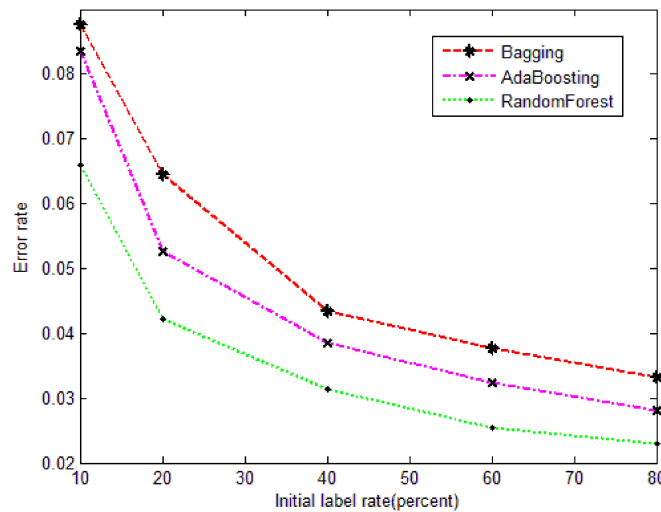


Fig. (5). Performance comparison of three ensemble strategies.

The layout of the split variables in a decision tree provides information about the importance of the features in the general classification model and in the classification of each category. However, it is almost impossible to carry out the interpretation with classifier ensembles based on multiple decision trees. The random forest assesses the importance of the variables by means of the GINI Index and the OOB subset. In the experiments, three kinds of features are adopted. The first is the MFCC features labeled as x_1, x_2, \dots, x_{13} in dataset. The second is the CELP features labeled as $x_{14}, x_{15}, \dots, x_{23}$. The third is the combination of the MFCC and CELP, ranged from x_1 to x_{23} . With random forest, three classification models are built based on different kind of features. Their performances is shown in Table 3. The model with MFCC+CELP is the best among them due to the 23 dimensional variables. From the results, it is found that the MFCC features play an importance role in the environmental audio classification, so they obtain better generalization than that of CELP features.

In each classification model, the variable importance could be obtained.

Fig. (6) shows the contribution of each variable to the classification model generated by considering MFCC, CELP and MFCC+CELP features, respectively. The variable im-

portance is ranked by the Mean Decrease Accuracy and Mean Decrease GINI. In the experiment, the GINI Index is considered only.

According to the GINI Index, the features of MFCC and CELP with higher contribution to the RF model are selected to form the new variable sets, which include $x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_9, x_{10}, x_{13}, x_{14}, x_{15}, x_{16}, x_{17}$ and x_{18} variable. The classification model is also built from the selected features. The comparison of two models classification performance and the time of building the model is illustrated in Table 4.

Seen from the Table 4, the accuracy of classification model with Selected Features is slightly superior to the whole variables (MFCC+CELP) except for the case of 60% training sample. The number of features involved in classification decreases, but error rate of classification dose not rise. And the time for building the classification model is shown in the Table 4. Due to reducing the number of variables, the time in selected features is lower than that of the MFCC+CELP. So the classification model built from the selected features obtained better performance in both accuracy and efficiency than that of all other features. The performance and efficiency of different features for RF classification are illustrated in Fig. (7) and Fig. (8), respectively.

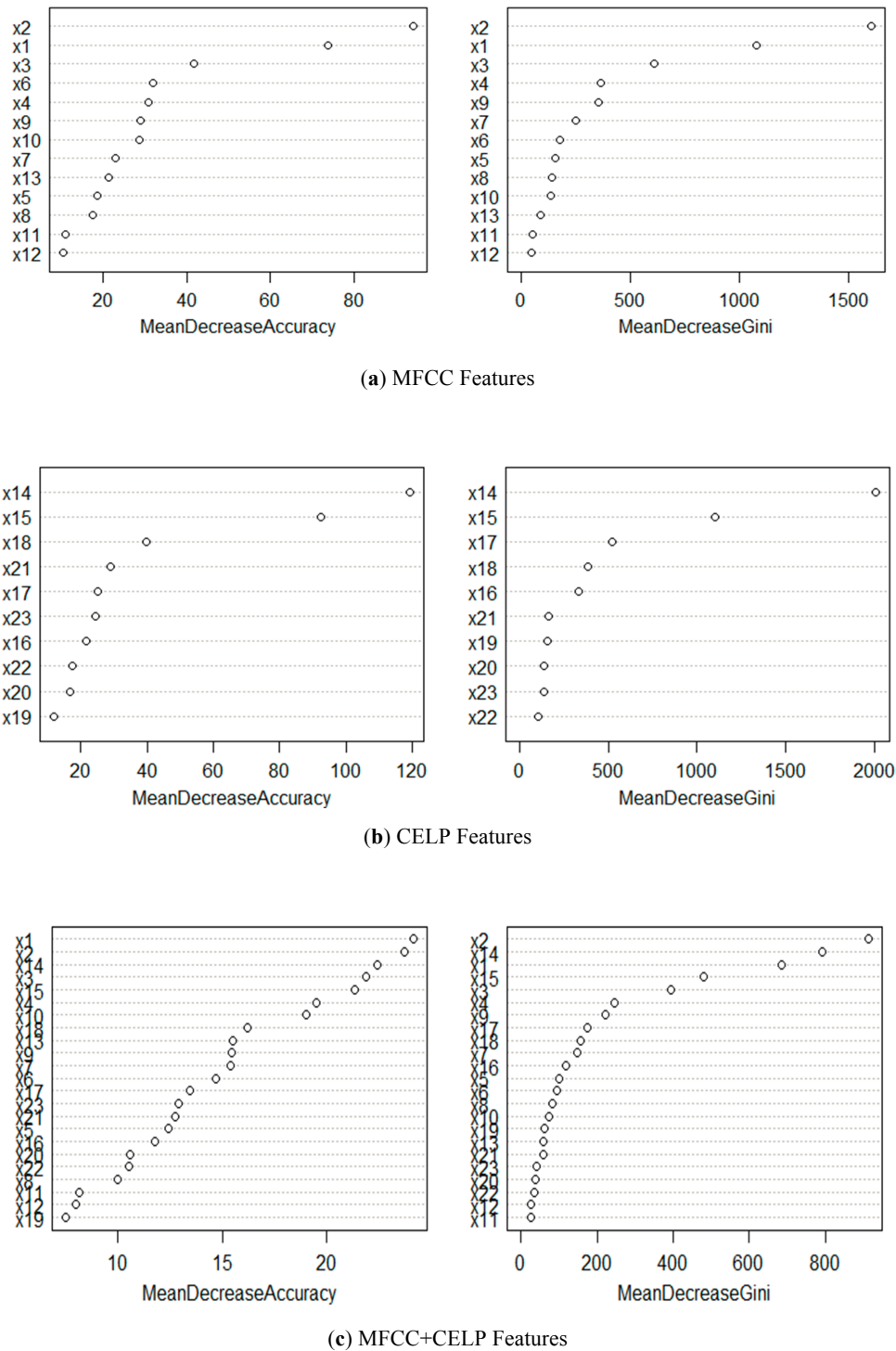


Fig. (6). Variable importance contribution in terms of OOB mean and GINI Index mean decrease in accuracy.

CONCLUSION

Random forest is a robust algorithm that can be used for classification application. According to the experimental results from the random forest and other ensemble methods such as Bagging, AdaBoost on environmental audio data, random forest outperforms others in most cases. The two kinds of features, MFCC and CELP, are involved in the ex-

periments. The results also show that the MFCC feature can obtain better prediction performance than that of CELP. In addition, random forest calculates useful information about errors and variable importance, which can be used to evaluate the performance of the model. Although the selected features by the GINI index (variable importance) make changes to the training data, reduce the dimensions of feature, yet

Table 4. Classification performance of RF classifier between two models.

Training Samples	Selected Features			MFCC+CELP		
	Kapp	err	time(s)	Kapp	err	Time(s)
10%	0.9107	6.51%	0.36	0.9100	6.59%	0.61
20%	0.9442	4.09%	0.83	0.9424	4.22%	1.37
40%	0.9595	2.98%	1.97	0.9574	3.13%	3.34
60%	0.9647	2.59%	3.22	0.9656	2.54%	5.74
80%	0.9695	2.24%	4.47	0.9688	2.29%	7.38

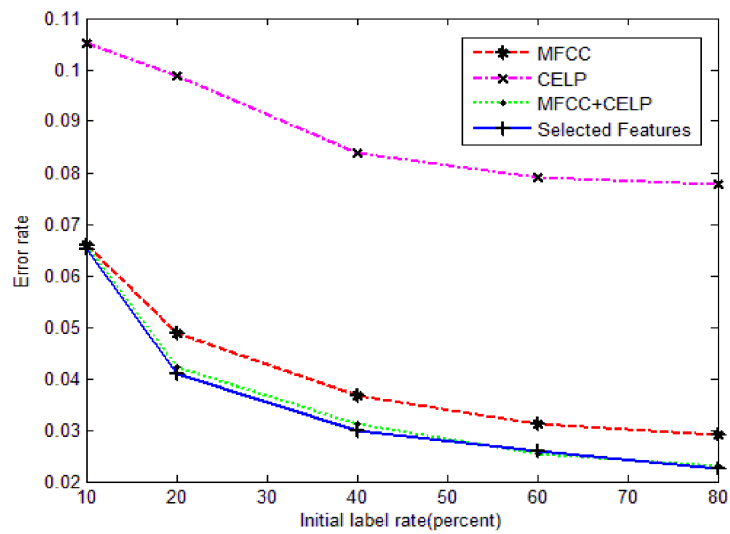


Fig. (7). Performance comparison of different features of RF classification model.

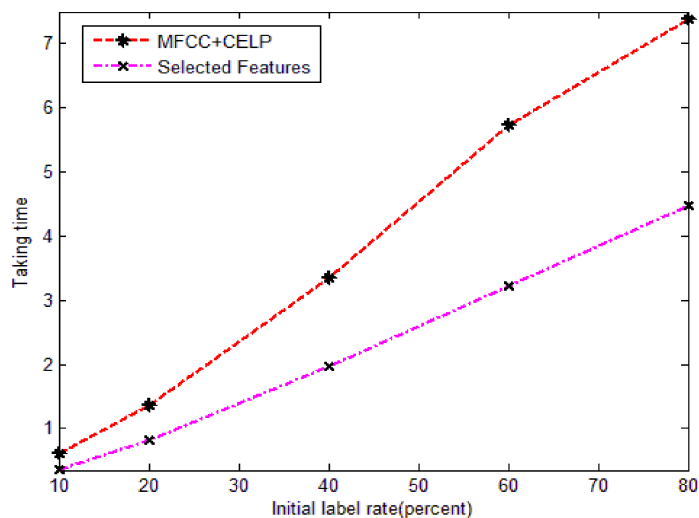


Fig. (8). Efficiency of selected features RF classification model.

they ensure the accuracy, and improve the efficiency in classifying environmental audio data.

For environmental audio data, various features can be extracted. How to select the optimal ones for classification is

an issue. Different kind of features form multiple views. Further research work about utilizing the multiple views in environmental audio data to build the better classification model is underway.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

This work was supported by the Research Fund Projects of Yunnan Provincial Department of Education under the Grant No. 2012C098.

REFERENCES

- [1] X.D. Yang. "Environmental audio classification based on support vector machine", *Electron. Measur. Technol.*, vol. 31, no. 9, pp. 121-123, 2008.
- [2] Y. Zheng, "Typical methods for audio classification", *JiSuanJi Yu Xian Dai Hua*, no. 8, pp. 59-63, 2007.
- [3] T.G. Dietterich, "Machine learning research: four current directions", *AI Magazine*, vol. 18, no. 4, pp. 97-136, 1997.
- [4] Ghimire, J. Roan, and J. Miller, "Contextual land-cover classification: incorporating spatial dependence in land-cover classification models using random forests and the Getis statistic", *Remote Sens. Lett.*, vol. 1, pp. 45-54, 2010.
- [5] L. Breiman, "Random forests", *Mach. Learn.*, vol. 45, no. 1, pp. 5-32, 2001.
- [6] V.Y. Kulkarni, and P.K. Sinha, "Random forest classifiers: a survey and future research directions", *Int. J. Adv. Comput.*, vol. 36, pp. 1144-115
- [7] G. Biau, "Analysis of a random forests model", *J. Mach. Learn Res.*, no. 13, pp. 1063-1095, 2012.
- [8] J.R. Quinlan, *C4.5 Programs for Machine Learning*. 1st ed. M. Kauffman, San Mateo, CA, 1993.
- [9] L. Breiman, *Classification and Regression Trees*, Chapman & Hall/CRC, 1984.
- [10] V.F. Rodriguez, B. Ghimire, and J. Rogan, "An assessment of the effectiveness of a random forest classifier for land-cover classification", *ISPRS J Photogramm.*, no. 67, pp. 93-104, 2012.
- [11] I.H. Witten, E. Frank, A. and M.A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*, 3rd ed., Morgan Kaufmann, 2011
- [12] L. Breiman, "Bagging predictors", *Mach. Learning*, vol. 24, no. 2, pp. 123-140, 1996.
- [13] R.O. Duda, P.E. Hart, and D.G. Stork, *Pattern Classification*, 2nd ed, China Machine Press, Beijing pp. 383-387, 2003.

Received: December 15, 2014

Revised: January 04, 2015

Accepted: February 25, 2015

© Zhang and Lv; Licensee Bentham Open.

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.