The Classification of Environmental Audio with Ensemble Learning *

Yan Zhang [1], Danjv Lv [1], Ying Lin[2]

[1]School of Computer and Information Southwest Forestry University, Yunnan Province, China
[2]School of Software Yunnan University, Yunnan Province, China
zydyr@163.com, lvdanjv@hotmail.com, linying@ynu.edu.cn

Abstract - Environmental audio classification has been the focus in the field of speech recognition. The paper proposes using ensemble learning to perform environmental audio classification. Ensemble learning is to train multiple component learners and combine their predictions to enhance the accuracy of classification. The paper explains the ensemble algorithms and strategies, and the comparison and analysis of classification results are given which obtain by employing the single classification algorithm, Bagging, Boosting, Random Forest and MCS ensemble classifiers for the given environmental audio data. The experimental results show that the ensemble technology can effectively improve the performance of environmental audio data classification. Even under the fewer number of the training examples, it provides a kind of effective approach to guarantee the performance and generalization of classification**.**

Index Terms - Ensemble learning, Environmental Audio, Bagging, Boosting.

I. Introduction

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

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

It is a good way to combine various algorithms and exploit complementary between different classifiers to boost the classification precision of environmental audio. The research on the theory and algorithm of ensemble learning has always been a hot spot in the field of machine learning. It was listed as four major research directions in the field of machine learning by Dietterich who is an authoritative scholar in the international machine learning field [3]. Ensemble learning technology can solve the issues above well when it is applied to speech recognition. The paper mainly focuses on the application of ensemble learning into the environmental audio classification.

II. Ensemble learning

Ensemble methods [4] are designed to train multiple learners to refine the classifier. In contrast to ordinary learning approaches which try to construct on one kind of learner from training data, ensemble methods try to construct a set of learners and combine them. Ensemble learning is also called committee-based learning, or learning multiple classifier systems.

Three strategies need to be chosen for building an effective ensemble system, which are referred to these as the three pillars of ensemble systems [5]: (1) data sampling or selection; (2) training member classifiers; and (3) combining classifiers. The key issue of ensemble learning is how to design the base classifiers with stronger generalization and diversity. The accuracy and diversity among base classifiers are crucial to ensemble performance. Therefore, it is desired that the individual learners should be accurate and diverse. The success of ensemble learning lies in achieving a good trade-off between the individual performance and diversity. Though diversity is the key factor, there is no well-accepted formal definition of diversity. How to measuring diversity in classifier ensembles and diversity for building the ensemble are presented in detail in the literature [10].

Currently, there are two ways to implement the ensemble system. One is the combined classification method based on the different training sample using the same learning algorithm to produce homogenous base learners. The other uses multiple learning algorithms on same training examples to produce heterogeneous learners, so the results are obtained from the decision fusion of component learners. The Bagging and Boosting algorithms are popular ensemble strategies in the application.

A. Bagging Algorithm

Bagging algorithm[6] is proposed by Breiman, the steps are as follows[7]:

Step 1 (Bootstrap) The data subset with n' size ($n'<n$) for

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

training the base learners is randomly obtained from the original data set containing *n* training examples;

Step2 by applying the step 1 *T* times, *T* samples of *n'* training examples are obtained, and they are independent each other;

Step3 from each sample a base learner can be trained by applying the base learning algorithm;

Step4 the final result is obtained by adopting the most popular strategies such as voting for aggregating the outputs of the base learners.

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 larger the performance improvement. If the base learners are stable, the improved performance of the bagging is not obvious. The bagging is not sensitive to noise data.

B. Boosting Algorithm

Boosting is a general method for improving the performance of a weak learner. The method works by iteratively invoking a weak learner, on training data that is taken from various distributions.

In order to make weak learners be boosted to strong ones, the Boosting algorithm is emerging. Boosting algorithm works by training a set of learners sequentially and combining them for prediction, where the later learners focus more on the mistakes of the earlier learners.

AdaBoost (Adaptive Boosting) is a popular ensemble algorithm that improves the simple boosting algorithm via an iterative process. The amount of focus is quantified by a weight that is assigned to every pattern in the training set. Initially, the same weight is assigned to all the patterns. In each iteration the weights of all misclassified instances are increased while the weights of correctly classified instances are decreased. As a consequence, the weak learner is forced to focus on the difficult instances of the training set by performing additional iterations and creating more classifiers. Furthermore, a weight is assigned to every individual classifier. This weight measures the overall accuracy of the classifier and is a function of the total weight of the correctly classified patterns. Thus, higher weights are given to more accurate classifiers. These weights are used for the classification of new patterns [7].

The iterative procedure provides a series of classifiers that complement one another. The AdaBoost algorithm is shown in Fig.1.

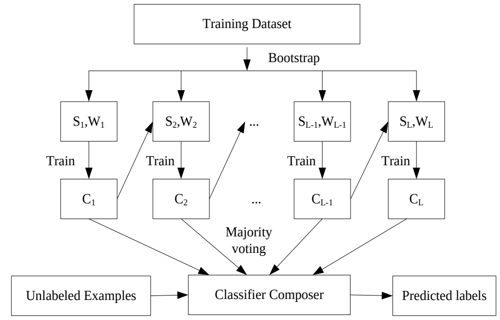


Fig.1 Flow chart of AdaBoost algorithm

C. Decision fusion based on different type base classifiers

Different type classifiers are trained by the same training data respectively, and combined with some strategies. The final classifier classifies new data and gives their by predictions. The multiple classifiers system is built by three classifiers with three learning algorithms and the model is shown as Fig. 2.

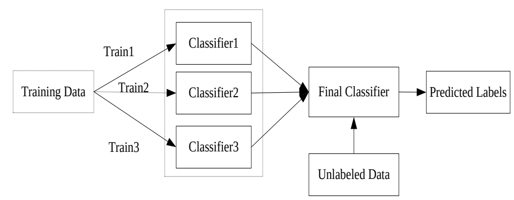


Fig. 2 The model of multiple classifiers system

D. Random Forest

Random Forest (RF) [9] 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 is illustrated the construction of RF.

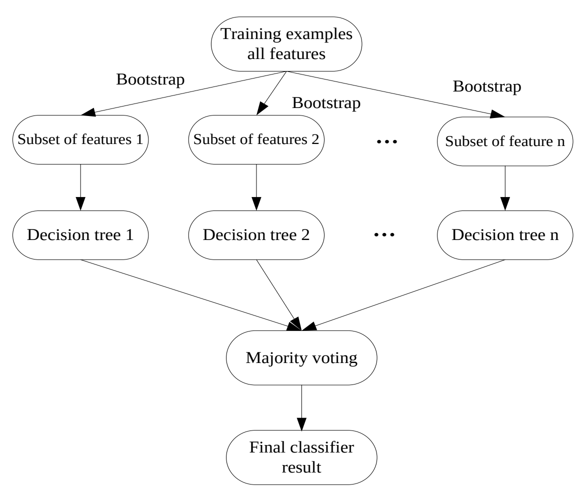


Fig.3 Construction of Random Forest

III. Experimental data and method

The experimental data are acquired from network and field recording, with 8k sampling rate, 16 bits and mono-track. The environmental audio data includes five classes, such as the sound of different kinds of birds, frogs, wind, rain and thunder. The speech sound length amounts to almost 10 minutes. The pre-processing includes silence and noise remove.

A. the feature extraction of Environmental audio

The feature extraction is executed based on the bit-stream through the G.723.1 data encoding on the Matlab 7.1 platform. LPC and pitch features are extracted at each bit-frame after the unpacked bit-stream. 10 order coefficients of LPC is obtained at each bit-frame, from 0 ~ 23bit (LPC0 ~ LPC2), which consists of the 10 dimensions of LPC features. 1 dimension pitch is extracted from the 24th to 38th bit at each bit-frame. Finally, the 11 dimension features of CELP audio are composed.

B. The method of experiment

The experiment carried out on the platform of development Weka[8]. At first, the environmental audio feature data is converted to an ARFF format file through program of Matlab. Then the ARFF format file can be obtained and classified with the module in development Weka. Finally, the results are obtained in ARFF file. The experiments are involved the single classifier and the ensemble classifiers.

The single classifier includes the algorithms such as decision Tree J48, Naïve Bayes, Radial basis function and neural network, while Bagging, Adaboost and Random forest are involved in the ensemble strategies. The flow char of environmental audio classification is shown in Fig. 4.

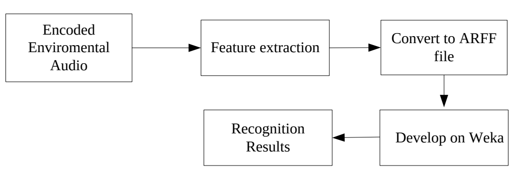


Fig. 4 Flow chart of environmental audio classification

IV . Analysis of the experimental result

In the experiment, the data are selected from the total according to a third of the total amount of data in each category. It is 10 times sampling randomly with 75% as the training sample, 25% as the testing sample. Five classes of audio signal frame situation and classification of samples as shown in TABLE I.

TABLE I The training and testing examples of CELP

class	Total frames	frames /time	75%-- training	25%-- test
bird	3900	3000	2250	750
wind	6434	2000	1500	500
rain	6200	2000	1500	500
frog	3334	1200	900	300
thunder	1634	1000	750	250
total			6900	2300

A. Single classifier experiment

The single classifier algorithm adopts the decision tree J48, Naïve Bayes (NB), Radial Basis Function (RBF) and neural network (BP). The rate of the training examples are 10%,20%,40%, 60%,80% and 100% respectively. The results of classification are shown in the TABLE II.

TABLE II Classification error rates of single classifier under different percent training samples

Training examples	Single-classifier			
	J48	NB	RBF	BP
10%	0.2159	0.2271	0.1823	0.1982
20%	0.1984	0.2216	0.1812	0.1875
40%	0.1810	0.2111	0.1866	0.1408
60%	0.1653	0.2062	0.1851	0.1550
80%	0.1659	0.2086	0.1875	0.1554
100%	0.1565	0.2100	0.1900	0.1579

According to the TABLE II, with the increasing of the training examples, the error rate of the classifiers decreased gradually. As the training examples are randomly selected from the labeled data, the distribution of training data is different from the labeled samples. Therefore, the error rate of the set of 80% training sample is higher than that of the set of 60%. In these methods, J48 and BP outperform the Naïve Bayes and RBF. BP is best while the performance of Naïve Bayes is lowest.

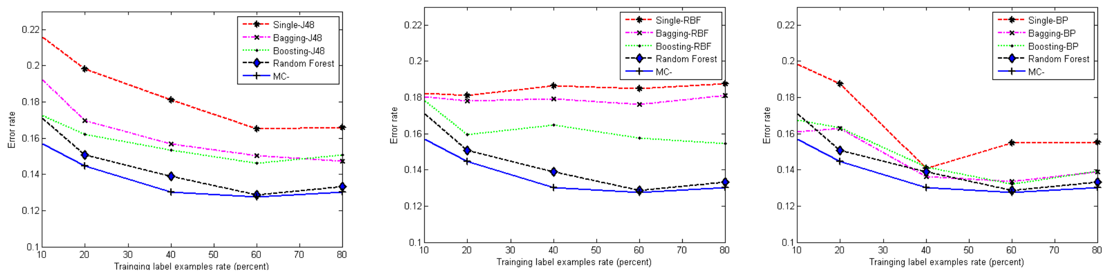
TABLE III Classification error rates of ensemble classifier under different percent training samples

Training samples	Bagging			Boosting			MCS	RF
	J48	RBF	BP	J48	RBF	BP	J48 RBF BP	
10%	0.1920	0.1803	0.1609	0.1725	0.1781	0.1675	0.1568	0.1709
20%	0.1696	0.1781	0.1629	0.1620	0.1595	0.1633	0.1447	0.1507
40%	0.1570	0.1793	0.1362	0.1535	0.1650	0.1417	0.1304	0.1389
60%	0.1505	0.1763	0.1337	0.1460	0.1577	0.1320	0.1277	0.1287
80%	0.1473	0.1813	0.1390	0.1508	0.1545	0.1392	0.1301	0.1331

B. Ensemble classification experiment

The outputs of multiple single classifiers trained respectively are fused by combination methods. This ensemble classification will compensate the shortcoming of single classifiers, and boost the classification accuracy. The results of Bagging, Boosting, three multiple classifiers (base classifiers are J48, RBF and BP, named as MCS) and Random forest (combined with 10 decision trees) are illustrated in TABLE III.

From the TABLE III, the performance of ensemble classification such as bagging and boosting are better than that of single classifier. Comparing with the figure of TABLE II, as the training samples reduce down to about 20%, the classification accuracy still can reach 100% training sample of single classifier. It shows that ensemble methods can achieve classification performance with small training examples (compared with the single supervised classification). Fig.5 shows the performance of four approaches of ensemble comparing with the single classifier. With the improvement of training examples rate, in most cases the error rate is on the decline. The MCS and RF are stable and make better results than others. Boosting is better than Bagging in the J48 and RBF learn algorithm.



(a) J48 (b) RBF (c) BP

Fig 5 Performance of different classifiers

V．Conclusion

Ensemble methods as the way to obtaining highly accurate classifiers by combining less accurate ones, are learning algorithms that construct a set of classifiers and then classify new data by taking a vote of their predictions. According to the experimental results from the single classifier and ensemble methods including Bagging, AdaBoost, MCS as well as Random forest on environmental audio, the ensemble methods are able to outperform any single classifier. Even with the small training examples, the performance of classification and generalization can be guaranteed. It is difficult to obtain the labeled data as training examples for environmental audio. Ensemble methods provide an effective way to perform the classification. Among the ensemble methods, the Random forest performs so well in environmental audio classification and obtains strong generalization.

Further research work about deriving effective diversity controls for ensemble learning, and combining ensemble learning with and semi-supervised learning to build the better learning model are underway.

Acknowledgments

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

References

[1] Xiao-mei Ding-cai Yang. Environmental audio classification based on support vector machine. Electronic Measurement Technology. 2008,31(9):121-123

[2] Yi-wen Zheng. Typical Methods for Audio Classification. JiSuanJi Yu Xian Dai Hua.2007,(8):59-63

[3] Dietterich T G. Machine learning research: Four current directions. AI Magazine, 1997,18(4):97-136

[4] Zhi-Hua Zhou. Ensemble Methods-Foundations and Algorithms. CRC Press, Taylor & Francis Group.2012

[5] Cha Zhang, Yunqian Ma. Ensemble Machine learning Methods and Application. Springer.2012

[6] Breiman L. Bagging predicators. Machine Learning, 1996, 24(2):123-140

[7] Duda,R.O. pattern classification(2nd ed.). Mechanical industry press. 2003

[8] Ian H. Witten, Eibe Frank, Mark A.Hall. Data Mining: Practical machine learning tools and techniques (3rd ed.). Morgan Kaufmann,2011

[9] Breiman L. Random forests. Machine Learning,45(1):5-32,2001.

[10] Kuncheva L.I. Combining Pattern Classifications: Methods and Algorithms. John Wiely & Sons, Hoboken, NJ,2004