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, 650224, China

Abstract: Environmental audio classification has been the focus in the field of speech recognition. For Environmental audio data, it is difficult to find an optimal classifier and select the optimal features from various features can be extracted. 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 are explored. The comparison and analysis of classification results are given which obtain by employing the Bagging, AdaBoost, and Random Forest for environmental audio data. The selection of importance variables in building the classification model and assessment variable importance are involved in the experiments. The experimental results show that the Random Forest method can effectively improve the performance of environmental audio data classification even under the fewer number of the training examples. According to the variable importance, the model built to improve on both the efficiency and the accuracy of classification based on the selected features in 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 becoming the hot spot for researchers increasingly [1]. Classification model selection has been the focus on the speech recognition and classification. 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 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.

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 between different classifiers to boost the classification precision. 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. Ensemble learning technologies use a combination of different base classifiers to generate multiple classifications of the same data or to the target different subsets of the data. Those 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 are 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 include five classes, such as the sound of different kinds of bird, frog, wind, rain and thunder. The speech 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, respectively. CELP is characterization by coefficient of short tube cascade channel model, while MFCC is based on human auditory representa-
tion speech signal characteristics by means of frequency conversion.

CELP features are mainly from LPC, Linear Prediction Coefficient analyses 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 combination of the signal to estimate. By making the actual speech 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 and pitch features are extracted at each bit-frame after the un-packed 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. One dimension pitch is extracted from the 24th to 38th bit at each bit-frame. Finally, the 10 dimension features of CELP audio are composed. The process of extracting CELP features is shown in Fig. (1).

The other kind 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 off 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 overfitting. 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 base for calculating the best split at that node. Once the random forest is formed, each sample is classified to the class taking the most popular votes from all the tree predictors in the forest.

3.1. Construction of Random Forests

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) is illustrated 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.

One of the remarkable benefits of random forests is to calculate useful information about errors, variable importance, and data outliers. This 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 forests only uses roughly two-thirds of the training data to build the random forests 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. This OOB error greatly simplifies the accuracy assessment portion of each analysis. The OOB error of Random Forest is given as [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 variable 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 of Gini Index decrease.

4. ANALYSIS OF THE EXPERIMENTAL RESULTS

4.1. Data of Experiments

The frame of the environmental audio data is the basis unit for data statistics and classification in the experiments. In order to avoid large amount of data overflowing in training, the experimental data is 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.
dimensionalities 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, 25% as the testing sample. Five classes of audio signal frame situation and classification of samples is shown in Table 1.

### 4.2. Performance of the Random Forests Classifier

The experiment carried out based on the platform of development Weka [11]. At first, the environmental audio feature data is converted to 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 char of environmental audio classification is listed in Fig. (4).

Bagging, AdaBoost and Random forests are involved in the ensemble strategies. In the experiment, the data are selected from the total according to one third of the amount of sampling as the total amount of data in each category.

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 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 adopts the decision tree J48. The rate 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 procedures are based on the cooperation of variable importance for ranking and model estimation to evaluate and compare a family of models.

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

<table>
<thead>
<tr>
<th>Class</th>
<th>Total Frames</th>
<th>75%--Training</th>
<th>25%--Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>bird</td>
<td>751</td>
<td>563</td>
<td>188</td>
</tr>
<tr>
<td>wind</td>
<td>3102</td>
<td>2326</td>
<td>776</td>
</tr>
<tr>
<td>rain</td>
<td>3000</td>
<td>2250</td>
<td>750</td>
</tr>
<tr>
<td>frog</td>
<td>1472</td>
<td>1104</td>
<td>368</td>
</tr>
<tr>
<td>thunder</td>
<td>808</td>
<td>606</td>
<td>202</td>
</tr>
<tr>
<td>total</td>
<td>9133</td>
<td>6849</td>
<td>2284</td>
</tr>
</tbody>
</table>

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

Table 1. Information of dataset in the experiments.

![Image](image-url)
Table 2. Classification performance of ensemble classifier under different percent training sample.

<table>
<thead>
<tr>
<th>Training Samples</th>
<th>Bagging</th>
<th></th>
<th>AdaBoost</th>
<th></th>
<th>Random Forest</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>kapp</td>
<td>err</td>
<td>kapp</td>
<td>err</td>
<td>kapp</td>
<td>err</td>
</tr>
<tr>
<td>10%</td>
<td>0.88</td>
<td>8.77%</td>
<td>0.8912</td>
<td>8.36%</td>
<td>0.9100</td>
<td>6.59%</td>
</tr>
<tr>
<td>20%</td>
<td>0.9121</td>
<td>6.45%</td>
<td>0.9284</td>
<td>5.26%</td>
<td>0.9424</td>
<td>4.22%</td>
</tr>
<tr>
<td>40%</td>
<td>0.9409</td>
<td>4.34%</td>
<td>0.9475</td>
<td>3.86%</td>
<td>0.9574</td>
<td>3.13%</td>
</tr>
<tr>
<td>60%</td>
<td>0.9488</td>
<td>3.78%</td>
<td>0.956</td>
<td>3.23%</td>
<td>0.9656</td>
<td>2.54%</td>
</tr>
<tr>
<td>80%</td>
<td>0.9565</td>
<td>3.33%</td>
<td>0.9619</td>
<td>2.80%</td>
<td>0.9688</td>
<td>2.29%</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Training Samples</th>
<th>MFCC</th>
<th>CELP</th>
<th>MFCC+CELP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>kapp</td>
<td>err</td>
<td>kapp</td>
</tr>
<tr>
<td>10%</td>
<td>0.9098</td>
<td>6.58%</td>
<td>0.8551</td>
</tr>
<tr>
<td>20%</td>
<td>0.9334</td>
<td>4.88%</td>
<td>0.8643</td>
</tr>
<tr>
<td>40%</td>
<td>0.9499</td>
<td>3.68%</td>
<td>0.8851</td>
</tr>
<tr>
<td>60%</td>
<td>0.9576</td>
<td>3.12%</td>
<td>0.8914</td>
</tr>
<tr>
<td>80%</td>
<td>0.9603</td>
<td>2.92%</td>
<td>0.8935</td>
</tr>
</tbody>
</table>

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

category. However, it is almost impossible to carry out the interpretation with classifier ensembles based on multiple decision trees. A RF allows for assessment of the importance of the variables by means of the Gini Index and the OOB subset. In the experiments, three kind features are adopted. First is the MFCC features such variables in dataset as $x_1, x_2, \ldots, x_{13}$. The second is the CELP features including $x_{14}, x_{15}, \ldots, x_{23}$. The third is the combination of the MFCC and CELP, ranged from $x_1$ to $x_{23}$. With random forests, three classification models are built based on the different kind features. The performances of them are shown in Table 3. The model with MFCC+CELP is best among them due to the 23 dimensions variables. From the result, the MFCC features play an importance role in the environmental audio classification, so it obtains better predict result 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 importance 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 includes $x1, x2, x3, x4, x5, x6, x7, x9, x10, x13, x14, x15, x16, x17$ and $x18$ variable.

The classification model is also built from the selected features. The comparison of two models classification performance and the time of built the model are illustrated in Table 4.

Seen from the Table 4, the accuracy of classification model with Selected Features is slightly superior to the
The performance of classification still keeps good. And the time for building the classification model is shown in the Table 4. Due to reduce 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 prediction performance in both accuracy and efficiency than that of all features. The performance and efficiency of different features for RF classification are illustrated in Fig. (7) and Fig. (8), respectively.

**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 such as MFCC and CELP are involved in the experiments. 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. The selected features by the GINI index (variable importance) make changes to the training data, reduce the dimensions of feature, but 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 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

© 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/3.0/) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.