ISSN 1746-7659, England, UK
**Journal of Information and Computing Science**
Vol. 1, No. 2, 2006, *pp*. 115-118

# Hierarchical Support Vector Machines for Audio Classification *

Xin He [1 +], Yaqin Zhao [1] and Xianzhong Zhou [2]

[1] School of Automatic Control, Nanjing University of Science and Technology, Nanjing 210094, China
[2] School of Management and Engineering, Nanjing University, Nanjing 210093, China

( *Received October 19, 2005, Accepted January 30, 2006* )

**Abstract.** Audio data is one of typical multimedia data and it contains plenty of information. Audio retrieval is becoming important content in multimedia information retrieval. In multimedia retrieval researches, it becomes more and more important research part how to construct better classifiers for audio classification and retrieval. Support Vector Machines, a novel method of the Pattern Recognition, presents excellent performance in solving the problems with small sample, nonlinear and local minima. But audio classification is a multi-class classification problem and it's just one of problems to be solved in SVM researches. In this paper, it compares several common Support Vector Machines and proposes a hierarchical Support Vector Machines based on audio features cluster method, combining audio features and hierarchical SVMS. It uses hierarchical classification method to classify audio data and it's proved better performance by experiments.

**Keywords:** support vector machines (SVMs), feature cluster, audio features

## 1. Introduction

The Statistical Learning Theory is the theory for research machine learning rules in small sample conditions. And Support Vector Machines is kernel content of the Statistical Learning Theory, which is based on one of kernel concepts of the Statistical Learning Theory—VC dimension and the theory of Structure Risk Minimization. It can overcome problems of dimension and local minimization, which exist in traditional machine learning methods [1, 2]. SVMs have been applied in Pattern Recognition, Regression Estimation and Probability Density Function Estimation, etc. For example, it uses SVMs to solve classification problems in Pattern Recognition, such as speech recognition, face recognition and handwritten numeral recognition, etc.

There are two main problems to be solved in the SVMs research. The first one is how to effectively apply separation methods used in two-class to multi-class classification problems, namely classification of multi-class. Some effective classifiers for multi-class are designed based on classifier for two-class, such as 1-v-1, 1-v-r and Decision Directed Acyclic Graph (DDAG) [3,4], etc. The second problem is the sensitivity to the noisy data.

In the paper, it discusses some usual SVMs classification methods [5-7] and presents a method, which applies hierarchical Support Vector Machines based on feature cluster to audio classification. The rest of this paper is organized as follows. In Section 2 some audio classification and audio extraction are introduced. Section 3 describes several common SVMs methods [5-7]. In Section 4, the method of hierarchical Support Vector Machines based on feature cluster is presented. Finally, in Section 5, experiments and evaluations on audio data are given.

## 2. Audio Features

It's always based on audio features for audio classification. Audio feature extraction becomes one of emphases in audio classification. Feature extraction includes the temporal and spectral characteristics. Some usual features will be described in this section.

* The research is supported by the Natural Science Foundation of Jiangsu province in China BK2004137).
+ Corresponding author. *E-mail address*: kernel_he@hotmail.com.

### 2.1. Temporal characteristics

It's one of usual methods to capture the temporal characteristics. And it's also used extensively. As a kind of temporal signal, it's intuitionistic to capture the temporal characteristics. These features are Short-Time Average Zero-Crossing Rate, Short-Time Energy, etc. For example, Short-Time Average Zero-Crossing Rate is proved to be useful in characterizing different audio signals, especially in speech/music classification algorithms.

### 2.2. Spectral characteristics

In audio theory, each audio signal is composed of sound wave in different time, frequency and energy amplitude. Different frequency components can be decomposed by the Fourier Transform. There are several Spectral characteristics such as LPC, MFCC, etc. LPC is extracted from every Shot-Time audio clip. Static characteristics of audio in short time can be described by LPC and dynamic characteristics can be described by difference of LPC. Audio data are processed with Z-transform and logarithm processing, and then MFCCs can be obtained, described as Figure 1. 12-order MFCCs are usually used because of its good discriminating ability.

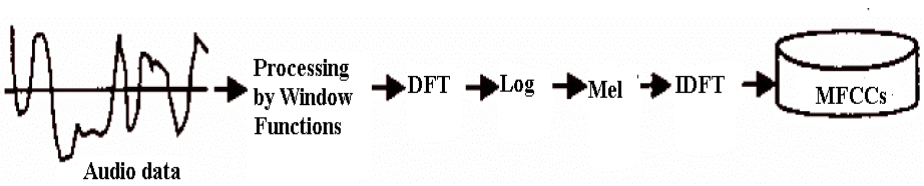


Fig. 1: Mel-scaled frequency Cepstral coefficients.

## 3. Typical SVM Methods

### 3.1. Standard SVMs classification of two-class

Supposing a set of training data,

$$D = \{(x_1, y_1), (x_2, y_2), \cdots, (x_m, y_m)\}$$

where $x_i \in R, y_i \in \{-1,1\}$, and $y_i$ is a class label.

The purpose of SVMs is to find decision function $f(x)$ by training, which can construct separating hyper-planes to classifier two different samples and maximize the margin. That is to say to solve the problem of quadratic programming [2], described as follows:

$$\min \varphi(\omega, \xi) = \frac{1}{2}(\omega, \omega) + C\sum_{i=1}^{m}\xi_i$$
$$\text{s.t.} \quad y_i[(\omega \bullet x_i) + b] \geq 1 - \xi_i$$
$$\xi_i \geq 0, i = 1,2,\cdots,m$$

It can be resolved into extreme value problem of quadratic function by seeking dual problem of the above equation:

$$\max \sum_{i=1}^{m}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{m} y_i y_j \alpha_i \alpha_j K(x_i, x_j)$$
$$0 \leq \alpha_i \leq C, i = 1,2,\cdots,m$$
$$\text{s.t.} \quad \sum_{i=1}^{m}\alpha_i y_i = 0$$

Then the separation decision function can be described as follows after finding the optimal solution:

$$f(x) = \text{sign}[\sum_{i=1}^{m}\alpha_i y_i K(x, x_i) + b]$$

### 3.2. Multi-class classification

There are mainly two solutions in multi-class separation. Firstly, several classifiers of two-class are combined to apply to separate multi-class, including 1-v-1, 1-v-r and DDAG, etc. If classifying N classes, it needs N two-class classifiers by using 1-v-1, N(N-1)/2 by 1-v-r, and it also needs N(N-1)/2 by DDAG. These methods are faster than typical algorithms, but the number of classifiers becomes more and more as the

number of classes increasing. Secondly, hierarchical classifier tree is constructed. In the tree, each node denotes a two-class classifier, which can complete a predefined separation. In this method, the structure of classifiers is simple and clear. The predefined separation can be defined according features of sample vectors. It also needs less number of classifiers than algorithm described above and the runtime is shorter. There exits obvious predominance as long as the layers of separation subtask is defined logically. In this paper, it's just based on features of audio data to cluster and complete separation.

# 4. Hierarchical SVM Audio Classification Method Based on Feature Cluster

## 4.1. Hierarchical SVM classification tree

One of important problems in multi-class pattern recognition is ambiguous class, which means that there is a set of subset including classes of 1~N, and all sample feature vectors are similar. Light error in metrical features can result in misclassification. The main idea of multi-class classifiers is first to separate ambiguous class roughly, which is to be classified, then fine classify.

The definition of separation subtask is the kernel section of the design of hierarchical classification tree. And it's referred to efficiency and performance of classification tree. Figure 2 describes this method.

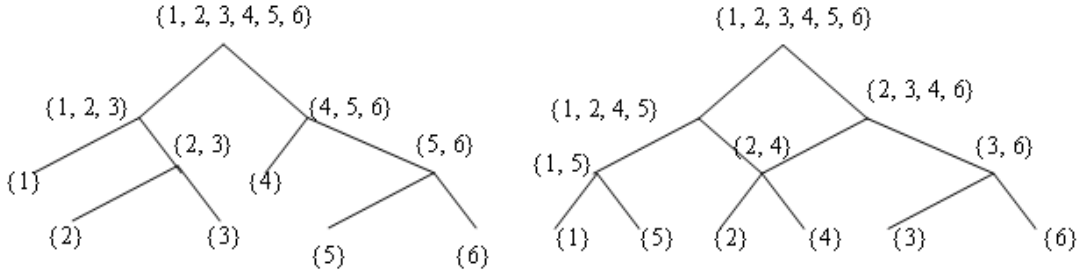


Fig. 1: Examples of hierarchical classification trees

## 4.2. Hierarchical SVM classification tree based on feature cluster

There are extensive sounds in the nature, such as speech, music and environment sound, etc. People can apperceive sounds by aural signature. Naturally, they hope to retrieve audio information by aural signature too. It can be realized to classify and retrieve audio data only when extensive aural signature is obtained from audio data. Main features of audio data are composed of temporal and spectral characteristics. Temporal characteristics mainly include average energy, zero-crossing rate (ZCR) and silencing ratio, etc. Spectral characteristics include bandwidth, frequency spectrum, harmonic and tone, etc. In this paper, these characteristics are used in audio feature cluster and hierarchical SVM is combined to classify audio data.

The method is described as follows. Supposing there are N classes audio to be classified.

Step 1: Dispose N classes audio data by the Fast Fourier Transform (FFT), then calculate audio features accordingly by different algorithms, including bandwidth, ZCR, mel-frequency cepstral coefficients (MFCCs), etc.

Step 2: According as different audio features, separate audio data into two classes by feature cluster.

Step 3: Repeat Step 2 to each sub-class until each there is only one class in each sub-class.

Step 4: Based on results of Step 2 and 3, define classification subtask respectively and construct the separation tree of hierarchical SVM, where each node is according to a two-class classifier.

In the Statistical Learning Theory, features between different samples can be enhanced in non-linear mapping. Then it can be accomplished in higher space to separate audio data and establish separation trees by using SVMs, and separation effects are more obvious.

# 5. Experiments

The database used in experiments is composed of recorded broadcasting program, including pure speech (male and female), speech with music, silence and music. Every audio class is about 20 minutes with sample rate at 16 kHz. The database is partitioned into a training set and a test set, it's drawn out 60-group audio data respectively from every audio class with 10 seconds long, 40 groups for training set and 20 groups for

testing group. The results are listed in Table 1.

Table 1.Experiment Results (N=5)

| Classify type / Classifying results | | Pure speech | | Speech with music | Silence | Music | Acc. |
|---|---|---|---|---|---|---|---|
| | | Male announcer | Female announcer | | | | |
| Pure speech | Male announcer | 10 | 0 | 0 | 0 | 0 | 100% |
| | Female announcer | 0 | 9 | 0 | 0 | 1 | 90% |
| Speech with music | | 0 | 0 | 19 | 0 | 1 | 95% |
| Silence | | 0 | 0 | 0 | 20 | 0 | 100% |
| Music | | 0 | 1 | 1 | 0 | 18 | 90% |

Furthermore, these data are also used test traditional SVM algorithms, in order to compare with the method used in this paper. In Table 2, results are presented.

Table 2.Result of Comparison (N=5)

| Classification algorithm | Number of Classifiers |
|---|---|
| Method in this paper | 4 |
| 1-v-1 | 10 |
| 1-v-r | 5 |
| DDAG | 10 |

## 6. Conclusion

In this paper, it's discussed in details a classification problem of multi-class audio data. Hierarchical SVM is used with audio features and it's presented hierarchical SVM based on feature cluster for audio classification and retrieval. It's also shown that the performance of this method achieve high accuracy in audio classification.

As for future research, we will improve this method for more audio classes and take emphases on feature set definition and how to apply audio analysis to video retrieval.

## 7. Acknowledge

The authors are very grateful to Dr. Yingchun Shi and Dr. Bing Huang who give many valuable comments and advices and the anonymous referees for their helpful comments.

## 8. References

[1] V. Vapnik, Statistical Learning Theory, New York: Springer Verlag, 1998.

[2] J. C. Burges, A tutorial on support vector machines for pattern recognition, Data Mining and Knowledge Discovery, 2(1998)2, 121-167.

[3] C. W. Hsu, C. J. Lin, A comparison of methods for multiclass support vector machines, IEEE Transactions on Neural Networks, 13(2002)2, 415-425.

[4] John C.Platt. Large Margin DAGs for Multiclass Classification, Neural Information Processing Systems 12, MIT Press, 2000, 547-553.

[5] Girolami M, Mercer Kernel Based Clustering in Feature Space, IEEE Transactions on Neural Networks, 13(2002)3, 780-784.

[6] X. Zhang, Using class-center vectors to build support vector machines. In Proceedings of NNSP'99, 1999.

[7] John C. Platt, Fast Training of Support Vector Machines using Sequential Minimal Optimization. In Scholkopf B. et al(ed.), Advances in Kernel Methods-Support Vector Learning, Cambridge, MA, MIT Press, 1999, 185-208.