AUDIO MUSIC GENRE CLASSIFICATION USING SUPPORT VECTOR MACHINE

JiePing Xu	Chao Zhen	Tao Zheng
School of Information	School of Information	School of Information
Renmin University of China	Renmin University of China	Renmin University of China
xjieping@ruc.edu.cn	zhenchao@ruc.edu.cn	w15784463@sina.com

ABSTRACT

The system submitted to the MIREX Audio Music Genre Classification task is described here. It uses a set of 78 features and a Support Vector Machine classifier to predict the genre cluster. The features are timbre texture features, rhythmic content features and pitch content features. The features were selected previously according to experiments on our annotated databases. And SVM is used as basic classifier.

1. INTRODUCTION

There has been a significant amount of researches about genre classification. Generally, intensity, timbre and rhythm are thought to be important factors in representing genre information. In this system, according to previous experiments and feature selection results, we extracted 78 features. SVM is chosen as basic classifier.

2. GENRE CLASSIFICATION

To know which features are relevant, we have made a previous analysis using different selection methods on George 2002 dataset. Using several feature selection methods, such as Sequence Feature Selection, ReliefF, Principal Component Analysis. At last, we have sorted out 78 features of different kind. Then we extracted these features and normalize them and finally we train a SVM model. The algorithm is implemented in C++ and complied as Win32 binary.

2.1 Feature Extraction

Each clip is divided into 0.5 overlapping 32ms-long frames. The extracted features fall into four categories: timbre, intensity, rhythm and others. Finally, we got 78 acoustic features to express mood information.

2.1.1 Timbre Features

In our experiments, timbre features were particularly helpful to classify by genre the exemplar songs provided. We decided to use:

-Spectral Centroid [1]
-Spectral Rolloff Point
-Spectral Flux
-MFCC (Mel-Frequency Cepstrum Coefficient)
-Zero Crossing
-Spectral Variability
-LPC (Linear Prediction Coefficient)

2.1.2 Pitch Features

Testing with our databases we discovered that pitch content features help to discriminate between some genre categories. Here is the list of features computed:

-RMS(Root Mean Square)
-Fraction of Low Energy Windows
-Compactness

2.1.3 Rhythm Features

Finally we also extract rhythm features as follows:

- Strongest Beat
- Beat Sum
- Strength Of Strongest Beat
- Method of Moments

2.1.4 Statistics

Most of these features are extracted using windowing. Afterward we compute statistics of these values (min, max, mean, variance, derivative variance, second-derivative variance). The decision to keep or not each value is made using feature selection methods as previously mentioned.
So the whole feature set and each feature's dimension are list in table 1.

Table 1

Label #	Feature Description	Dimension
1	Spectral Centroid Overall Average/ Derivative Variance	1
2	Spectral Rolloff Point Overall Average/ Derivative Variance	1
3	Spectral Flux Overall Average/ Derivative Variance	1
4	Compactness Overall Average/ Derivative Variance	1
5	Spectral Variability Overall Average/ Derivative Variance	1
6	Root Mean Square Overall	1

	Average/ Derivative Variance	
7	Fraction Of Low Energy Windows Overall Average/ Derivative Variance	1
8	Zero Crossings Overall Average/ Derivative Variance	1
9	Strongest Beat Overall Average/ Derivative Variance	1
10	Beat Sum Overall Average/ Derivative Variance	1
11	Strength Of Strongest Beat Overall Average/ Derivative Variance	1
12	MFCC Overall Average/ Derivative Variance	13
13	LPC Overall Average/ Derivative Variance	10
14	Method of Moments Overall Average/ Derivative Variance	5

2.2 Classifier

Once the features extracted from the audio and normalized, we train a Support Vector Machine model. We use the libSVM [2] library.

3. ANALYSIS OF THE RESULTS

In the experiment, we choose the George 2002 dataset. It includes ten genres, every genre is consist of almost 100 segments whose length is 30 seconds. In our experiment, we choose eight genres from the George 2002. They are Blues, Classical, Country, Jazz, Metal, Pop, Reggae and Rock.

Table 2

train:test =8:2	first	second	third	average
	81.87%	83.13%	79.37%	81.46%

Table 3

	Blues	Classical	Country	Jazz	Metal	Pop	Reggae	Rock
Blues	19	0	0	0	0	0	0	1
Classical	0	18	1	0	0	0	0	1
Country	2	0	15	0	1	0	0	2
Jazz	0	0	0	19	0	0	0	1
Metal	0	0	0	0	19	0	0	1
Pop	0	0	2	1	0	17	0	0
Reggae	2	0	2	0	0	1	13	0
Rock	1	0	2	0	1	2	0	14

From the above table, we can see that the accuracy of the Rock and Reggae genre is lowest among the other genres. This makes the overall accuracy rate of the classification decrease in much degree. Therefore, we investigate which features are good at classifying the music in these two genres. And the result is very interesting. If we only use the 13-dimensional feature MFCC Overall Standard Deviation, then we can get the success rate of 85%. Based on this discovery, in the future we can devise some new classification algorithm to improve the accuracy of these two genres.

4. FUTURE WORK

Many things can be done to improve the system quality. Firstly, more useful features can be explored and used. Secondly, more classifier can be tested and combined to do this experiment. Thirdly, feature selection methods can be improved and find an approach to add into this system, then the selected feature may more effective for classification.

5. ACKNOWLEDGMENTS

We gratefully acknowledge the support of the National Science and technology Basic Works Foundation under grant 60873098.

6. REFERENCES

[1] George Tzanetakis, Perry Cook :“Music genre classification of Audio Signals” *IEEE transactions on speech and audio processing*,Vo1.10, No.5:293-302,2002

[2] C.-C. Chang, C.-J. Lin. “LIBSVM: a library for support vector machines” Software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm/, 2001.