MUSIC MOOD CLASSIFICATION BASED ON VOTING

Yi Liu
School of Information
Renmin University of China
liuyi_lee@126.com

Tao Zheng
School of Information
Renmin University of China
w15784463@sina.com

Yue Gao
School of Information
Renmin University of China
gaoyue-yx@ruc.edu.cn

ABSTRACT

Our system used 78-dimension feature vector to represent audio segments. These features expressed intensity, timbre and rhythm information. Each 30s-long segment is also divided into three 10s-long clips. SVM classifier is used to get the predict mood cluster. And the final result is acquired through voting.

1. INTRODUCTION

There has been a significant amount of researches about mood classification [1] [2] [3] [4] [5] [7]. Generally, intensity, timbre and rhythm are thought to be important factors in representing mood character. Segments duration also affects the predict results [8]. In this system, according to previous experiments and feature selection results, we extracted 78 features on 10s-long clips which are come from 30s-long segments. SVM is chosen as basic classifier.

2. MOOD CLASSIFICATION

From audio data, we extracted 92-dimension features. In order to know which features are relevant, we do some feature selection experiments including ReliefF, Fisher Rule, Step Forward Selection (SFS) and Active Selection on our own dataset [9], and got a feature ranking with relation value. According to a threshold value, we removed the low relation ones, and finally got 78-dimension features in this system. At last we normalized them and 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

Positive songs usually sound brighter, negative ones sound dark and melancholy. The features extracted in our system are as follows. For each clip, we calculated the mean and standard deviation over all frames. This led to 56 acoustic features.

- Centroid (2)
- Rolloff Point (2)
- Flux (2)
- Zero Crossings (2)
- Compactness (2)
- MFCC(26)
- LPC(20)

2.1.2 Intensity Features

Intensity features can be used to judge whether the emotion is very strong or not. In this system, the intensity features we extracted are as follows, by calculating the mean and standard deviation over all frames for each clip, we get 4 acoustic features.

- RMS (2)
- Fraction Of Low Energy Windows (2)

2.1.3 Rhythm Features

Fast songs tend to be happier than slow ones. rhythm features can express some mood information. Here we extracted following one, and finally get 6 acoustic features.

- Beat Sum (2)
- Strongest Beat (2)
- Strength of Strongest Beat (2)

2.1.4 Other Features

According to our previous experiments, some other features also play an active role in affecting music mood. This system used following ones and finally get 12 acoustic features.

- Spectral Variability (2)
- Method of Moments (10)

2.2 Classifier

We used Gauss Mixture Model (GMM) with Expectation Maximum (EM) algorithm and LibSVM in previous experiments on our own datasets, and SVM performed better [9]. In MIREX 2008 audio mood classification task [6], using LibSVM as classifier performed very well. So in this system, we chose LibSVM [10] as basic classifier.

2.3 Post-processing Methods

After classification, we get the predict labels of clips in test set. Each segment's predict label is decided by three clips which are generated by the segment. A difficult situation is that the three clips' labels are different from each other. In this case, the second label would be chosen as final label for the segment. Because the first and last one's mood expression may not very stable attributed to their transition position

3. 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.

4. ACKNOWLEDGMENTS

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

5. REFERENCES

[1] D. Liu, L. Lu, H.J. Zhang, Automatic Mood Detection from Acoustic Music Data, Proc. of the 4th Int. Conf, Music Information Retrieval (ISMIR'03), Washington, DC, USA, October 2003: 81-87

[2] L. Lu, D. Liu, H.J. Zhang, Automatic Mood Detection and Tracking of Music Audio Signals, IEEE Transaction on Audio, Speech, and Language Processing, 2006: 5-18

[3] Feng Y.Z., Zhuang Y.T., Pan Y.H., Popular Music Retrieval by Detecting Mood, Music Information Retrieval (ISMIR'03), July 2003: 375-376

[4] Yang Y.H., Liu C.C., Chen H.H., Music Emotion Classification: A Fuzzy Approach, Association for Computing Machinery (ACM'06), October 2006:81~84

[5] K. Trohidis, G. Tsoumakas, G. Kalliris, and I. Vlahavas, Multi-label Classification of Music into Emotions, Proceedings of the 9th International Conference on Music Information Retrieval, ISMIR'08, Philadelphia, USA, 2008

[6] Hu X., J.Stephen Downie, The 2007 Mirex Audio Mood Classification Task: Lessons Learned, Proceedings of the 9th International Conference on Music Information Retrieval, ISMIR'08, Philadelphia, USA, 2008

[7] T. Li, M. Ogiharai, Detecting Emotion in Music, Johns Hopkins University, 2003

[8] Z. Xiao, E. Dellandrea, W. Dou, and L. Chen, What is the Best Segment Duration for Music Mood Analysis, Content-Based Multimedia Indexing, 2008: 17-24

[9] Y. Liu, Y. Gao, Classification of Music into Emotions based on Keyword Model, proceedings of 17th National Conference of Multimedia Technology, Oct. 25, 2009, Wuhan, China

[10] C.-W. Hsu, C.-C. Chang, C.-J. Lin. A practical guide to support vector classification. Technical report, Department of Computer Science, National Taiwan University. July, 2003.