AUDIO CLASSIFICATION TASKS: MIREX 2015 SUBMISSIONS

Kang Cai, Deshun Yang, Xiaoou Chen
Institute of Computer Science and Technology, Peking University, Beijing, China

ABSTRACT

In this submission system, we extract novel features based on general features such as chroma-based features, Mel Frequency Cepstrum Coefficients (MFCCs) and other spectral features. As for classification method, we use support vector machine (SVM) as the classifier, combining the one-vs-all classification with the two-class classification as the strategy.

1. INTRODUCTION

In our system the same set of features is extracted for all tasks. The feature extraction is implemented in MATLAB and our SVM is implemented with LIBSVM [1]. The detailed information of audio features set is discussed in Section 2. Section 3 introduces the genre/mood/composer modeling and classification method.

2. FEATURE EXTRACTION

2.1 Feature set

The feature set consists of the following features:

1. Harmonic Ratio (1)
2. Fundamental Frequency (1)
3. Chroma Feature [2] (12)
4. MFCCs [3] (13)
5. Short Time Energy (1)
6. Spectral Centroid (1)
7. Spectral Entropy (1)
8. Spectral Flux (1)
9. Spectral Roll Off (1)
10. Zero Cross Rate (1)

All the features are extracted in a short hamming window of around 46ms with 50% overlap. For each window, 33 dimensional features are extracted.

2.2 Statistical method

An efficient and effective method of statistics for features of all the windows in a piece of music is to calculate the means and variances. However, the windows of a piece of music construct a time series and the inner-connection between those windows cannot be revealed simply through means and variances, we, therefore, seek a proper way to reflect this connection in terms of time.

In this system, we build an Auto-Regressive (AR) and Moving Average (MA) Model [5] to sort out the relationship between windows in terms of time. First of all, we analyze the features of all windows and sequence them in the light of time. Each dimension of the features forms an independent time series. Then, we gain new parameters by modeling those time series using the AR and MA Model. These parameters, together with means and variances, form the new features, among the 165 dimensions of which, means amount to 33 dimensions, variances 33 dimensions, AR model 33 dimensions and MA model 66 dimensions.

3. CLASSIFICATION

We used the commonly-used support vector machine (SVM) as our classifier. To deal with the imbalance of positive and negative training samples, we set the penalty factor according to the proportions of both samples.

The strategies we adopt include one-vs-all (OvA) [4] and two-class classification [6]. First, We use OvA to predict the test instance to get its the most likely Class C_1 and the second possible Class C_2 initially. If the probability of Class C_1 is significantly higher than that of Class C_2, then the result of classification is C_1. While if the difference between the probabilities of Class C_1 and C_2 is not significant, we will use two-class classification and further judge the class of test instances according to the results of both OvA and two-class classification.

4. SUBMISSIONS

Three submissions are as follows:

Submission 1 uses total 165 dimensional features and one-vs-all classification.

Submission 2 uses total 165 dimensional features and the strategy combining one-vs-all classification with two-class classification.

Submission 3 uses 66 dimensional features of means and variances, and one-vs-all classification as the baseline.

5. ACKNOWLEDGEMENTS

This work was supported by National Hi-Tech Research and Development Program (863 Program) of China under Grant 2014AA015102 and the Natural Science Foundation of China (Multimodal Music Emotion Recognition technology research No.61170167).

6. REFERENCES

[1] Chih-Chung Chang and Chih-Jen Lin. Libsvm: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3):27, 2011.

[2] Daniel PW Ellis. Classifying music audio with timbral and chroma features. In *ISMIR 2007: Proceedings of the 8th International Conference on Music Information Retrieval: September 23-27, 2007, Vienna, Austria*, pages 339–340. Austrian Computer Society, 2007.

[3] Beth Logan et al. Mel frequency cepstral coefficients for music modeling. In *ISMIR*, 2000.

[4] Ryan Rifkin and Aldebaro Klautau. In defense of one-vs-all classification. *The Journal of Machine Learning Research*, 5:101–141, 2004.

[5] Said E Said and David A Dickey. Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika*, 71(3):599–607, 1984.

[6] David MJ Tax and Robert PW Duin. Using two-class classifiers for multiclass classification. In *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, volume 2, pages 124–127. IEEE, 2002.