Combining Spectral And Rhythmic Features FOR Music Genre Classification

Xinquan Zhou
Georgia Institute of Technology
Center for Music Technolgy
roy.zxq@gmail.com

Guang Li
Tianjin University
lguang@live.cn

Shumin Shu
Rochester University
Shumin.xu@rochester.edu

Yating Gu
Rochester University
yatinggu@rochester.edu

ABSTRACT

Music genre classification is an interesting and challenging task in the field of music information retrieval. And feature extraction is significantly crucial for this task. In this paper, we not only use the traditional MFCCs, but also use Logarithmic Fluctuation Pattern (LFP) in the Mel-domain and MFCCs extracted from percussive detection. Based on these features, we then employ SVM to do the final classification task. Experimental results shows that it will increase the accuracy by more than 8% compared with only using MFCCs.

1. INTRODUCTION

There has been lots of research on music genre classification over decades. Classical method is to extract some spectral features and train classifiers to make decision. Noticed that rhythm matters a lot for music genre, features that can represent the rhythm information are suitable for this particular task. LFP can represent the rhythmic structure of a song proposed by [1]. Also, percussive information determines the rhythm for some songs in some sense. Therefore percussive detection can help inspired by [2]. Final we employ SVM to perform the final classification.

2. FEATURE EXTRACTION

The first feature we use is traditional 20-dimonsion MFCCs. MFCCs is demonstrated to be really useful in many music and speech applications. Secondly, we extract LFP based on the cent-scaled magnitude spectrum in the block level: Resampling the input wav to 22050 Hz, applying STFT using a window size of 1024 samples, a hop size of 512 samples. Then mapping the magnitude spectrum onto the logarithmic Cent scale. And we use a block size of 128 and a hop size of 32 to extract LFP. Finally inspired by [2] who extracted MFCCs based on rhythm domain. We extracted 20-dimonsion MFCCs (MFCCs_percussive) from the results of percussive detection, which we hope to represent the some kind of fluctuation pattern of percussive information. Since both MFCCs and MFCCs_percussive are frame-based features, and LPF is block-based feature. To combine these features, we use the same block as LFP to pool the MFCCs and MFCCs_percussive. We calculated mean and standard deviation within the pooling windows. And standard deviations of the first derivative as well as the second derivative of pooling windows are also computed.

3. MODEL

3.1 SVMs

Supported Vector Machines (SVMs) is a really powerful classifier in many cases. Since our feature dimension is more than 100 and SVMs is not very sensitive to feature dimension, it becomes the suitable classifier for our scenario.

3.2 GMM

Gaussian Mixture Model can provide some complementary information for SVMs for that GMM is not a discriminant model and only focus on each class independently. We use MFCCs to train 10 GMM for each class. And use the posterior probability of the input as a kind of high-level features (we called it GSV inspired by [3]) combined with the features above to train SVMs.

4. CLASSIFICATION

We use two classification scenarios. One is just use the combined features to train SVMs and do the prediction. We use the libSVM toolkit for this task. Secondly we use the combined features to train the first SVMs, and use the output probability of this SVMs as in new features to train a second layer SVMs, which is a stack of 2 SVMs.

4.1 Single SVMs

Since after pooling, all features become block level. Therefore each block can be represented by a feature vector, which constitutes a single training sample for SVMs. In the test phase, input wav is also divided into several blocks and each block is a single test sample for SVMs. And we use a simple voting mechanism to decide which class is the whole input song.

4.2 Stack SVMs

The first SVMs is trained as exactly the same way as the single SVMs. And we use the output probability of the first SVMs as the high-level features to train the second Footnotes and Figures

5. REFERENCES

[1] Seyerlehner K, Widmer G, Schedl M, et al. Automatic music tag classification based on block-level[J]. Proceedings of Sound and Music Computing 2010.

[2] i Termens E G. Audio content processing for automatic music genre classification: descriptors, databases, and classifiers[D]. Universitat Pompeu Fabra, Barcelona, 2009.

[3] Wu M J, Chen Z S, Jang J R, et al. Combining visual and acoustic features for music genre classification[C]//Machine Learning and Applications and Workshops (ICMLA), 2011 10th International Conference on. IEEE, 2011, 2: 124-129.