MIREX AUDIO GENRE CLASSIFICATION

Huaxin Wang
School of Electronics Engineering and Computer Science, Peking Univisity
wanghxcis@gmail.com

ABSTRACT

This extended abstract details a submission to the Music Information Retrival Evaluation eXchange in the Audio Genre classification task, including the feature set used, the feature selection methods and the final classification method.

Index Terms— Audio, Genre, Classification

1. FEATEURE SET

We mainly use timber features to classify the music genre. The features included are Mel-frequency cepstral coefficients (MFCC), the spectral centroid, the MPEG 7's audio spectral flatness(ASF), and some AM modulation features. The MFCC feature is about to characterize the whole spectral envelop and we select the first 13 coefficients as our feature. The spectral centroid is about to characterize the music brightness and darkness. The other two features are more used in the music instrument recgnition area(Antti 2001 and Slim 2006), and we belive the genre is much more related with the instrument used in the music. The audio spectral flatness is about to charaterize the flatness of each spectral band. We devide the whole frequency band into 20 subbands and the band width is in logarithmic scale, then we calulate the ASF of each band. The last feature is about AM modulation. Strenth, frequency and heuristic strength of amplitue modulation is measured at two frequency band, 4~8Hz and 10~40Hz. In each band, FFT analysis is carried a widnowed envelope and the maxima peak is located. The peak frequency , the difference between the maxima and the band average amplitude and the difference between the maxima and the average amplitude of the whole frequency range under consideration are used as AM features. In summary, we have 13 dimensions for MFCC, 1 dimension for spectral centroid, 20 dimensions for ASF and 3 dimensions for AM modulation for each frame. The frame length is 46.4ms, and the frame shift is 23.2ms. We calucate the mean and variance of each dimension of the feature in 1s and use those as the input for feature selection. At last, for 1s there are 74 real numbers as the output featuer.

2. FEATURE SELECTION

We use a simple and intuitive approach called inertia ratio maximization using feature space projection(IRMFSP) as our feature selection method(Geoffroy 2003). This method is about to maximize the ratio between the inter-class distance and the inner-class distance on each dimension of the feature set. In the total 74 dimensions we selected only 40 dimensions which maximize the ratio. We select the feature for each pair of the class, so in some cases one dimesion of a certain class is used when compared with some class and is not used when compared with others.

3. CLASSIFATION

We use the support vector machine as our classifier. We modified the libsvm software package to suit our feature selection method. We use the rbf kennel, and find the best parameters of svm by 5-fold crossvalidaion. We also use voting method for the last classification stage.

4. References

[1] Geogre Tzanetakis, Perry Cook "Music Genre Classification of Audio Signals" IEEE Transaction on Speech and Audio Proceeding, VOL. 10, NO.5, JULY 2002

[2] Antti Eronen "Automatic Music Instrument Recognition" Master Thesis 2001

[3] Slim Essid, Gael Richard, and Bertrand David "Instrument Recognition in Polyphonic Music Based on Automatic Taxonomies" IEEE Transaction on Audio, Speech, and Language Processing, VOL.14, NO.1, JAN 2006

[4] Geoffroy Peeters "Automatic classification of large musical instrument databased using hierarchical classifiers with inertia ratio maximization" AES 115th Convertion, New York, NY, USA, 2003