Fast Genre Classification and Artist Identification

George Tzanetakis and Jennifer Murdoch
University of Victoria
Computer Science Department (also in Music)
gtzan@cs.uvic.ca

ABSTRACT

This abstract describes the audio feature extraction and classification algorithm used for the University of Victoria submission to the MIREX (Music Information Retrieval Exchange) 2005. The same audio features and classification algorithm are used for the audio genre classification and artist identification tasks and arge therefore combined in this abstract. The feature set used is similar to one described in Tzanetakis and Cook (2002) but not including the Beat and Pitch Histogram features. This decision was made in order to address stability problems with the above features and computation speed requirements. Even though our submission did not win the contest it was by far the fastest one.

Keywords: audio classification, genre classification, artist identification

1 INTRODUCTION

The approach used by the University of Victoria submission to the MIREX 2005 audio genre classification, and artist identification is based on the calculation of the features described in Tzanetakis and Cook (2002). The *Marsyas-0.2* free software framework for audio analysis and synthesis has been used for the implementation [1].

The features used to represent timbral texture are based on standard features proposed for music-speech discrimination, speech and general audio and music classification. They consist of a set of 4 features computed based on the Short Time Fourier Transfrom (STFT) magnitude spectrum such as the Spectral Centroid (defined as the first moment of the magnitude spectrum) as well as the 13 Mel-Frequency Cepstral Coefficients (MFCC) S. and Mermelstein (1980)]. These features are computed using an analysis window of 20 milliseconds. Means and variances of the features over a larger texture window (1 second) with a hop size of 20 milliseconds are computed resulting in a set of 18 features every 20 milliseconds. An additional feature (the percentage of low energy frames over the texture window) results in a timbral texture feature vector of 19 dimensions. These features are described in more detail in Tzanetakis and Cook (2002).

[1] http://marsyas.sourceforge.net

Classification decisions and training are made for audio segments of approximately 3 seconds. The 18 features are summarized using mean and variances over the 3 seconds resulting in a single feature vector for each 3 second snippet. A linear support vector machine classifier (LSVM) was trained using the SMO (Sequential Minimum Optimization) provide in the Weka machine learning toolkit. The outputs of the classifier were mapped to probabilities using logistic regression and the classification decision over the entire song was done by taking weighted (by the classifier outputs) sums for each class and selecting the one with the highest sum.

2 DISCUSSION OF EVALUATION RESULTS

The results of the genre classification and artist identification contest are not yet completed but from the current numbers the classification accuracy of our submission was not as good as most of the other entries. This is good news as it shows that research in genre classification and artist identification has made progress since 2002 when Tzanetakis and Cook (2002) was published. Unfortunately, we didn't manage to include the Beat and Pitch Histogram features which possibly could provide better classification accuracy. The reasons include: time pressure, instability of the algorithms (they wouldn't work properly for certain soundfiles) and high computational cost.

On the positive side, our algorithm was the fastest in terms of run-time performance. As our goal is to scale to very large datasets and have real-time classification systems this is good news. It also demonstrates some of the performance advantages of using *Marsyas-0.2*. As most of the proposed algorithms really on similar feature extraction algorithms and machine learning blocks I believe that it would be straightforward to port the other submission to *Marsyas-0.2* achieving significant speedups in runtime performance. We are looking forward to such collaborations.

ACKNOWLEDGEMENTS

Many thanks to Stuart Bray, Ajay Kapur, Adam Parkin, Adam Tindale, Richard McWalter and Antonin Stefanutti for help with the development of Marsyas.

References

Davis S. and P. Mermelstein. Experiments in syllable-based recognition of continuous speech. *IEEE Transcactions on Acoustics, Speech and Signal Processing*, 28:357–366, August 1980.

G. Tzanetakis and P. Cook. Musical Genre Classification of Audio Signals. *IEEE Trans. on Speech and Audio Processing*, 10(5), July 2002.