USING TIMBRE, RHYTHM AND TEMPO MODELS FOR MUSIC GENRE CLASSIFICATION

Franz de Leon
University of Southampton
fadl1d09@ecs.soton.ac.uk

Kirk Martinez
University of Southampton
km@ecs.soton.ac.uk

ABSTRACT

In this submission, audio features that approximate timbre, rhythm and tempo are used for genre classification and music similarity estimation. This abstract describes the feature set, distance computation method, and classifier model used for the submitted algorithms.

1. INTRODUCTION

In our system, audio data are modeled as long-term accumulative distribution of frame-based spectral features. This is also known as the "bag-of-frames" (BOF) approach wherein audio data are treated as a global distribution of frame occurrences. This approach is widely used in MIREX submissions. For MIREX 2010 genre classification and audio similarity estimation task, the BOF approach was used for some of the top performing systems[1][2] .

The features that are extracted from audio files are approximations of timbre, rhythm and tempo. The feature extraction, distance computation and classification algorithms are implemented in MATLAB®.

2. FEATURE EXTRACTION

This section describes the processes involved in feature extraction. More detailed explanation can be found on the cited references.

2.1 Audio Preprocessing

The input signal is assumed to be sampled at 22050 Hz, as specified in MIREX wiki[1]. The audio signal is normalized and preprocessed to remove inaudible parts. The signal is then cut into frames with a window size 512 samples (~23 msec.) and hop size 512 samples.

2.2 Timbre Component

The timbre component is represented by the Mel-Frequency Cepstral Coefficients (coefficients 2:20) [3]. We then model the distribution of the MFCCs for the audio file using a Gaussian mixture model (GMM). In this work, we use a single Gaussian represented by its mean μ and covariance matrix Σ. This feature is complemented by appending its time derivative.

2.3 Rhythm Component

The rhythm component based on the Fluctuation Patterns [4] (FPs) of the audio signal. Fluctuation patterns describe the amplitude modulation of the loudness per frequency band.

For each frame, the fluctuation pattern is represented by a 12x30 matrix. The rows correspond to reduced Mel-frequency bin while the columns correspond to modulating frequency bands. To summarize the FPs, the median of the matrices is computed. Additional features derived are FP mean and FP standard deviation.

2.4 Tempo Component

The tempo component is derived from a technique using onset autocorrelation [5]. The tempo is computed by taking the first-order difference along time of a Mel-frequency spectrogram then summing across frequency. A high-pass filter is used to remove slowly-varying offsets. The global tempo is estimated by autocorrelating the onset strength and choosing the period with the highest windowed peak.

3. GENRE CLASSIFICATION

The timbre, rhythm, and tempo distances are calculated separately. Before they are combined, each distance component is normalized by removing the mean and dividing by the standard deviation of all the distances. Symmetry is obtained by summing up the distances in both directions for each pair of tracks [6].

Distances between timbres are computed by comparing the GMM models. We use symmetric Kullback-Leibler (SKL) distance between two models [7]. The SKL distances are transformed into metric by getting the root of the logarithm for each distance measure. The Euclidean distance is used to compute distance between rhythms. For tempo distances, a simple absolute distance is computed.

A direct approach to combine timbral similarity with other features is to compute a weighted sum of the individual distances. Each distance component is normalized by removing the mean and dividing by the standard deviation of all the distances. The system is then optimized by

[1] http://www.music-ir.org/mirex/wiki/2011:Audio_Music_Similarity_and_Retrieval

determining the appropriate weights for each distance component. Finally, all the distances are tabulated to form a full distance matrix. For a given unlabeled track, the genre of the nearest track based on the weighted sum of distance features is used to classify it.

4. RESULTS

This submission is an updated version of the algorithm submitted to MIREX 2011 Train/Test task. From the MIREX 2011 data, the classification accuracy obtained was 58.51%. For comparison, the best performing system [8] in MIREX 2011 had a classification accuracy of 80%.

5. ACKNOWLEDGMENTS

Mr. Franz de Leon is supported by the Engineering Research and Development for Technology Faculty Development Program of the University of the Philippines, and DOST.

6. REFERENCES

[1] K. Seyerlehner, M. Schedl, T. Pohle, and P. Knees, "Using Block-Level Features for Genre Classification , Tag Classification and Music Similarity Estimation," in *Submission to Audio Music Similarity and Retrieval Task of MIREX 2010*, 2010.

[2] K. Seyerlehner, G. Widmer, M. Schedl, and P. Knees, "Automatic Music Tag Classification Based on Block-Level," in *Proceedings of Sound and Music Computing 2010*, 2010.

[3] B. Logan and A. Salomon, *A music similarity function based on signal analysis.* IEEE, 2001, pp. 745-748.

[4] E. Pampalk, "Computational Models of Music Similarity and their Application in Music Information Retrieval," Vienna University of Technology, 2006.

[5] D. P. W. Ellis and G. E. Poliner, "Identifying `Cover Songs' with Chroma Features and Dynamic Programming Beat Tracking," in *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07*, 2007, p. IV-1429-IV-1432.

[6] E. Pampalk, "Audio-Based Music Similarity and Retrieval : Combining a Spectral Similarity Model with Information Extracted from Fluctuation Patterns," in *Submission to MIREX 2006*, 2006.

[7] M. I. Mandel and D. P. W. Ellis, "Song-level Features and Support Vector Machines for Music Classification," in *Submission to MIREX 2005*, 2005, pp. 594-599.

[8] P. Hamel, "Pooled Features Classification MIREX 2011 Submission," in *Submission to Audio Train/Test Task of MIREX 2011*, 2011.