Classification of modern pop songs into ‘Day’ and ‘Night’ songs

Raunaq Aman Jaswal

Department of Electrical and Computer Engineering

ABSTRACT

This paper attempts to classify audio tracks into ‘day’ or ‘night’ songs based on the features of its musical instrumentation. Timbral, spectral, as well as perceptual features of the audio tracks are extracted and then fed to an SVM classifier which classifies ‘Night’ songs from the ‘Day’ songs. Expansion and future work of the paper will address classification of songs to a particular time of day, and improvement of feature selection and the classification method.

BACKGROUND

There are songs that are evocative of or better suited to a particular time of day, either achieved intentionally by the composers of the song, or the songs may contain latent features in its instrumentation and/or lyrics.

Some subjective differences noticed:

- ‘Morning’ and ‘Day’ songs tend to be more ‘acoustic’ in its instrumentation and the vocals seem to be cleaner. The songs seem perceptually sharper as well.
- In ‘Evening’ and ‘Night’ songs, where the instrumentation seems much more electronic and filled with echo and ‘reverb’, and the vocals seem to be dripped in effects.

DATASET

- GZTAN Dataset, consisting of 1000 song snippets
- Each song snippet is 30 seconds long, sample rate = 22050 Hz, 16-bit mono audio files in .wav format.
- The dataset was labelled according to its genres, with 10 genres consisting of 100 songs each.
- No dataset available consisting of these label, and the only place these labels were available were in curated playlists on music streaming services.
- The dataset was thus labelled personally, potentially leading to bias.
- Training/Testing split – 80/20 of the total dataset

PREPROCESSING

- Amplitude of each track was normalized
- The following representations of the audio track were computed for further feature extraction
  - STFT – For spectral features
  - Mel-Spectrogram – To extract the mel-frequency cepstral coefficients (MFCC).
  - Bark Spectrogram – To extract perceptual features.
- Prior to calculating the bark spectrogram, a filter simulating auditory response of the mid-ear was applied to the magnitude spectrogram.

FEATURES AND CLASSIFICATION

The following features were extracted from the following representation.

Representation	Features
STFT	Spectral flatness, centroid, flux, roll-off, kurtosis, crest
Mel-spectrogram	13 MFCCs
Bark Spectrogram	Sharpness, Spread

The features were selected using principal component analysis to reduce feature dimensionality.

Classification was done using SVM with a Gaussian Kernel.

RESULT

The confusion matrix for a test is given below:

	Predicted Day	Predicted Night
True Day	62.98%	37.02%
True Night	42.35%	57.65%

- The results aren’t too promising using conventional audio classification features.

CHALLENGES AND FUTURE WORK

- Despite attempts to objectify features that make a song ‘Day’ or ‘Night’, it is still a very subjective opinion.
- Earlier direction of project involved extracting more latent and objective features
- General features, used for other classification problems, such as genre classification and mood classification may not work very well and thus may not be sufficient when approaching this particular problem.
- Future work involve extracting features which are better representative.
- Extracting such features may help in algorithmic composition of music consisting of these features

REFERENCES

[1] G.Peeters. A large set of audio features for sound description (similarity and classification) in the CUIDADO project. Technical report, IRCAM, 2004.

[2] Emmanouil Theofanis Chourdakis and Joshua D. Reiss: “Automatic Control of a Digital Reverberation Effect using Hybrid Models*,” AES 60TH INTERNATIONAL CONFERENCE, Leuven, Belgium, 2016 February 3–5*

[3] T.Ganchev, N. Fakotakis, and G. Kokkinakis. “Comparative evaluation of various MFCC implementations on the speaker verification task.” *Proceedings of the SPECOM-2005, 2005*

[4] G. Tzanetakis P. Cook, “Automatic Musical Genre Classification of Audio Signals,” IEEE Transactions on Speech and Audio Processing, 2002