Viktor Pavlov 230335TAF

1 Music Genre Classification via MFCC Analysis

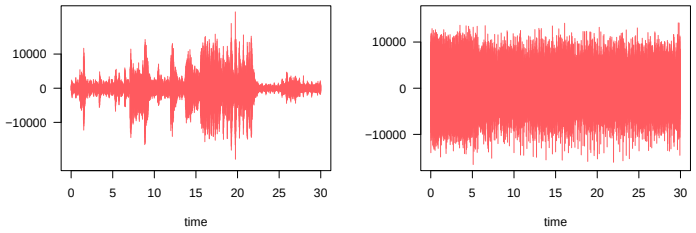

Figure 1: Audio Waveforms of the files `jazz.00000.wav` (left) and `metal.00001.wav` (right).

The goal of this project is to try music genre classification, utilizing audio feature extraction techniques together with clustering and classification data mining techniques. For this purpose, the GTZAN Genre Collection dataset [5] will be used. The dataset consists of 1000 audio tracks, each 30 seconds long. It contains 10 genres, each represented by 100 tracks. The tracks are all 22050 Hz monophonic 16-bit audio files in `.au` format. It's important to note, that this is a legacy dataset, collected from various sources, including personal CDs, radio and microphone recordings, with inconsistent quality [4]. Nevertheless, it's sufficient for our use case. Visualizations of audio signals from the dataset are displayed in Figure 1.

2 MFCC Audio Feature Extraction

Mel-frequency cepstrum coefficients (MFCCs) are a representation of the audio signal in the form of a unique spectrum. It's known to taking a complex audio signal and translating it into a simplified version that captures the essential components to the human ear. Applications of MFCCs include speech recognition [1] and music genre classification [2]. Various literature suggests that the first 12 coefficients are sufficient for speech recognition, while it's recommended to use $20 - 40$ coefficients for musical audio signals. The precise number of coefficients can be determined by means of experementation, and in context of this project, the first 30 coefficients were used. The `tuneR` package [3] was utilized for extracting the MFCCs. Visualization of MFCCs is displated in Figure 2.

3 K-Means Clustering

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi

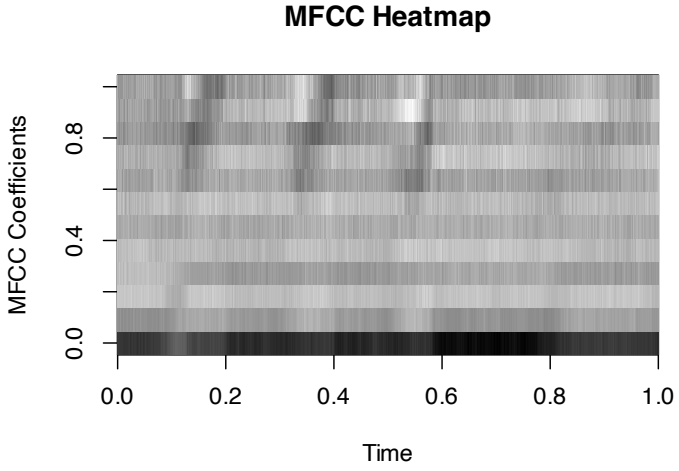

Figure 2: MFCC heatmap of the file `classical.00015.wav`.

sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

4 Classification using Support Vector Machines (SVM)

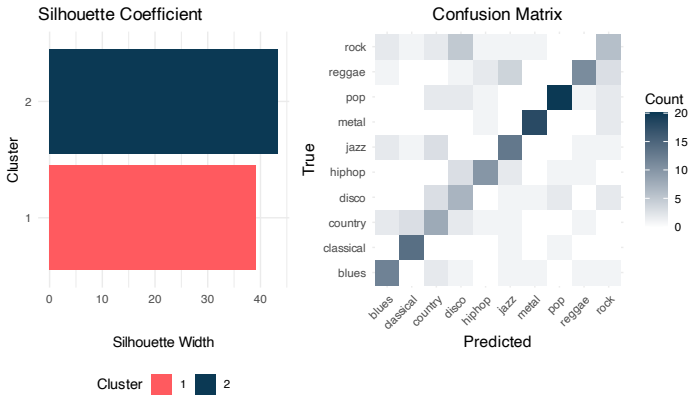

Figure 3: Correlation heat map of the Boston Housing dataset.

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

References

[1] Todor Ganchev, Nikos Fakotakis, and George Kokkinakis. Comparative evaluation of various mfcc implementations on the speaker verification task. In *Proceedings of the SPECOM*, volume 1, pages 191–194, 2005.

[2] Alexander Lerch. *An Introduction to Audio Content Analysis: Applications in Signal Processing and Music Informatics*. Wiley-IEEE Press, 2012.

[3] Uwe Ligges, Sebastian Krey, Olaf Mersmann, and Sarah Schnackenberg. *tuneR: Analysis of Music and Speech*, 2023.

[4] Bob L Sturm. The gtzan dataset: Its contents, its faults, their effects on evaluation, and its future use. *arXiv preprint arXiv:1306.1461*, 2013.

[5] George Tzanetakis, Georg Essl, and Perry Cook. Automatic musical genre classification of audio signals, 2001.