Viktor Pavlov 232211IAPM

1 Music Genre Classification via MFCC Analysis

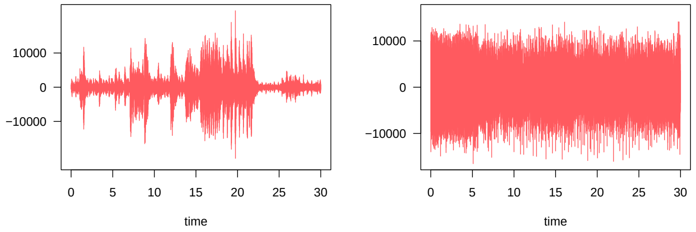

Figure 1: Audio Waveforms of the files `jazz.00000.wav` (left) and `metal.00001.wav` (right).

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 [6] 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 [5]. 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) convert complex audio signals into a compact spectral representation that highlights components critical to the human ear. Applications of MFCCs include speech recognition [1] and music genre classification [3]. In particular, MFCCs can encapsulate the timbral texture of audio, capturing subtle nuances, useful for distinguishing different musical styles. 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 experimentation, and in context of this project, the first 30 coefficients were used. The `tuneR` package [4] was utilized for extracting the MFCCs. Visualization of MFCCs is displayed in Figure 2.

3 K-Means Clustering

While clustering all 10 genres by MFCC features produced adequate results at first sight, only two genres were picked for simplification purposes. Results are displayed in Table 1. Additionally, silhouette coefficient plot of the Pop and Metal genres clustering results is displayed in Figure 3.

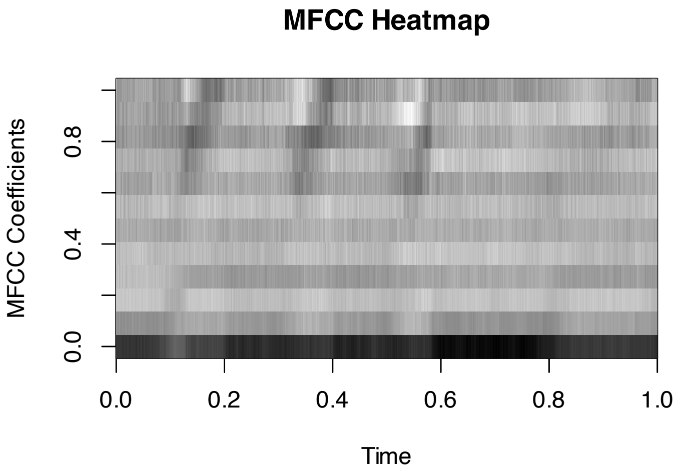

Figure 2: MFCC heatmap of the file `classical.00015.wav`.

Genre Pair	Cluster Sizes	Variance Ratio (%)
Pop vs Metal	106, 94	46.8
Jazz vs Rock	84, 116	34.3
Reggae vs Classical	95, 104	26.1
Country vs Disco	104, 96	40.9

Table 1: K-Means Clustering Results for Different Genre Pairs.

4 Classification using Support Vector Machines (SVM)

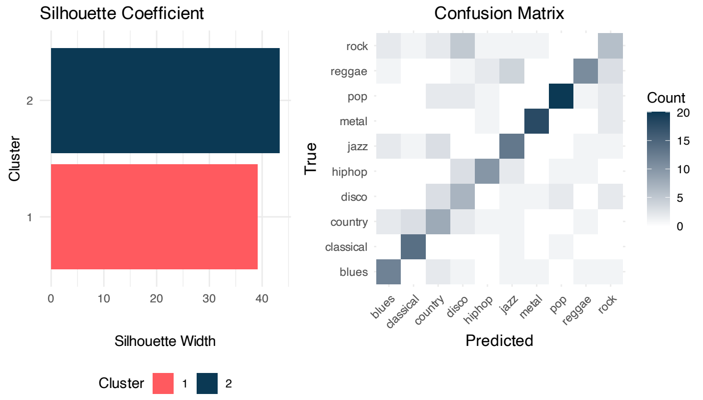

Figure 3: Pop vs Metal silhouette plot (left), Confusion matrix for all genres (right).

For the classification task SVM classifier was utilized with the *radial* kernel, yielding a 0.604 accuracy. This result represents the optimal outcome obtained after experimenting with various numbers of MFCC coefficients. The model can potentially be improved by adding features like the spectral centroid and zero crossing rate, that capture nuances related to dynamics and tonality of the given audio sample. The `seewave` package [2] can be utilized for extraction of these additional features.

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] J. Sueur, T. Aubin, and C. Simonis. Seewave: a free modular tool for sound analysis and synthesis. *Bioacoustics*, 18:213–226, 2008.

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

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

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

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