EXAMENSARBETE INOM TEKNIK,
GRUNDNIVÅ, 15 HP
STOCKHOLM, SVERIGE 2018

The Effect of Audio Snippet Start Positions Based on Audio Signal Features on Genre Classification Accuracy

NICKLAS HERSÉN

AXEL KENNEDAL

KTH
SKOLAN FÖR ELEKTROTEKNIK OCH DATAVETENSKAP

Contents

1 Introduction

When analyzing audio to perform genre classification it is standard to not analyze entire songs, but to choose small segments ("snippets") of them. This reduces the amount of data that needs to be processed, but may potentially lead to less accurate results. How to choose which part of a song that is to represent the entire song therefore becomes an important question. Interestingly enough this is a commonly overlooked aspect of genre classification; in our research of previous work we found that most often the researchers would simply choose the first X seconds of the songs (or use a library of snippets ready to go), without any solid argument for that choice.

1.1 Problem Statement

How audio snippet start positions based on audio signal features affect genre classification accuracy.

1.2 Scope

To limit our scope we will attempt to produce one algorithm for analyzing the spectrum of a song in order to find a good snippet start position.

1.3 Approach

We will perform a literature analysis and find relevant material about previous studies in the field of machine learning based music genre classification. Based on the results of the literature study identify and apply the current state of the art methods for music genre classification to a high quality dataset. Compare the performance and results of the classification using different starting points based on different features. Was there any difference between the results, if so why?

1.4 Thesis Outline

2 Background

2.1 Definitions

2.1.1 Genre

The Oxford Dictionary of English defines genre as "a style or category of art, music, or literature" [2]. Music genres don't even necessarily correlate to the sound of the music - they may be more related to the culture of the musicians creating it, and the time and place where it happened [8]. Kriss notes that there are some problems with determining genres: Since there are no exact confines between separate genres the perception of them are rather subjective. In recent times the rapid development of subgenres and hybrid genres continues to blur these lines even further. Tzanetankis, Essl and Cook state that even though the classification of music into genres is subjective there are perceptual criteria related to the texture, instruments and rythmic structure that can be used to characterize music [4].

Even as an experienced listener it may sometimes be difficult to determine the genre of a song one is listening to. Most people would probably agree that music taste is subjective, and so is the way that we perceive music. As a result of this there is no standard for classifying music into genres - it is done on a basis of what the listener knows about genres. Songs can be constructed of the same instruments and with the same emotion attached to it yet still be considered different genres. Although music is meant to be enjoyed by humans - perhaps machines are better at classifying it.

2.1.2 Feature

An individual measurable property or characteristic of a phenomenon being observed [1]. Selecting informative and independent features is crucial for developing effective algorithms in pattern recognition and machine learning. A set of numeric features can be described by a feature vector and is used as input data for classification algorithms.

2.2 Previous Work

From what our research has told us most work in the field of music genre classification thus far has been concerned with optimizing parameters for related algorithms and choosing the best algorithms and features and combinations of such algorithms, features and parameters. In order to be able to compare songs to each other there is a need for some sort of numerical description of sound content, which characterizes the sound. This is what is referred to as feature extraction [4].

Feature extraction is a central part during the construction of machine learning algorithms. Deciding on which features to extract depends on its computability and its relation to the specified problem. By categorizing audio features into different subcategories based on Weihs et al. [7] and Scaringella [6] work, the audio characteristic are assigned into certain categories based on different perspectives (short-, long term) and levels (low-, mid level). Low-level features can efficiently be extracted from the audio using different spectral analysis techniques such as fast Fourier transform (FFT) or discrete wavelet transform (DWT). These kind of features describe the characteristics of an audio signal and are usually called timbre (low-level short-term). Mid-level features attempt to describe patterns such as rhythm, pitch and harmony (low-level long-term) rather than the characteristics of the signal. Even though mid-level features are easily identified by music listeners, they are not straightforward to define. Low-level features such as timbre are currently the most widely used during music genre classification, however there are also some algorithms that rely on mid-level features (mostly rhythm) [3].

One of the largest challenges about comparing the similarity between songs is choosing which features to extract, and how well you are able to classify songs into genres depends very much upon this step. Songs that are considered similar must be described by features that have a relatively small distance between them in feature space, while songs that are considered to not be similar need to be described by features far apart. Another important aspect is that the chosen features should preserve relevant information from the original data (the music snippet). Often these features are chosen by trial and error [4].

After features have been extracted they are used as input to some sort of classification algorithm, for an example a machine learning algorithm or a neural network.

3 Method

Algorithms for music genre classifications must take the characteristics of a song into account when deciding which genre the song belongs to. Machine learning algorithms attempt to emulate this by comparing the distance between spectral features of the songs. In order to achieve maximum precision we base our feature extraction and classifier parameters on previously conducted research.

For feature extraction we have decided to use MFCC, where we use 20 coefficients. The frame size used to to derive each coefficient is 30 ms with 50% overlap between frames, which is recommended by Aucouturier and Pachet [5].

4 Results

5 Discussion

5.1 Result discussion

5.2 Possible improvements and future work

6 Conclusion

References

[1] C. M. Bishop. *Pattern recognition and machine learning.* Information science and statistics. Springer, New York, NY, 2006.

[2] O. L. D. English. Definition of genre. *Oxford Dictionary*, 2018.

[3] Z. Fu, G. Lu, K. M. Ting, and D. Zhang. A survey of audio-based music classification and annotation. *IEEE Transactions on Multimedia*, 13(2):303–319, April 2011.

[4] P. Kriss. Audio based genre classification of electronic music. Master's thesis, University of Jyväskylä, June 2007.

[5] F. Pachet and J.-J. Aucouturier. Improving timbre similarity: How high is the sky. *Journal of negative results in speech and audio sciences*, 1(1):1–13, 2004.

[6] N. Scaringella, G. Zoia, and D. Mlynek. Automatic genre classification of music content: a survey. *IEEE Signal Processing Magazine*, 23(2):133–141, March 2006.

[7] C. Weihs, U. Ligges, F. Mörchen, and D. Müllensiefen. Classification in music research. *Advances in Data Analysis and Classification*, 1(3):255–291, Dec 2007.

[8] J. Wong. Visualising music: the problems with genre classification. *Masters of Media*, 2011.