Musical Genre Classification

MUS-15

Jan Müllers

July 10, 2015

Introduction

Musical Genres are categorical labels created by humans to characterize pieces of music. They are caracterized by instrumentation, rhythmic structure and harmonic content and are used to structure large collections of music. Doing this manually is slow and expensive. Therefore automatic genre classification systems are developed which can assist or replace the human user. Those systems generally have some limitations. They only use a small size of different genres (10 or less) and focus on the hard decision problem where only one label is given to each song. Those limitations may make those systems not usable in praxis, where you could have up to 500 different genres and songs belonging to differnt genres at the same time, expecialy when subgenres are taken into account. The Problem is generally solved by exploiting different low-level features from the audio signal and using different classification methods to classify those features. Those features can be melody, rhythm, pitch or Mel Frequency Cepstral Coefficients as used in automatic speech regocnition. One Paper presented in detail in this report uses Locality Preserving Non-Negative Tensor Factorization as a way to transform audio signals into features vectors. It produces an accuracy of around 95% on two different testing sets. This is the highest accuracy I could find in all papers.

History

Automatic Genre Classification first became a point of interest when large digital music libarys got commen around 15 years ago when mp3s, mp3-players, napster and similiar services became available. The first approach was presented by Tzanetakis and Cook in 2002 in a paper called 'Musical Genre Classification of Audio Signals' which focused on using methods from automatic speech recognation for automatic genre classification. This could be seen as a promising strategy due to the similarity of the problems and the research already done for automatic speech recognition since the 1970s. Both problems want to apply a class to an audio signal (word/sentence ¡-¿ genre) and try to do so by extracting a feature vector from the audio signal. But those problems are also different since you have to take context (surounding words) into account for automatic speech recognition. This paper is still refered to today and used for comparision in new papers. Papers after that tried to use different features and classification methods to increase accuracy but used the same method of transforming the song into a feature vector and using standart classification methods to classify this vector.

Locality Preserving Non-Negative Tensor Factorization

Locality Preserving Non-Negative Tensor Factorization (LPNTF) is a method introduced by Panagakis, Kotropoulos and Arce in 2009. With this method errorrates around 95% were archived on two different testing sets. the idea of the method is to first use the LPNFT and afterwards a Sparce Representation-Based Classification. First we take a look at what LPNFT means. A tensor is the multidimensional equivalent of matrices and vectors. Non-negative means all tensors have no negative elements. Factorization is dividing a tensor in several vectors which when linear combined give the tensor. To be locality preserving the method takes the nearest neighbour graph into account. So each song is represented as a tensor and for the training the tensor is factorizied. The vectors created through the fatorization are then used for the classification of other songs. Sparce Representation-Based Classification is a classification method first introduced for automatic face regocnition. The idea is to have a dictionary created in training. In this case consisting of vectors calculated through the LPNTF from the training data. Each new songs is then again represented as a tensor and a linear combination for that tensor consisting of vectors from the dictionary that all belong to the same genre is searched for. The genre which can linear combine the new song with the fewest Vectors is the label given to the song.

Conclusion

The general idea for automatic genre classification is to transform the audio-signal into a feature vector and afterwards classifying the feature vector. With the approach shown in the last paragraph accuracys of 95% can be achieved. However this is still no replacement for human experts due to the small number (10) of different genres used in the test. But the system could be used to assist human experts. An easy way in practical applications to improve the accuracy for a larger number of genres would be to take the meta data of the song, especially the artist into account since most artist always make songs from the same genre. This however would not be part of the problem researched which wants to give genre labels only from the audio signal.

References

[1] George Tzanetakis, Perry Cook. *Musical Genre Classification of Audio Signals*. IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING, VOL. 10, NO. 5, JULY 2002.

[2] Yannis Panagakis, Constantine Kotropoulos, Gonzalo R. Arce. *MUSIC GENRE CLASSIFICATION USING LOCALITY PRESERVING NON-NEGATIVE TENSOR FACTORIZATION AND SPARSE REPRESENTATIONS*. 10th International Society for Music Information Retrieval Conference (ISMIR 2009).