Musical Genre Classification

Jan Müllers

RWTH Aachen, 2015

Musical Genres

- categorical labels created by humans to characterize pieces of music
- caracterized by the instrumentation, rhythmic structure, and harmonic content
- used to structure large collections of music

Automatic Genre Classification

- manual genre classification is slow and expensive
- automatic genre classification can assist or replace the human user

Limitations

- only small size of different genres (10 or less)
- only hard decision problem (only one label for a song)
- maybe not close to praxis (over 500 different genres, songs can belong to different gernres, subgernres)

History

- first ideas: 15 years ago when large digital music libarys got common
- first approach: using ideas from automatic speech recognition [1]
- each paper after that used new features and different classification methods to increase the accuracy

Automatic Speech Regocnition

- already researched since 1970s
- similar problem: also applying a class to an audio signal (word/sentence <-> genre)
- both need a way to represent the audio signal as an feature-vector
- no context needed for genre classification

2 Steps

- The Problem can be divided into 2 differnet parts:
  - transforming the audio-signal into a feature vector
  - classification of the vector
- for step 2 one can use standart classification methods, research: which is the best?
- for step 1 new features can be developed and tested

Features that can be used

Different papers introduced features that can be used:

- Mel Frequency Cepstral Coefficients (as used in Speech Regocnition)
- melody
- rhythm
- pitch

Locality Preserving Non-Negative Tensor Factorization

- a state of the art approach with low error-rates
- introduced by Yannis Panagakis, Constantine Kotropoulos and Gonzalo R. Arce in 2009 [2]

The Idea

- the first part is the *LPNTF*
- the second part is *Sparse Representation-Based Classification*

How LPNTF works

- tensor: multidimensional equivalent of matrices and vectors
- non-negative: all tensors have no negative elements
- factorization: a tensor is divided in several vectors, which linear combined give the tensor
- locality preserving: take the nearest neighbor graph into acount

Sparse Representation-Based Classification

- a classification method first introduced for automatic face regocnition
- idea:
  - we have an dictionary created in training
  - we presented a song as a linear combination of atoms from the dictionary which all belong to one genre

Results

- Errorrates are around *95%* with this approach
- not all papers are comparable due to different test settings (number of genres, different databases)

Conclusion

- Two steps:
  - transforming audio-signal in feature vector
  - classifing feature vector
- still low number of different genres
- No replacement for human experts yet, can only assist

References I

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

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).