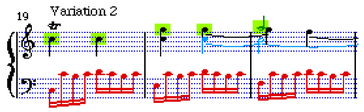
From Multi-labeling to Multi-domain-labeling: A Novel Two-dimensional Approach to Music Genre Classification



Hanna Lukashevich, Jakob Abeßer, Christian Dittmar, and Holger Großmann

Fraunhofer Institute for Digital Media Technology, Ilmenau, Germany

Main Idea

Problem:
- Often multiple genre labels are assigned to music tracks due to multiple genre influences
- Problematic ground-truth for training classification models

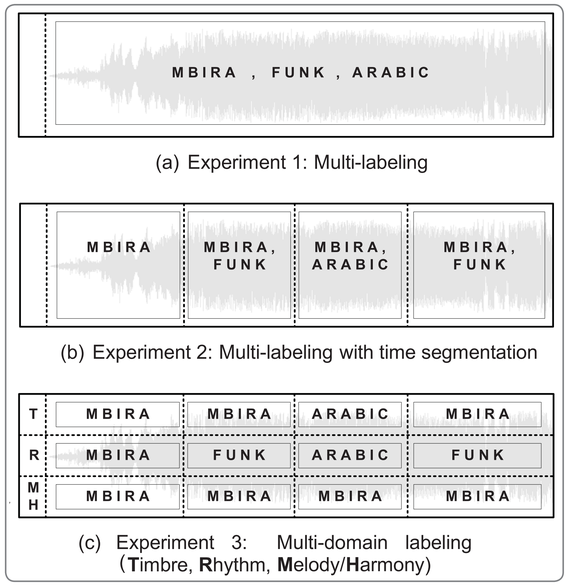
Proposed solution:
- Apply temporal segmentation
- Use musical domains ("dimensions"):
  - Timbre
  - Rhythm
  - Melody/Harmony
- Assign genre labels to the most appropriate domain in each time segment
- Break down the multi-label into multiple single-label classification tasks

Gain about 10% of F-measure!

Features & Classifiers

Features:
- Besides low-level acoustic features, several mid-level representations are extracted
- Features are assigned to the musical domains
- Timbre (233 dimensions): Mel-Frequency Cepstral Coefficients, Audio Spectrum Centroid, Spectral Crest Factor, Spectral Flatness Measurement, Modulation Spectral Features
- Rhythm (768 dimensions): Audio Spectrum Envelope, Percussiveness, Envelope Cross-Correlation, derived from Auto Correlation Function, log-lag Auto Correlation Function, etc.
- Melody/Harmony (187 dimensions): Circular Pitch Space, features derived form Enhanced Pitch Class Profiles

Classifiers:
- Principal Component Analysis (PCA)
  - Reduction to 100 dimensions for each domain
- Gaussian Mixture Models (GMM)
  - Diagonal covariance matrix; 1, 5, 20 or 50 mixtures
- Post-processing via simple thresholding



Experimental Framework

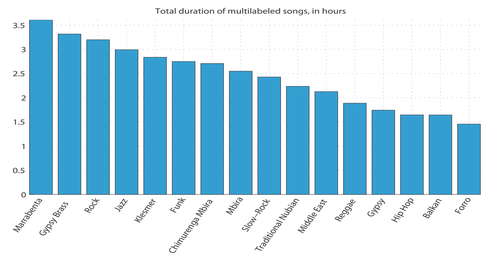
Experiment 1: Multi-labeling (ML)
- Multiple labels (may) be assigned to each song
- No time segmentation available
- Set of binary GMMs, one model per genre

Experiment 2: ML + time segmentation (TS)
- Provide time segmentation
  - Set of labels is not changed within each time segment
- Multiple labels (may) be assigned to each segment
- Set of binary GMMs, one model per genre

Experiment 3: Multi-domain-labeling + TS
- Provide time segmentation
- Assign each label to one of musical domains
  - "This segment belongs to *Samba* because of *rhythmics*"
- Up to 4 genre labels for each time segment
  - Global, timbre-related, rhythm-related, tonality-related
- One "discriminative" GMM for each domain

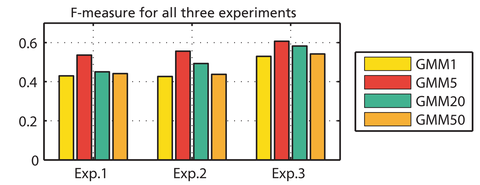
Database

- Music collection of 430 full-length tracks
- Flat taxonomy of 16 world music genres
  - Balkan, Gypsy, Forro, Klesmer, Mbira, Nubian, etc.
- Manual annotations by educated musicologists
- Follow the multi-domain-labeling principle
- Label cardinality of 1.1 to 2.0
- Split in training (70%) and test set (30%)



Results

- F-measure values for each number of mixtures in GMM are increased for Exp. 3 in comparison to Exp. 1 and Exp. 2.
- Optimal number of mixtures for all three experiments: 5 mixtures
- The best performance in Exp. 3 for GMM with 5 mixtures: F-measure of 0.61
- In Exp. 3 the classifier includes about two times fewer free parameters than in Exp. 1 and Exp. 2.



Conclusions & Future Work

Conclusions:
- Novel two-dimensional approach to music genre classification
- Decompose a multi-label classification problem by breaking it down in two dimensions
- Gain about 10% of F-measure!
- Extensible to other music genres
  - Partitioning into musical domains is universal

Future Work:
- Apply alternative classification techniques such as Support Vector Machines
  - Suppose to perform better for binary classification
- Make use of supervised feature selection and feature space transformation methods within each domain
- Add additional domains:
  - Vocals and instrumentation

Acknowledgments

This work is part of the German research project *GlobalMusic2One* (www.globalmusic2one.net) funded by the German Ministry of Education and Research. Special word of praise goes to the music experts of *Piranha Music & IT* for providing music content and manual annotations.

Selected References

[1] C. Dittmar, C. Bastuck, and M. Gruhne. Novel midlevel audio features for music similarity. In Proc. of the Intl. Conf. on Music Communication Science (ICOMCS), 2007.
[2] K. Trohidis, G. Tsoumakas, G. Kalliris, and I. Vlahavas. Multi-label classification of music into emotions. In Proc. of Intl. Conf. on Music Information Retrieval (ISMIR), 2008.
[3] G. Tzanetakis and P. Cook. Musical genre classification of audio signals. IEEE Transactions on Speech and Audio Processing, 10(5):293–302, 2002.

Fraunhofer IDMT

Bundesministerium für Bildung und Forschung