ADAPTING METRICS FOR MUSIC SIMILARITY USING COMPARATIVE RATINGS

Daniel Wolff and Tillman Weyde
Music Informatics Research Group
Northampton Square, EC1V 0HB London
daniel.wolff.1@soi.city.ac.uk
t.e.weyde@soi.city.ac.uk

CITY UNIVERSITY LONDON

Motivation

This poster presents a machine learning approach for analysing user data that specifies song similarity. **Understanding how we relate and compare music** has been a topic of great interest in musicology as well as for business applications, such as music recommender systems. The way music is compared seems to vary between different cultures. Adapting a generic model to user ratings is useful for personalisation and can help to better understand such differences.

In our experiments we find that a significant amount of information can be gained from comparative similarity ratings, allowing for an improved similarity estimation on seen and unseen data.

Audio and Similarity Dataset: MagnaTagATune [E. Law et al. 2009]

Online Song excerpts from the **Magnatune** label
- 30 seconds long, can be divided into 4 broad categories: "electronica" (30%), "classical" (28%), "world" (15%) and "rock" (17%)
- Annotation data (user tags) and **similarity ratings** from the human computation game **„TagATune"**

Features

The clips in our database are described using a combination of content-based and genre features:

Chroma and **timbre** features precomputed by "TheEchoNest"
- Postprocessing:
  - K-means: **4 clusters per clip** and feature type,
  - 12-dim. chroma features are transposed to root note C
  - 12-dim. timbre features are clipped
  - Both normalised to a maximum value of 1

2-3 genres per clip are annotated in the **Magnatune catalogue**
- Each clip is assigned a 44-dim. binary genre vector

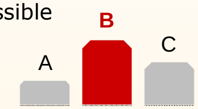
Chroma and timbre centroid information and genre features are combined into one 148-dim. vector per clip

Similarity Data

- TagATune gamers have to **agree** on the "outlier" clip out of 3

- Data for 533 clip triplets
  - Avg. 14 votes per triplet
  - 1019 clips included

Postprocessing:
- Consider the triplet histograms as voting
  - Determine winning **outlier (B)** where possible
  - Discard votings featuring no clear winner



- Derive relative clip similarity constraints:
  - (A, **B**, C), **B** being the outlier implies
  - **sim(A, C) > sim(A, B)** AND **sim(A, C) > sim(B, C)**
- Derive binary rankings
  - Alternative representation of constraints
  - Inconsistent constraints are removed (where clips are similar and dissimilar at the same time)

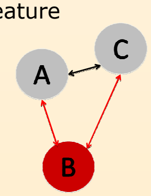
Similarity Model and Adaptation

- **Mahalanobis metric** for measuring clip similarity:

$$d_W(x,y)=\sqrt{(x-y)^{\mathrm{T}}W(x-y)}$$

Matrix W defines the similarity measure, clip feature vectors x and y

Generalised Euclidean metric
- allows for geometric interpretation
- psychological validity has been questioned

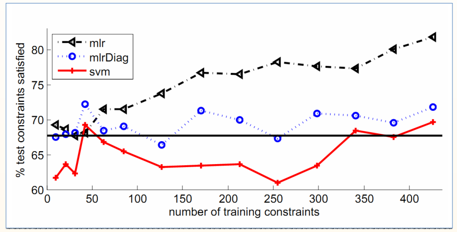


- We compare **two different algorithms** for optimising d_W
  1. **MLR**: [McFee and Lanckriet 2010] optimise a full W to binary rankings
  1.1. mlrDiag: MLR variant restrained to a diagonal matrix W
  2. **SVM**: [Schultz and Joachims 2003] optimise a weighted Euclidean metric using a diagonal matrix W

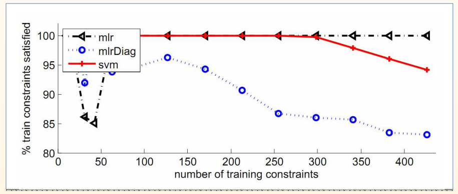
Experiments

- 5-fold cross validation with test-sets of ~106 binary rankings, evaluate fulfilled rankings

Test Set:
- MLR: **82%**
- mlrDiag: 71%
- SVM: 70%
- Eucl.: 67%
- ($W_{ij} = \delta_{ij}$)



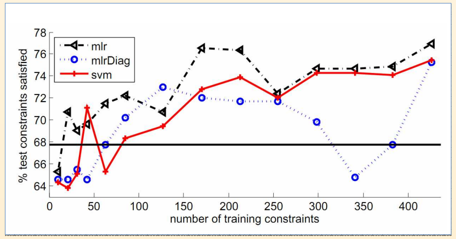
Training Set:
- MLR: Best, but bad for <50 constr.
- mlrDiag: weak adaptation
- SVM**:** good on training data, bad generalisation



Feature dimension / PCA feature test

- Features reduced to 20 –dim using Principal Component Analysis (PCA)

- MLR: **77%**
- mlrDiag: 76%
- SVM: 76%



Conclusion

- Similarity constraints contain generalisable information, which can be trained using the tested methods.
- MLR works well on both feature types tested
- mlrDiag tradeoff for regularisation and constraints has to be investigated
- Faster SVM works comparably well for low-dimensional feature space

For references and details, please ask or see our paper in the proceedings.