Adapting Similarity on the MagnaTagATune Database:

Effects of Model and Feature Choices

Daniel Wolff, Tillman Weyde
City University London
Department of Computing
daniel.wolff.1@soi.city.ac.uk, t.e.weyde@city.ac.uk

CITY UNIVERSITY
LONDON

Introduction

CITY UNIVERSITY LONDON

- **Music similarity measures:**
  - Central in MIR: recommendation, analysis, indexing, ...
  - Important in musicology: repetition / variation, citations, categorisation into style / genre

- **Goal: Learn human similarity judgements** from a human computation game.

- Compare two modelling approaches on the same similarity data.
  - **Facet-based similarity measures:**
    Stober and Nürnberger 2011 (ST11)
  - **Mahalanobis Metrics:**
    Wolff and Weyde 2011 (W11)
- Evaluate applicability of different algorithms and feature types

Structure

CITY UNIVERSITY LONDON

- Introduction
- Data
  - The MagnaTagATune Dataset
  - Similarity data
- Similarity Models: ST11 / W11
- Features: ST11 / W11
- Experiments
  - Results
- Conclusion

CITY UNIVERSITY LONDON

Data

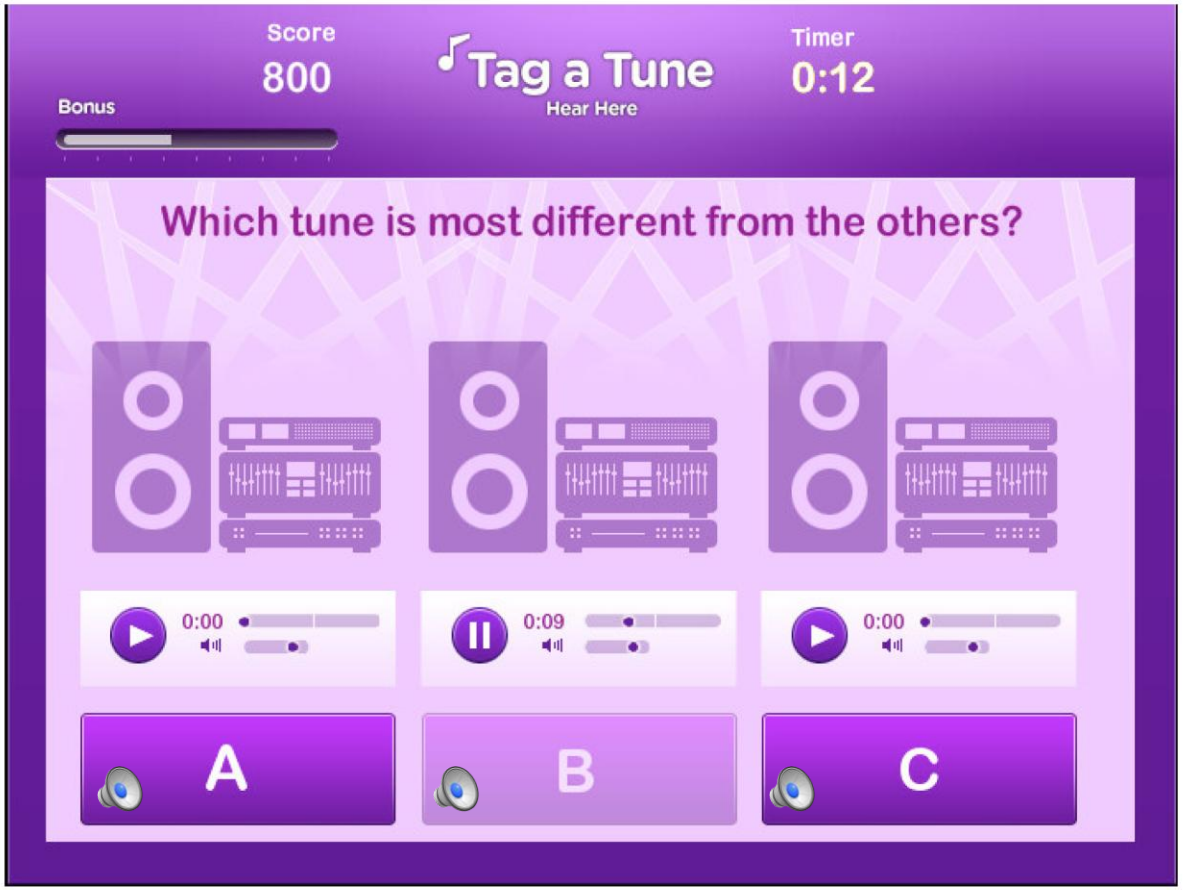
Dataset: MagnaTagATune

CITY UNIVERSITY LONDON

- Subset of 1019 Song excerpts from the Magnatune label
  - about 30 seconds long, most prominent genres:
    - ”**electronica**”, ”**classical**”, ”**world**” and ”**rock**”
  - **Similarity judgements** from the human computation game „TagATune“

  - **Tag features** from „TagATune“
  - **Audio features:**
    - Precomputed by the echonest

Similarity data



Law et al. 2009

Similarity data

CITY UNIVERSITY LONDON

- Data collected via bonus round in TagATune game
  - Users aim to agree on **outlying (most dissimilar) clip** out of three
  - **533 triplet histograms**, 1019 clips
  - On average **14 votes** per histogram
  - Some triplets reappear as permutation
    - (186 appear twice)
  - Most triplets contain 2 or 3 genres

Similarity Constraints

CITY UNIVERSITY LONDON

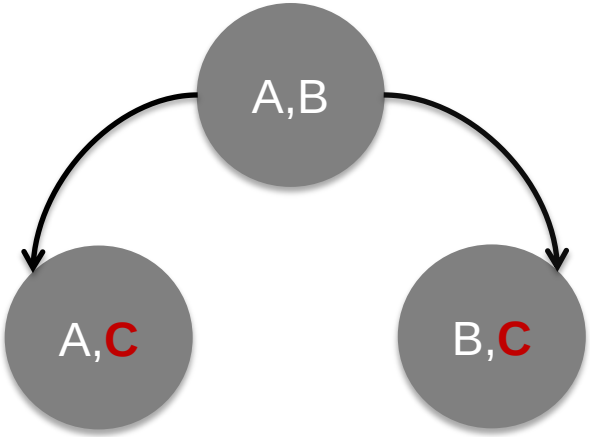
- Model similarity through distance measure
  - **d** is prospective distance measure
  - low distance ⇔ high similarity

- For each **outlier** vote **C,** given a triplet (A, B, **C**):
  - Derive similarity constraints
    - (A, B, **C**), **C** being the outlier implies
    - **d**(A, B) < **d**(A, **C**) AND **d**(A, B) < **d**(B, **C**)

Similarity Graph (Stober11)

CITY UNIVERSITY LONDON

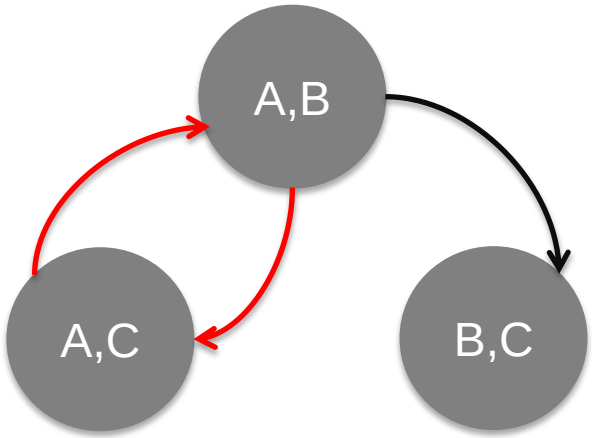
- Build a similarity multigraph (McFee et al. 2009)
  - Vertices: pairs of clips { (A, B), (A, **C**) …}
  - Directed edges: similarity constraints
    - (A, B) => (A, **C**) ⇔ d(A, B) < d(A, **C**)
    - (A, B) => (B, **C**) ⇔ d(A, B) < d(B, **C**)



Filter Similarity Data

CITY UNIVERSITY LONDON

- Graph is filtered to remove any inconsistencies
  - Remove cycles of length 2
    - Balance contradictory edges
    - Equal connections disappear



  - Further filtering:
    - Designed to remove cycles of greater length
    - Randomised process returns acyclic subgraph
      - **674 unique constraints remain**
    - Actually removes more edges than necessary
      - Future work (ISMIR2012)

CITY UNIVERSITY LONDON

Similarity Models

Similarity Models

CITY UNIVERSITY LONDON

- Goal: Learn / Model similarity votes
  - Find distance measure satisfying all constraints
  - Predict similarity votes on unknown data

- 2 approaches applied on MagnaTagATune:
  - W11: Mahalanobis Metrics
    - Metric Learning to Rank
  - ST11: Facet-based Distance
    - Quadratic optimisation
    - Linear SVM
    - Others

W11: Mahalanobis Metrics

CITY UNIVERSITY LONDON

- Generalised weighted Euclidean metrics,
  - Weight matrix W allows for transformations of the comparison space:
    - Rotations
    - Translations
    - Dilations

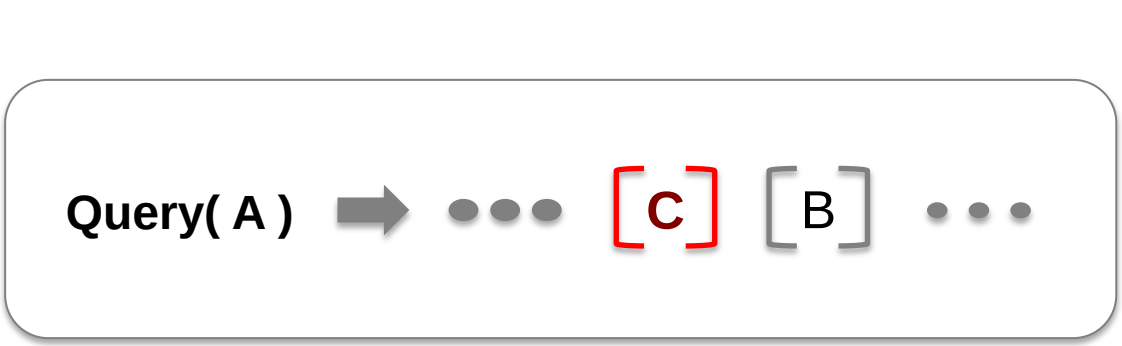
$$d_W(x,y) = \sqrt{(x-y)^T W (x-y)}$$

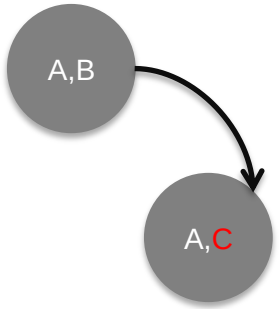
    - with feature vectors x,y $\in R^N$
    - pos. semidefinite W $\in R^{N\times N}$ defines the metric
    - W can be **restricted to diagonal shape**

W11: Metric Learning to Rank

CITY UNIVERSITY LONDON

- McFee and Lanckriet (2010): Metric Learning to Rank.
- Metric learning formulated as constrained regularisation
  - Structural SVM framework is used,
  - Optimisies Malalanobis distance measure
    - Constraints are defined by training rankings
    - A soft-margin approach allows some constraints to be violated in the final solution





ST11: Facet-based Distance

CITY UNIVERSITY LONDON

- Weighting of **predefined facet distances**

- Instead of directly weighting in feature space:
  - Assign specialised distance measure δ_{f_i} to each feature
  - Positive weights w_i determine a linear combination of individual distance results

$$d_w(A,B) = \sum_{i=1}^{l} w_i \delta_{f_i}(x_A, x_B)$$

(for clips A, B, features x_A, x_B)

ST11: Learning Facet Weights

CITY UNIVERSITY LONDON

- Various approaches have been compared
  (Stober and Nürnberger 2011)

- Compare the most successful ones:
  - **LIBLINEAR**
    - Learns a SVM which distinguishes between constraints
      - (A, B) => (A, **C**) vs (A, **C**) => (A, **B**)
    - Produces some slightly nonnegative weights w_i
    - Popular toolbox can be downloaded online
  - **Quadratic programming with slack**
    - Their own solver with quadratic optimisation of squared slack values, returns non-negative w_i

CITY UNIVERSITY LONDON

Features

Acoustic Feature Data

CITY UNIVERSITY LONDON

- Extractor: The Echo Nest „Analyse“ API

- Chroma & timbre features (segment-level, St11+W11)
- Aggregated to clip level:
  - ST11**: Single mean and variance** vectors per feature & clip
  - W11: **4 weighted cluster centroids** per feature & clip

- Clip-level information (ST11, relevant only)
  - key, mode,
  - loudness, energy,
  - time signature, tempo, “danceability”

ST11: Tag Feature Data

CITY UNIVERSITY LONDON

- STOB11: TagATune tag annotations
  - 188 unique tags provided in the dataset.
  - distributed rather sparsely, combine several tags:
    - singular/plural forms,
    - spelling correction and
    - semantic similarity.

- Result: Vocabulary of **99 tags,**
  - represented by binary values per clip

W11: Genre Features

CITY UNIVERSITY LONDON

- Genre information from the Magnatune label
  - Online catalogue annotates all Magnatune songs!
    - Small vocabulary: 44 genres for the whole set

- **Binary vector** $\in \{0,1\}^{44}$ per clip (1 dimension per genre)

Facets, Features, Parameters

CITY UNIVERSITY LONDON

Features	Facets Stober 11	Param. Wolff 11 MLR	Param. Wolff11 DMLR
chroma	2	4 · 12	4 · 12 · 148
timbre	2	4 · 12	4 · 12 · 148
clip-level audio	7	/	/
tags	99	44	44 · 148
	110	148	21904

CITY UNIVERSITY LONDON

Experiments

Experimental Setup

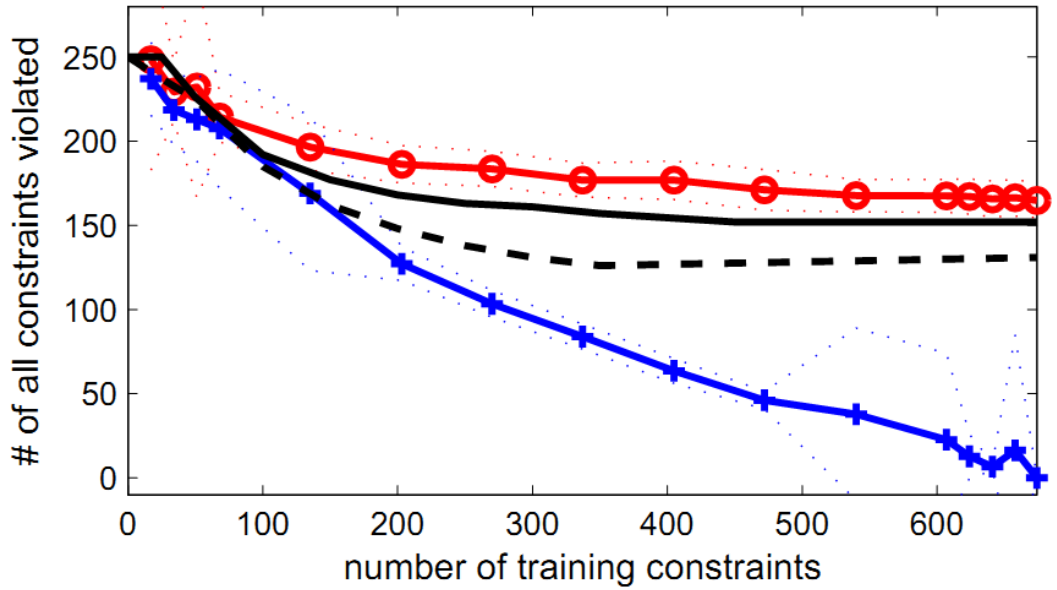
CITY UNIVERSITY LONDON

- Generate 10 randomly extracted *all constraints* sets
  - using the methods from ST11
- Use different numbers of training constraints:
  - For each size, 5 training subsets are selected randomly (for each of 10 *all constraints* sets)
- Evaluate W11 training success on *all constraints* sets,
  - including the training data
- Results are compared to the numbers in ST11

Algorithms training performance

CITY UNIVERSITY LONDON

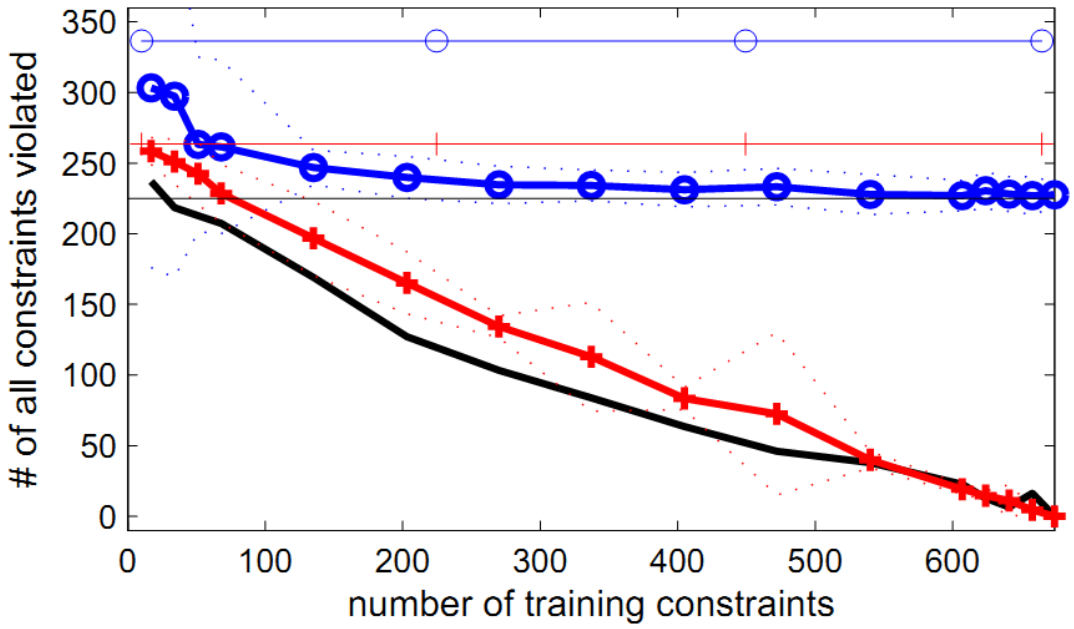
- **MLR** achieves **100%** top performance (no training constr. violated)
  - variance shows dependency on sampling
- **Quadratic** programming slightly better than **DMLR**
- **LIBLINEAR** (130 violated) achieves best facet-based result
  - but includes negative facet distance weightings



Results: W11 Feature Types

CITY UNIVERSITY LONDON

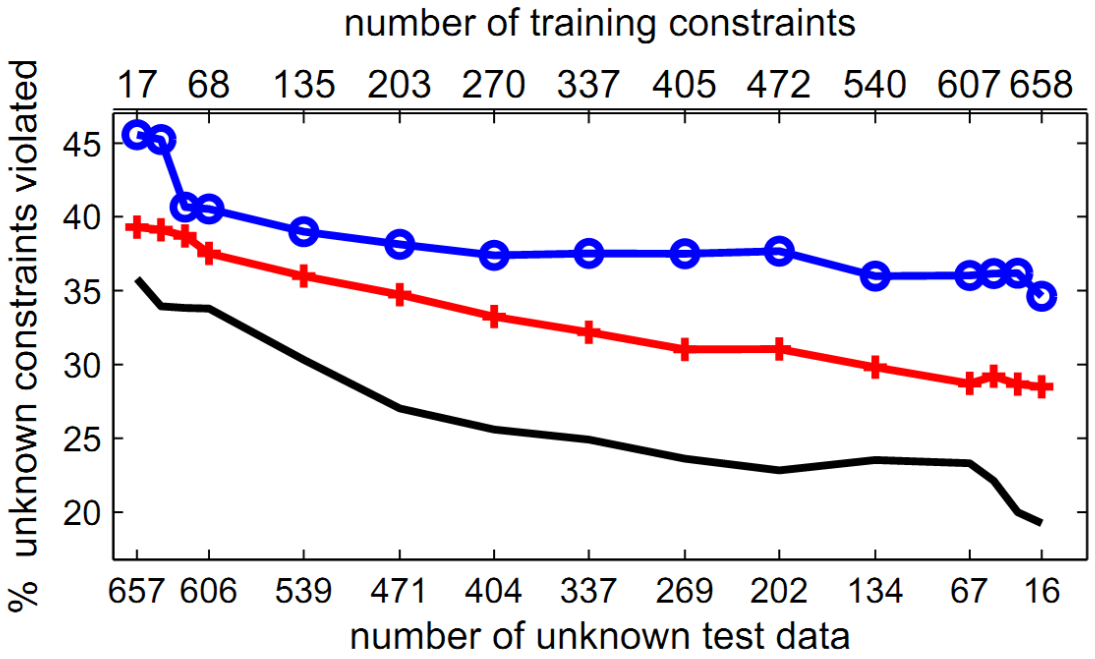
- **Baselines:** unweighted Euclidean distances for feature types
- **Combined features**: Best results (fast and complete learning)
- **Genre features:** features fail at learning, worst baseline
- **Acoustic features:** slower learning but can learn all constraints, **better performing baseline**



Results: W11 Generalisation

CITY UNIVERSITY LONDON

- **Combined features**: Best results **(20% violated)**
- **Acoustic features:** continuous improvement, but lower in general
- **Genre features:** some early learning, then no impact
  - information still valuable in combined features



Conclusions

CITY UNIVERSITY LONDON

- **Constraints from similarity votings contain generalisable information**, which can be modelled using the tested methods.
  - MLR with **full W** matrix learns all constraints
  - **Facet-based** approaches outperform diagonal MLR
- **Combined features** outperform single-source features
  - **Effectiveness of features is not necessarily reflected** in unweighted Euclidean distance
  - **Feature type** strongly affects performance (training and generalisation)
  - Genre features too sparsely located in vector space

Future Work

CITY UNIVERSITY LONDON

- Submitted for ISMIR 2012: Systematic comparison of algorithms with common features and extended similarity data
- Currently testing
  - training with more elaborate features
- Coming soon
  - Gather similarity data with more context information
  - Comparison of user groups

Fin

CITY UNIVERSITY LONDON

Thank you