ISMIR 2015 MÁLAGA

16th International Society for Music Information Retrieval Conference - October 26-30, 2015 Málaga, Spain

Comparative Music Similarity Modelling using Transfer Learning across User Groups

Daniel Wolff, Andrew MacFarlane and Tillman Weyde, Music Informatics Research Group, daniel.wolff.1@city.ac.uk

CITY UNIVERSITY LONDON

Music Similarity and User Data

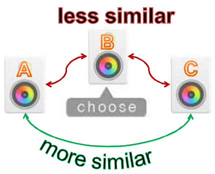
We present first results of experiments using music similarity ratings from human participants for group-specific similarity prediction. Music similarity is a key topic of research in music psychology and ethnomusicology. Perceived similarity is **specific to the individual user** and influenced by a number of factors such as **cultural background**, age and education.

Our goal is to adapt similarity models to similarity data of users sharing common attributes such as age. To this end we use age data reported from participants of a music similarity game. We finally compare the role of musical features in the specific and general models.

The CASimIR Dataset

Our Spot the Odd Song Out game **collects relative similarity judgments** of users on triplets of songs, where they are asked to **choose one song as the "odd song out"**.

- If song B is chosen, then:
  sim (A,C) > sim (B,C) AND
  sim (A,C) > sim (B,A)



Data is annotated with anonymised **user attributes** including: Age group, gender, occupation (sector), spoken languages, current location (city), birth location (city)

Music Database

- More than 11000 clips in total, each min. 30 seconds long
- Datasets:
  - **1.Million Song Dataset Subset**
    - Mostly Pop/Rock music, streamed by 7digital
  - **2.MagnaTagATune**
    - Only "classic" genre subset used.

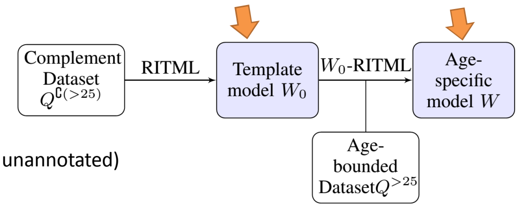
The Game Interface

- HTML5 web application
- **Odd One Out** scenario
- Rewards **agreement**
- 45 seconds **time limit**
- **Multi-player**

mirg.city.ac.uk/casimir/game/

Transfer Learning

- We model music similarity using a **generalisation of the Euclidean distance**: The M**ahalanobis distance**. This allows for training **a weighting and combination** of **music features** that correspond to the collected similarity data.
- **Challenge**: Age-specific **datasets are small** – direct training becomes difficult.
- **Solution**: Use **transfer learning** – models are **initialised on a large dataset**, then **fine-tuned** to age-specific data.
- Idea: Train general model, then adapt to specific group.

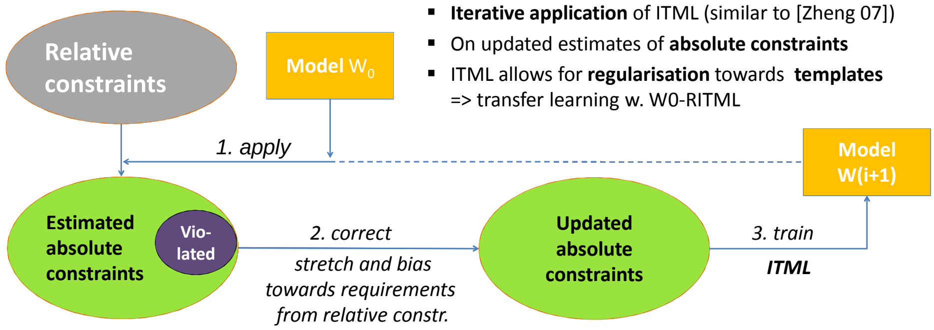


Process

1. Use data from all remaining ages (including unannotated) for training a general model "W_0".
2. Adapt model to specific age group data with W_0-RITML method.

Model Training: R(elative-input)ITML

- Training of model achieved with new RITML algorithm, adapted from ITML [Davis 2007].
  - **Iterative application** of ITML (similar to [Zheng 07])
  - On updated estimates of **absolute constraints**
  - ITML allows for **regularisation** towards **templates** => transfer learning w. W0-RITML



- **Preliminary evaluation** of RITML on full datasets:
- Test results with 10-fold cross-validation
- Results equal or slightly better than state-of-the-art

	MagnaTagATune	CASimIR
Euclidean	59.80%	59.75%
RITML	71.12%	**64.23%**
SVM	**71.20%**	63.22%
MLR	68.90%	62.79%

Modelling Age-Based Subsets

- Split data into **user groups** by **reported** age of participants
- Frequently reported attribute
- Only slight bias towards users > 25
- Experiments with smaller categories showed strong variation

	R	$R^{\leq 25}$	$R^{>25}$	$R^{\complement(\leq 25)}$	$R^{\complement(>25)}$
ratings	2102	919	644	1183	1458
constr.	914	723	576	732	809
clips	180	171	163	175	176

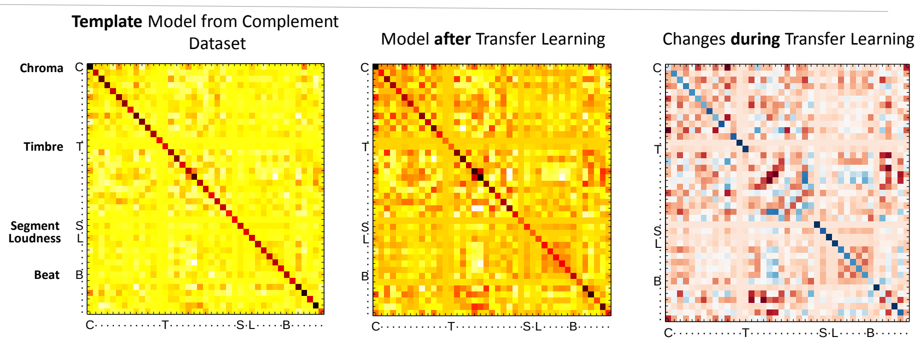
Performance for Age-Specific Data

Generalisation performance in percentage of similarity ratings fulfilled by the models on the age-specific datasets (10-old cross-validation). We compare: No training (Euclidean), Support Vector Machine (SVM), Metric learning to Rank (MLR), direct training (RITML), transfer learning (W0-RITML) and training with all countries' data (JOINT)

	Q <= 25	Q >=25	Average
Euclidean	59.32%	59.15%	59.23%
SVM	61.56%	61.34%	61.45%
MLR	62.06%	62.58%	62.32%
RITML	63.69%	61.02%	62.35%
W_0-RITML	**65.53%**	**67.07%**	**66.30%**

- Training on individual datasets achieves little improvement over Euclidean baseline.
- Simple RITML varies in performance.
- W0-RITML / transfer learning achieves best results.

Analysis of Specific Models



- Analyse the difference of fine-tuned model to template model
- Specific model has stronger off-diagonal values
  - Raised correlation of **timbre** and **beat**, **chroma**, **tatum confidence** (C11C1; T6T5; B4T4 and B4T5)

⇒ RITML and transfer learning improve training on smaller datasets.
⇒ Analysis shows features' correlation with similarity data.
⇒ Can be extended to ethnomusicology with different user groups.

ISMIR

This work was supported by the AHRC grants "An Integrated Audio-Symbolic Model of Music Similarity" AH/M002454/1 and "Digital Music Lab - Analysing Big Music Data" AH/L01016X/1 and the EPSRC Platform Grant "Digital Music" EP/K009559/1.