Adapting Computational Music Similarity Models to Geographic User Groups

Daniel Wolff and Tillman Weyde
City University London, Music Informatics Research Group, daniel.wolff.1@city.ac.uk

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 location. To this end we use information on the country where the data was provided. We finally compare the role of musical features in the specific and general models.

Collected Data

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:

| Age group | Music education |
| Gender | Listening habits |
| Occupation | Favourite genres |
| Languages | Music activities |
| Current location | Favourite media items |
| Birth location | Religious affiliations |
| Geoprofile of friends | Political affiliations |

Users can login with their Facebook account. Given prior permission, user attributes are extracted from their profile data.

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

The Game Interface

- HTML5 web application runs on many devices w. browser
- Odd One Out scenario
- 45 seconds time limit
- Decision and timings are logged
- Game look and feel
- Multiplayer
- Motivation through rewarding of agreement
- Hop-on – hop off: enter or leave anytime

Single Country Datasets

- For first experiments with user groups we use the most frequently annotated attribute: Input location
- Select datasets from the 4 European countries with most data
- Only minor differences expected due to the macroculture character of pop music, but regional microcultures exist.

Modelling Strategy: Transfer Learning

- We model music similarity using the generalisation of the Euclidean distance: The Mahalanobis distance. This allows for training a weighting and combination of music features that correspond to the collected similarity data.
- Training of model achieved with new RITML algorithm, adapted from RITML [Davis et al. 2007]
- Challenge: single country datasets are very small – direct training (RITML) becomes difficult
- Solution: Use transfer learning – models are initialised on a large dataset, then fine-tuned to single country data.
- Idea: Train general model, then adapt to specific group

Process (for DE dataset)

1. Use data from 3 countries (FR + SW + UK) for training a general model "W0".
2. Adapt model to single country (DE) data with W0-RITML method.

Results

<table>
<thead>
<tr>
<th></th>
<th>DE</th>
<th>FR</th>
<th>SE</th>
<th>UK</th>
<th>AVERAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>W0-RITML</td>
<td>69.28%</td>
<td>64.34%</td>
<td>64.40%</td>
<td>70.36%</td>
<td>67.09%</td>
</tr>
<tr>
<td>JOINT</td>
<td>67.80%</td>
<td>67.39%</td>
<td>64.05%</td>
<td>70.46%</td>
<td>67.43%</td>
</tr>
<tr>
<td>RITML</td>
<td>64.35%</td>
<td>62.71%</td>
<td>61.75%</td>
<td>63.78%</td>
<td>63.15%</td>
</tr>
<tr>
<td>Euclidean</td>
<td>60.79%</td>
<td>62.09%</td>
<td>58.11%</td>
<td>62.65%</td>
<td>60.91%</td>
</tr>
</tbody>
</table>

- Training improves performance, but general models provide better results than specific ones
- Used datasets too variable / small to train robust models per country and to analyse model differences
- Fine-tuning only effective on DE dataset
- Resulting model for use as example for country-specific adaptation
- Diagonal -> most of the features are assigned an equal weight
- Larger weights for timbre, timbre&tempo, timbre&tatum

Analysis of Specificities

- Analyse the difference of fine-tuned model to general model
- For DE dataset, exemplary character as more data needed for modelling
- Red entries for specifically strong correlation of features with similarity data, blue for lesser importance
- Specific importance of timbre
- Heightened correlation of tempo and tatum confidence with timbre
- Less correlation to Segment duration, #TatumsPerBeat and loudness factors

Conclusion:
Method allows analysis of features’ influence on similarity, can be extended to ethnomusicology with different user groups.
Future work: Collection of more data with user attributes.