LEARNING A FEATURE SPACE FOR SIMILARITY IN WORLD MUSIC

Maria Panteli, Emmanouil Benetos, Simon Dixon
Centre for Digital Music
Queen Mary University of London
the goal

- study similarity in a large world music collection
related research

music universals

world music features & classification
- [Gomez et al., ISMIR 2009, Kruspe et al., AES 2011, Zhou et al., ICDM 2014]

corpus analysis
- [Serra et al., Nature 2012, Mauch et al., Royal Society 2015, Moelants et al., ISMIR 2009]
world music corpus -> audio descriptors -> feature learning

similarity
dissimilarity
world music corpus → audio descriptors → feature learning → similarity, dissimilarity
what is world music?

“basically, all the music of the world” [Bohlman, Oxford University Press 2002]

Smithsonian Folkways Recordings

- folk and traditional, from as many countries as possible
- 30-second audio previews
- balanced dataset: 31 countries x 70 recs = 2170 recs
world music corpus → audio descriptors → feature learning → similarity, dissimilarity
which descriptors?

“style can be recognized by characteristic uses of form, texture, harmony, melody, and rhythm” [Sadie et al., Oxford University Press 2001]

timbre: MFCCs [Aucouturier et al., IEEE-Multimedia 2005]
rhythm: scale transform [Holzapfel & Stylianou, IEEE-ASLP 2011]
melody: pitch bihistogram [Van Balen et al., ISMIR 2014]
harmony: average chroma [Bartsch & Wakefield, IEEE-Multimedia, 2005]
why these descriptors?

- low-level representations more likely to be robust to the diversity of world music
- state-of-the-art performance in related music similarity tasks
- invariance - check out the poster today! [Panteli & Dixon, ISMIR 2016]
- feature vector for 8-second frames
  - rhythm, melody, timbre, harmony

- from \(~1000\) dimensions to \(~30\) using
  - Principal Component Analysis (PCA)
  - Non-negative Matrix Factorization (NMF)
  - Linear Discriminant Analysis (LDA)
similarity via *country* classification

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Transform. Method</th>
<th>Frame Accuracy</th>
<th>Recording Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>KNN</td>
<td>–</td>
<td>0.175</td>
<td>0.281</td>
</tr>
<tr>
<td></td>
<td>PCA</td>
<td>0.177</td>
<td>0.279</td>
</tr>
<tr>
<td></td>
<td>NMF</td>
<td>0.139</td>
<td>0.214</td>
</tr>
<tr>
<td></td>
<td>LDA</td>
<td>0.258</td>
<td><strong>0.406</strong></td>
</tr>
<tr>
<td>LDA</td>
<td>–</td>
<td><strong>0.300</strong></td>
<td>0.401</td>
</tr>
<tr>
<td></td>
<td>PCA</td>
<td>0.230</td>
<td>0.283</td>
</tr>
<tr>
<td></td>
<td>NMF</td>
<td>0.032</td>
<td>0.032</td>
</tr>
<tr>
<td></td>
<td>LDA</td>
<td><strong>0.300</strong></td>
<td>0.401</td>
</tr>
<tr>
<td>SVM</td>
<td>–</td>
<td>0.038</td>
<td>0.035</td>
</tr>
<tr>
<td></td>
<td>PCA</td>
<td>0.046</td>
<td>0.044</td>
</tr>
<tr>
<td></td>
<td>NMF</td>
<td>0.152</td>
<td>0.177</td>
</tr>
<tr>
<td></td>
<td>LDA</td>
<td>0.277</td>
<td>0.350</td>
</tr>
</tbody>
</table>

Table: Classification accuracies for the predicted frame labels and the predicted recording labels based on a vote count (31 classes x 70 instances, baseline at 0.03)
Figure: Confusion matrix for best classifier in the country classification task.
world music corpus → audio descriptors → feature learning → similarity, dissimilarity
outliers: “recordings that stand out in the collection”

detection via Mahalanobis distance [Aggarwal & Yu, ACM SIGMOD 2001]
Figure: Mahalanobis distances and outliers at the 99.5% quantile of chi-square distribution.
Figure: Number of outliers for each country in our collection (grey areas denote missing data). Botswana had the most outliers (39 out of 70) and Germany the least (0 out of 70).
- **listening test: odd one out** [Wolff & Weyde, ISMIR 2011]
  - 60 outliers x 10 triads each
  - 53% agreement (random baseline 33%)
- some remarks
  - representative dataset
  - reliable descriptors
  - advanced embeddings
  - quantitative vs qualitative evaluation
conclusion

- feature learning for world music content description
- classification confusion between countries of cultural proximity
- geographical regions with most and least outliers
  - moderate agreement in perceptual evaluation of outliers
- …world music analysis is interesting for MIR research

Thank you!
demo: http://tinyurl.com/hk6kq6k