Identifying Cover Songs using NCD

Teppo E. Ahonen, Kjell Lemström

9.7.2008
MML '08
Cover Song Identification

- a MIR task where the goal is to determine whether two songs are different versions of same composition or not
- not an easy task: the covers are usually intentionally different from the original songs
- successful cover song identification yields important information on the similarity between musical pieces
Cover Song Identification

- a common approach: extract the harmonic features (i.e. chords) and compare them
  - chord extraction: using a Hidden Markov Model to estimate chord sequence from chromagram
  - comparison: comparing chord sequences using different sequence matching techniques, for example DTW
- problems: song structure changes, key transpositions, etc.
Normalized Compression Distance

- a similarity metric that is *(quasi)*universal
  - the dominant similarity of the compared features is captured
- used successfully for classifying and clustering composers based on Parsons coded melodies
  - also various other merits, for example clustering genome data and Russian authors
NCD: a very brief explanation

- Kolmogorov complexity $K(x)$: the length of the smallest binary program that produces string $x$
- conditional Kolmogorov complexity $K(x|y)$: the length of the smallest binary program that produces $x$ given the input $y$
  - note: Kolmogorov complexity is incomputable
NCD: a very brief explanation

- Normalized Information Distance (NID): a distance measure based on the Kolmogorov complexity
  - incomputable, as $K$ is incomputable
- Formally,

$$NID(x, y) = \frac{\max \{K(x|y), K(y|x)\}}{\max \{K(x), K(y)\}}$$
incomputable Kolmogorov complexity can be approximated using a standard lossless compression algorithm
- the compressor discovers patterns from the data and thus can compress similar files into smaller size when they are concatenated

the more we can compress the data, the closer we are to the Kolmogorov complexity
NCD: a very brief explanation

- denote $C(x)$ as the size of the string $x$ when compressed using compression algorithm $C$
  - and: $C(x/y) = C(xy) - C(y)$
- Now, approximate $K$ with $C$ and NID turns to NCD:

$$NCD(x, y) = \frac{C(xy) - \min\{C(x), C(y)\}}{\max\{C(x), C(y)\}}$$
NCD and Cover Song Identification

- for testing, we build a system that employs the CompLearn toolkit for NCD calculation
- the chord sequences were extracted by a 24-state HMM, with initial parameters selected according to the work of Bello and Pickens
- as testing material we used the covers80 data set, a collection of 80 pairs of songs
Testing the NCD

- the chord sequences were written into text files using notation that summarizes the chord change into three characters
  - the root note semitone difference and the change between major and minor chords
- distance for every pair of songs was calculated and written into a distance matrix
- tested on two standard compression algorithms: gzip and bzip2
## Distance Matrix Excerpt

<table>
<thead>
<tr>
<th>Song Name</th>
<th>Robert Palmer</th>
<th>The Beatles</th>
<th>Fleetwood Mac</th>
<th>Eric Clapton</th>
<th>Beach Boys</th>
</tr>
</thead>
<tbody>
<tr>
<td>God Only Knows (Brian Wilson)</td>
<td>0.880074</td>
<td>0.843373</td>
<td>0.801205</td>
<td>0.892491</td>
<td>0.777108</td>
</tr>
<tr>
<td>Addicted To Love (Tina Turner)</td>
<td><strong>0.880368</strong></td>
<td>0.918712</td>
<td>0.898773</td>
<td>0.900307</td>
<td>0.900307</td>
</tr>
<tr>
<td>Come Together (Aerosmith)</td>
<td>0.876384</td>
<td><strong>0.850515</strong></td>
<td>0.801546</td>
<td>0.895904</td>
<td>0.827320</td>
</tr>
<tr>
<td>Cocaine (Nazareth)</td>
<td>0.906349</td>
<td>0.919048</td>
<td>0.914286</td>
<td><strong>0.919048</strong></td>
<td>0.903175</td>
</tr>
<tr>
<td>Gold Dust Woman (Sheryl Crow)</td>
<td>0.900369</td>
<td>0.799363</td>
<td><strong>0.770701</strong></td>
<td>0.899317</td>
<td>0.796178</td>
</tr>
</tbody>
</table>
Results

- Out of 80 queries, the system was able to identify the exact match on 10 cases using the bzip2 algorithm
  - with gzip, 9/80 – mostly same songs as with bzip2
- With an answer set size of three, using bzip2 the system found the cover in 20 cases
  - with gzip, 17/80
Remarks on results

- clearly, the results are slightly disappointing
- however, they imply that the NCD approach has potential
  - also, the choice of the compression algorithm had only a slight effect on the results
- current work is focused on improving the chord sequence estimation and searching for a more suitable notation for the chord changes
Conclusion

- improving cover song identification by replacing the string matching with the NCD metric
  - tested on a system that uses HMMs to estimate the chord sequences and two standard compression algorithms for the distance matrix calculation
- results are slightly disappointing but the tests imply that the NCD approach could work, after a suitable notation for the chord changes is discovered
  - thus, the work continues
References

- CompLearn toolkit: http://www.complearn.org/
- Bello and Pickens: A Robust Mid-Level Representation for Harmonic Content in Music Signals. *Proc. ISMIR’05*
Questions?

Thank you.