CONTEXT-BASED MUSIC SIMILARITY ESTIMATION

Peter Knees
Markus Schedl

Department of Computational Perception
Johannes Kepler University (JKU)
Linz, Austria

LSAS 2009, 3rd Workshop on Learning the Semantics of Audio Signals, Graz, Austria, December 2009
Learning Semantics of Audio Signals

Which semantics?

The “semantics” are determined by the outside world, i.e., the context

“Similarity” as perceived by humans is not exclusively determined by the audio

Influenced by, e.g., marketing strategies, political views of artists, language of songs, decade of activity, listening context, mood, peers, etc., etc., etc.

Analysis of audio alone probably won’t do
Audio similarity may find that these two sound similar:

- Foxboro Hot Tubs “Ruby Room”
- The Staggers “Little Boy Blue”

But, for example, it won’t tell you that...

- “Foxboro Hot Tubs” are better known as “Green Day”
- “The Staggers” are a band from Graz
What do these songs have in common?

NOFX
“Idiot Son of an Asshole”

Eminem
“Mosh”

Answer:
Both are Anti-Bush protest songs.
What do these artists have in common?
(Example borrowed from Lamere & Celma’s Music Recommendation Tutorial)

Ravi Shankar  Norah Jones

Answer:
Half of their DNA. Norah Jones is Ravi Shankar’s daughter.
What do these songs have in common?

Antonio Carlos Jobim “Insensatez”
Rammstein “Rammstein”

Answer:
Both were featured on the Soundtrack of David Lynch’s movie “Lost Highway”
Advantages of Content Analysis

• Features can be extracted from any audio file
• No other data or community necessary
• No cultural biases (i.e., no popularity bias, no subjective ratings etc.)

Advantages of Context Analysis

• Capture aspects beyond pure audio signal
• No audio file necessary
• Usually, user-based features are closer to what users want

Challenge for both Content and Context Analysis

• Extraction of relevant features from (noisy) signal
Manifold applications what can be done with contextual data. In this talk we focus on approaches that explicitly target the **estimation of similarity** of musical entities

**Not in this talk**…
- Collaborative Filtering, User modeling, Foafing
- Server Log Mining (difficult to obtain)
- Social Network Mining (only preliminary results)

**In this talk**…

**Text-based Similarity**
- Web-Terms
- Tags
- Lyrics

**Co-Ocurrence-based Similarity**
- Playlists
- Page Counts
- P2P Networks
Using **traditional Text-IR concepts** to deal with textual data related to music

**Bag-of-Words approach**

Text chunked into words (or n-grams): text = unsorted accumulation of terms

**Part-of-Speech (POS) Tagging**

Determines the linguistic category for each word in a text; e.g., used to extract all adjectives

**Term weighting**

Assigns a score to each term for each document. Very frequently a variant of the **TF-IDF scheme**:

\[
w_{t,d} = tf_{t,d} \cdot \log \frac{N}{df_t}
\]

**Vector Space Model**

Each term represents a dimension, value = weight

Each doc is represented as a vector which dimensionality equals the number of distinct terms
Latent Semantic Analysis (LSA)

Similarity/Distance Calculation

• Euclidean Distance
  \[ d(a, b) = \sqrt{\sum (a_i - b_i)^2} \]

• Cosine Similarity
  \[ sim(a, b) = \cos \theta = \frac{\sum a_i b_i}{\sqrt{\sum a_i^2 \sum b_i^2}} \]

• L₁ (Manhattan Distance)
  \[ d(a, b) = \sum |a_i - b_i| \]

• Overlap Score
  sum up terms \( t_i \) for which \( a_i \neq 0 \) && \( b_i \neq 0 \)
Basic Idea

• Analyze arbitrary texts about music (artists) from Web
• Create characteristic term profiles for each piece

The “Let’s Use Google” Approaches

[Whitman & Lawrence, 2002], [Baumann & Hummel, 2003], [Knees et al., 2004]

• Take name of artist, add some constraints (e.g., +music +review), send it to Google
• Retrieves a list of relevant Web pages
• Optionally, apply filtering of noisy pages
• Throw retrieved texts together and calculate features (TF-IDF)

• Works best on artist level
• No restriction to specific sites
• Unstructured text data
• Number of terms very high (dimensionality!)
The “Let’s Use Google and a Dictionary” Approaches

Only specific terms are considered in order to lower dimensionality and exclude noisy features
- manually compiled list of (musically relevant) terms [Pampalk et al., 2005]
- automatically generated using tags from Last.fm [Pohle et al., 2007]

Not necessarily better, especially important if features are presented to user

Retrieving Texts from Specific Sources

- Mining texts from review pages such as epinions.com [Hu et al., 2005]
- Mining texts from mp3-Blogs (RSS feeds) [Celma et al., 2006]

Structure of data known, facilitates extraction of relevant text, limited to tracks/album/artists included in the service
Basic Idea

• Use user/community generated tags for music as textual input

Steps

• Retrieve tags for artist or track from Last.fm
• Cleaning of noisy and redundant tags
  - manually
  - automatically [Geleijnse et al., 2007]
• List of collected terms is treated as text document and TF-IDF’d [Levy & Sandler, 2007]
• Optionally, LSA to reduce dimensionality
• Comparison of vectors via cosine similarity (or overlap score)

• Data available in standardized fashion
• Dedicated terms for music
• Lower dimensionality (13,500 tags vs. >200,000 Web terms [Levy & Sandler, 2007])
• Depends on community
Basic Idea

• Analyze the lyrics for a song (lyrics are usually easily available)

Topic Features

[Logan et al., 2004]
• Typical topics for lyrics are distilled from a large corpus using (P)LSA
  (“Hate”, “Love”, “Blue”, Gangsta, Spanish)
• Lyrics are transformed to topic-based vectors, similarity is calculated via $L_1$ distance

• Alternative approaches use TF-IDF with optional LSA and Stemming for Mood Categorization
  [Laurier et al., 2008], [Hu et al., 2009]

Rhyme Features [Mayer et al., 2008]

• Phonetic transcription is searched for patterns of rhyming lines (AA, ABAB, AABB)
• Frequency of patterns + statistics like $words per minute$, $punctuation freq.$ etc.

Other Features [Mahedero et al., 2005]

• Language, Structure
<table>
<thead>
<tr>
<th></th>
<th>Web-Terms</th>
<th>Tags</th>
<th>Lyrics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Source</strong></td>
<td>arbitrary Web pages</td>
<td>Web service</td>
<td>lyrics portal</td>
</tr>
<tr>
<td><strong>Community-based</strong></td>
<td>depends</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td><strong>Level</strong></td>
<td>artists</td>
<td>artists (tracks)</td>
<td>tracks (artists)</td>
</tr>
<tr>
<td><strong>Feature Dimensionality</strong></td>
<td>very high</td>
<td>moderate</td>
<td>possibly high</td>
</tr>
<tr>
<td><strong>Specific Bias</strong></td>
<td>low</td>
<td>community</td>
<td>none</td>
</tr>
<tr>
<td><strong>Potential Noise</strong></td>
<td>high</td>
<td>moderate</td>
<td>low</td>
</tr>
</tbody>
</table>
Approaches based on Co-Occurrences: Playlists

[Pachet et al., Proc. WEDELMUSIC, 2001]

analysis of co-occurrences of artists and songs on
- radio station playlists (French radio station *Fip*)
- compilation CD databases (CDDB)

"co-occurrence" of an entity with itself is its number of occurrence

normalizing in order to account for different frequency/popularity of entities

similarity of 2 entities $A_i$ and $A_j$:

$$\text{sim}_{pl\_cooc}(A_i, A_j) = \frac{1}{2} \cdot \left[ \frac{\text{cooc}(A_i, A_j)}{\text{cooc}(A_i, A_i)} + \frac{\text{cooc}(A_j, A_i)}{\text{cooc}(A_j, A_j)} \right]$$

shortcoming: cannot capture indirect links
Approaches based on Co-Occurrences: Playlists

[Pachet et al., Proc. WEDELMUSIC, 2001]

correlation on vector representation to capture indirect links:

\[
sim_{pl\_corr}(A_i, A_j) = \frac{\text{Cov}(A_i, A_j)}{\sqrt{\text{Cov}(A_i, A_i) \cdot \text{Cov}(A_j, A_j)}}
\]

**insights:**
- co-occurrence performed better than correlation
  (too much irrelevant info in feature vectors?)
- rather small test collection
Approaches based on Co-Occurrences: Playlists

[Cano and Koppenberger, Proc. ISMIR, 2004]

29,000 playlists from "Art of the Mix"

Focus on creating and analyzing a similarity network based on co-occurrences (straightforward def.: edge if two artists co-occur in playlist)

Insights:
- Small average degree of a node (12.5) → Sparse similarity matrix
- Similarity measure can only capture positive relations (high similarity)
- Only one large cluster connecting more than 99% of artists (population bias, hubs, popularity bias?)
- Average shortest path is only 3.8 → Probability of indirect links (clustering coefficient) is 0.1
Approaches based on Co-Occurrences: Playlists

[Baccigalupo et al., Proc. ISMIR, 2008]

> 1 mio. playlists from "MusicStrands"

subset of 4,000 most popular artists

distance between the occurrence of two artists in a playlist taken into account → $d_h(A_i, A_j)$: co-occurrence count of song by $A_i$ and song by $A_j$ at distance of $h$

\[
dist_{\text{pl-d}}(A_i, A_j) = \sum_{h=0}^{2} \beta_h \cdot [d_h(A_i, A_j) + d_h(A_j, A_i)]
\]

$\beta_0 = 1, \beta_1 = 0.8, \beta_2 = 0.64$
Approaches based on Co-Occurrences: Playlists

[Baccigalupo et al., Proc. ISMIR, 2008]

normalization w.r.t. popularity:

\[ \text{dist}_{pl-d}(A_i, A_j) = \frac{\text{dist}_{pl-d}(A_i, A_j) - \text{dist}_{pl-d}(A_i)}{\max \left( \text{dist}_{pl-d}(A_i, A_j) - \text{dist}_{pl-d}(A_i) \right)} \]

no evaluation relevant to similarity measurement
Approaches based on Web Co-Occurrences: General Remarks

• Web as very rich source of context information

• context defined similarly to "text-based approaches" as textual surrounding of a music entity on a Web page (artist name)

• BUT: how to find the musically relevant Web pages?

  - build own crawler and indexer (focused crawling)
    + high quality, relevant pages
    – slow

  - rely on results of search engines
    – noisy
    – "black box"
    – restricted
    + easy
    + can be very fast
Approaches based on Web Co-Occurrences: Challenges

Peter Knees and Markus Schedl

music@jku.at | http://www.cp.jku.at

3rd Workshop on Learning the Semantics of Audio Signals, Graz, Austria, December 2009

Britney Spears - official web site and blog
20 Nov 2000 ... Access Britney Spears photos, galleries and videos. Get the latest news direct from Britney on her official blog
Photos - Tour - Videos - Got Naked (They Had A Plan)
www.britneyspears.com - Cached - Similar

Concert Photos. Brisbane Night 3 - BRITNEY SPEARS
29 Nov 2009 ... Here's your second-to-last gallery of Circus concert photos! Flip through to see the images from Britney's third night in Brisbane.
www.britneyspears.com/.../concert-photos-brisbane-night-3.php - 16 hours ago

Britney Spears - Wikipedia, the free encyclopedia
Britney Jean Spears (born December 2, 1981) is an American singer and entertainer. Born in Mississippi and raised in Louisiana, Spears first appeared on ... Discography - Videography - Products - Filmography
en.wikipedia.org/wiki/Britney_Spears - Cached - Similar

Image results for britney spears - Report images
Web Co-Occurrences / Page Counts: Simple Approach


"Alice Cooper"
"Alice Cooper" + "BB King"
"Alice Cooper" + "Beethoven"
...
"Alice Cooper" + "ZZ Top"
"BB King"
...
"ZZ Top"

[Google page counts]

+music+review

3rd Workshop on Learning the Semantics of Audio Signals, Graz, Austria, December 2009

Peter Knees and Markus Schedl
Web Co-Occurrences / Page Counts: Simple Approach


Web services for artist recommendation

retrieve possibly related artists to seed artist from "Amazon Listmania!"

Google Web API to obtain a (sparse) co-occurrence matrix

similarity normalized for popularity via $\min()$

\[
sim_{pc-min}(A_i, A_j) = \frac{pc(A_i, A_j)}{\min(pc(A_i), pc(A_j))}
\]

only empirical evaluation

(co-occurrence) page counts

<table>
<thead>
<tr>
<th></th>
<th>100</th>
<th>3</th>
<th>5</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>91</td>
<td>27</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>8</td>
<td>96</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>12</td>
<td>84</td>
<td></td>
</tr>
</tbody>
</table>
Web Co-Occurrences / Page Counts: Simple Approach

[Schedl et al., Proc. CBMI, 2005]

aim: artist similarity estimation

model of conditional probability

normalization and symmetrization via cross-artist-probability

\[
\text{sim}_{pc, cp}(A_i, A_j) = \frac{1}{2} \cdot \left( \frac{pc(A_i, A_j)}{pc(A_i)} + \frac{pc(A_i, A_j)}{pc(A_j)} \right)
\]

complete similarity matrix

also combination of different query schemes ("googling-approaches")

evaluation via genre classification task (224 artists, 14 genres): 85% acc.
Web Co-Occurrences / Page Counts: Fetching Pages

[Schedl et al., Proc. CBMI, 2005]

**problem:** quadratic complexity in #artists

**solution:** retrieve content of top-ranked pages and index them with artist names

→ #queries linear with #artists

[Schedl, PhD thesis, 2008][Chapter 3]
Web Co-Occurrences / Page Counts: Data Processing Pipeline

100 top-ranked URLs

Alice Cooper
http://www.geocities.com/sfloman/alicecooperband.html
... retrieve Web pages

indexing

+music

calculate DFs

(co-occurrence) page counts

Peter Knees and Markus Schedl
3rd Workshop on Learning the Semantics of Audio Signals, Graz, Austria, December 2009
Approaches based on Co-Occurrences: P2P Networks

make use of meta-data transmitted as files names or ID3 tags in peer-to-peer networks

information gathered from users' shared folders (no file downloads)

early work using "OpenNap":

[Whitman and Lawrence, Proc. ICMC, 2002]
[Ellis et al., Proc. ISMIR, 2002]
[Logan et al., Eval. MIR Systems Workshop @ SIGIR, 2003]
[Berenzweig et al., Proc. ISMIR, 2003]
Approaches based on Co-Occurrences: P2P Networks

[Logan et al., 2003] and [Berenzweig et al., 2003]

- 400 most popular artists in mid 2002
- 175,000 user-to-artist relations from 3,200 shared collections

similarities via artist co-occurrences in collections (cond. prob.)

sparsity of the co-occurrence matrix

compared with "Art of the Mix" co-occurrences, AMG's similar artists, similarity data from a Web survey and content-based similarity (MFCCs)

evaluation via overlap between top N most similar artists for different sources

insights:
- "OpenNap" and "Art of the Mix" revealed a high overlap (0.6)
- Web survey and AMG showed a moderate overlap (0.4)
- MFCCs had a low agreement with all other sources (0.1)
→ content-based features capture very different similarity aspects
Approaches based on Co-Occurrences: P2P Networks

[Whitman and Lawrence, Proc. ICMC, 2002]

alleviate popularity bias in the similarity measure:

\[ \text{sim}_{p2p-wl}(A_i, A_j) = \frac{C'(A_i, A_j)}{C(A_j)} \cdot \left(1 - \frac{|C'(A_i) - C'(A_j)|}{C'(A_k)}\right) \]

- \( C(A_i) \) number of users that share artist \( A_i \)
- \( C(A_i, A_j) \) number of users that have both \( A_i \) and \( A_j \) in their shared collection
- \( A_k \) most popular artist in the corpus
Approaches based on Co-Occurrences: P2P Networks

[Ellis et al., Proc. ISMIR, 2002]

400,000 user-to-song relations from "OpenNap"

about 3,000 artists

similarity defined as Erdős distance in artist-similarity-graph:
\[ d(A_i, A_j) \] equals number of intermediate artists on shortest path from \( A_i \) to \( A_j \)

accounts for indirect links (per definition)

alternative formulation "Resistive Erdős":
assumes that 2 artists are more similar if they are connected via many paths

\[
\text{dist}_{p2p_{res}}(A_i, A_j) = \frac{1}{\sum_{p \in \text{Paths}(A_i, A_j)} \frac{1}{|p|}}
\]

no improvement with "Resistive Erdős" due to popularity bias
Approaches based on Co-Occurrences: P2P Networks


meta-data of shared files in "Gnutella" network retrieved in November 2007 (.mp3 and .wav)

1.2 million users; 530,000 songs

distance measure on the song level

accounts for popularity bias

$$dist_{p2p-pop}(S_i, S_j) = -log_2 \left( \frac{uc(S_i, S_j)}{\sqrt{C_i \cdot C_j}} \right)$$

$uc(S_i, S_j)$ number of users that share songs $S_i$ and $S_j$

$C_i, C_j$ popularity of $S_i$ and $S_j$, measured as total number of occurrences

evaluation in a recommendation setting (30% of songs of each user’s collection used to predict remaining 70%): about 12% prec., 13% rec.

heavy inconsistencies in meta-data (ID3 tags)
Co-Occurrence-based Approaches: Summary

<table>
<thead>
<tr>
<th></th>
<th>Playlists</th>
<th>Web Co-Ocs</th>
<th>P2P nets</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Source</strong></td>
<td>radio, compilation CDs, Web services</td>
<td>search engines, Web pages</td>
<td>shared folders</td>
</tr>
<tr>
<td><strong>Community-based</strong></td>
<td>depends on source</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td><strong>Level</strong></td>
<td>artists (tracks)</td>
<td>artists</td>
<td>artists (tracks)</td>
</tr>
<tr>
<td><strong>Specific Bias</strong></td>
<td>low</td>
<td>&quot;wikipedia&quot;-bias</td>
<td>community</td>
</tr>
<tr>
<td><strong>Potential Noise</strong></td>
<td>low</td>
<td>high</td>
<td>high</td>
</tr>
</tbody>
</table>
Summary and Discussion

• estimating similarity based on the context of a music entity
• "context" can be defined in various ways
• "cultural knowledge", "community meta-data", "context-based features"
• context of music entity as complementary source of information
• combination of context- and content-based features

• still no common evaluation data set like in TREC
  → each publication uses its own data set
  → need to establish common "ground truth"
Challenges / Shortcomings of Context-based Methods

- **data sparsity** (especially in "long tail")

- **popularity bias**: disproportionately more info is available for popular artists than for lesser known one

- **community/population bias**: only participants of the community under consideration are taken into account (e.g., certain P2P network, last.fm, myspace, ...);

  users of certain communities may not represent the average music listener, but rather share similar music tastes
Future Challenges: What we believe

• establish a differentiated, multi-granular concept of music similarity that takes into account
  cultural areas,
  regional particularities,
  individual views and tastes (→ personalization)

• transcend the current focus of research on Western music

• integrate multi-faceted music similarity measures in music applications that should combine both content- and context-based information

  → enable retrieval systems to deal with queries like
  "give me rhythmically similar stuff to the most recent chart hits in Canada, but which was released in the 1970s (probably somewhere else)"

• more (richer?) info sources to discover and make accessible
This is the end!

thank you!

peter.knees@jku.at

markus.schedl@jku.at