Extracting Semantic Information from on-line Art Music Discussion Forums.

Mohamed Sordo, Joan Serrà, Gopala Krishna Koduri and Xavier Serra
Outline

• Introduction
• Background
• Methodology
• Experimental Results
• Conclusions
• Future Work
Introduction (I)

• Understanding music requires (also) understanding how listeners
  • perceive music
  • consume it or enjoy it
  • share their tastes among other people.

• The online interaction among users results in the emergence of online communities.
Introduction (II)

• Online community:
  • “a persistent group of users of an online social media platform with shared goals, a specific organizational structure, community rituals, strong interactions and a common vocabulary” (Stanoevska-Slabeva [2002])
Introduction (3)

• By mining UGC (text) we can obtain music-related information that could not otherwise be extracted from audio signals or symbolic score representations.

• We propose a methodology for extracting music-related semantic information from online art music discussion forums.
Background

• Extracting semantic information from online forums -> only in text mining.
  • Structured data (Yang et al. [2009]), detect high quality posts and topics (Weimer et al. [2007], Chen et al. [2008]), topic and opinion leader detection (Zhu et al. [2010])

• Mining UGC in Music Information Retrieval
  • Reviews (Whitman et al. [2002]), Blogs (Celma et al. [2006]), Social tags (Lamere et al. [2008]), Web documents (Schedl et al. [2010]), etc.
  • No approach in MIR has analyzed discussion forums
Methodology

• Step 0: dictionary definition
• Step 1: text processing
• Step 2: network creation
• Step 3: network cleaning
Step 0: dictionary definition (I)

- Flat taxonomy (category - word)
  - MusicBrainz
    - **per-song:**
      - composers
      - lyricists
      - performers
      - recordings
      - works
      - instruments
    - **intra-song:**
      - e.g.: ragas, talas, makams, usuls
Step 0: dictionary definition (II)

- Flat taxonomy (category - word)
  - DBpedia

Seed category

Carnatic music
Step 0: dictionary definition (II)

- Flat taxonomy (category - word)
  - DBpedia

  Carnatic Ragas

  Sangeetha Kalanidhi recipients

  Carnatic music

  Carnatic compositions

  Carnatic musicians

  Carnatic music instruments

  Carnatic music terminology

  Carnatic classical Music festivals

Seed category
Sub-categories
Step 0: dictionary definition (II)

- Flat taxonomy (category - word)
  - DBpedia
    - Carnatic Ragas
    - bhairavi
    - Carnatic music
    - Sangeetha Kalanidhi recipients
    - Carnatic classical Music festivals
    - Carnatic musicians
    - Carnatic compositions
    - Carnatic music instruments
    - mridangam
    - Carnatic music terminology
    - Instrumentalists
    - Singers
    - Composers

Seed category
Sub-categories
Sub-sub-categories
Articles
Step 0: dictionary definition (III)

- Dictionary examples

<table>
<thead>
<tr>
<th>Category</th>
<th>Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>Composer</td>
<td>Dede Efendi</td>
</tr>
<tr>
<td>Performer</td>
<td>Bhimsen Joshi</td>
</tr>
<tr>
<td>Raga</td>
<td>Bhairavi</td>
</tr>
<tr>
<td>Makam</td>
<td>Hicaz</td>
</tr>
<tr>
<td>Tala</td>
<td>Ektal</td>
</tr>
<tr>
<td>Instrument</td>
<td>Mridangam</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Step 1: text processing

• Match dictionary with the text of forum posts
  • NLP techniques: Tokenization + Part-of-Speech Tagging

IMO there is no reason for the experts committee to interfere in this since one of the 'supreme experts' 😝 Sri Dikshitar has codified clearly the difference between AbhEri and dEvagAndhAram [recently known as karnATaka dEvagAndhAri] by including the rAga mudras in his krithis.

viNAbhEri is the AbhEri krithi and panCAshatpITha [rAga mudra is in the line 'dEvagAndhAra rAga tOSiNi '] is one of the krithis is dEvagAndhAram.

Since AbhEri and dEvagAndhAram are two clearly separate rAgAs, the current trend of singing naguMoMu ganalEni in dEvagAndhAram is quite sad.

the difference between AbhEri and dEvagAndhAram

DT  NN  IN  NN  CC  NN

DT: determiner, NN: noun, IN: preposition, CC: coordination conjunction
Step 1: text processing

• Match dictionary with the text of forum posts
• NLP techniques: Tokenization + Part-of-Speech Tagging

IMO there is no reason for the experts committee to interfere in this since one of the 'supreme experts' 😊 Sri Dikshitar has codified clearly the difference between AbhEri and dEvagAndhAram [recently known as karnATaka dEvagAndhAri] by including the rAga mudras in his krithis.

vINabhEri is the AbhEri krithi and panCAshatplItha [rAga mudra is in the line 'dEvagAndhAra rAga tOSiNi '] is one of the krithis is dEvagAndhAram.

Since AbhEri and dEvagAndhAram are two clearly separate rAgAs, the current trend of singing nagumOmu ganalEni in dEvagAndhAram is quite sad.

the difference between AbhEri and dEvagAndhAram
difference: noun, AbhEri: noun, and: coordination conjunction

DT: determiner, NN: noun, IN: preposition, CC: coordination conjunction

dictionary

nouns and adjectives
Step 1: text processing

- Match dictionary with the text of forum posts
- NLP techniques: Tokenization + Part-of-Speech Tagging

**Example**

IMO there is no reason for the experts committee to interfere in this since one of the 'supreme experts' 😊 Sri Dikshitar has codified clearly the difference between AbhEri and dEvagAndhAram [recently known as karnATaka dEvagAndhAri] by including the rAga mudras in his krithis.

vINabhEri is the AbhEri krithi and panCAshatplTha [rAga mudra is in the line 'dEvagAndhAra rAga tOSiNi '] is one of the krithis is dEvagAndhAram.

Since AbhEri and dEvagAndhAram are two clearly separate rAgAs, the current trend of singing nagumOmU ganalEni in dEvagAndhAram is quite sad.

| * | difference | * | AbhEri | * | dEvagAndhAram |
| DT | NN | IN | NN | CC | NN |

DT: determiner, NN: noun, IN: preposition, CC: coordination conjunction

dictionary

* non-eligible words

nouns and adjectives
Step 2: network creation

- Undirected weighted network:
  - nodes: terms in the dictionary + nouns & adjectives
  - edges: if the two nodes are close in the text (close?)

* difference * AbhEri * dEvagAndhAram

DT NN IN NN CC NN

- Link Threshold ($L$)

$L = 2$
Step 3: network cleaning (I)

• The previous step can yield a very dense network
  • Very high avg. degree (num. of edges per node)
  • Noise

• Possible solutions:
  • Remove less frequent terms (Frequency threshold, $F$)
  • Apply disparity filter ($\rho$, Serrano et al. [2010])
Step 3: network cleaning (II)
Step 3: network cleaning (II)

$L$: Link threshold
$F$: Frequency threshold
Step 3: network cleaning (II)

$L$: Link threshold
$F$: Frequency threshold
$\rho$: Disparity filter
Evaluation measures

- **Network-related measures:**
  - Centrality (betweenness, closeness, katz, degree, etc.)
  - Communities/clusters

| Nodes    | - Centrality (betweenness, closeness, katz, degree, etc.)
<table>
<thead>
<tr>
<th></th>
<th>- Communities/clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Edges</td>
<td>- Frequency (degree)</td>
</tr>
<tr>
<td></td>
<td>- Relevance (disparity filter)</td>
</tr>
</tbody>
</table>

- **Other measures:**
  - semantically connecting terms
  - e.g.: lineage, musical influence
  - ranking measures to compare different networks

Newman [2010]
Experimental results (I)

- rasikas.org:
Experimental results (I)

- rasikas.org:

Discussion thread on rasikas.org regarding RV Vedavalli's performance:

1. **sankark**: Thanks be to Nick H. Believe he sponsored today's concert; also for the Sri Krishna Sweets sweet/savory package 😊

It was an experience that can't be written about, even with thousands of words. So, a simple "Thank You" to RV & team (Sumitra Vocal Support, RKSK Violin, KAP Mridangam & N Guruprasad Ghatam). Without much further ado, here goes the list of songs:

- mAnasa guruguha – Anandabhairavi – rUpakam – MD (S@manasa).
- kamalIapatakula – brndavana sAranga – Adi – T
- parama pAvana – pUrvikalyani – Adi – Poochi Srinivasa lyengar (R NS@parama pavana, sublime NS)
- soundarAjam – brndavana sAranga – rUpakam – MD
- nenarunci nAnu – malavasri mAlavi – Adi – T – what a contrast compared to the previous slow tempo song, this was in jet speed.
- bhajarE rE cItta – kalyAni – m cApu – MD (RNS@devlm sakthi, T. Energetic and well appreciated thani)
- mAmava pattAbi rAma – manirangu – m cApu – MD
- avalokayE sri gDevindam – kurinji? – Adi – T
- slokam (vandE brmdAvana) in sAvi, sahAna, surati followed by pathiki hArathi re – Adi – T mangalam

2. **ganeshkant**: Really missed this one since away from Chennai for a week. Two Brin.sAranga and manirangu that too my favourite mAmava.„My thanks to Nick for arranging such a wonderful concert.

3. **cRama**: RV's concert sponsored by Nick was an enlightening and enjoyable experience in itself. The start was with Manasguruguha – not with a varnam or a Canapathi kriti as is normally done. I have observed this trend in her many recent concerts – For eg, Mukhathai kaati in Academy 2010, Meenakshiranjana in MMI Remembrance day concert and many other recent concerts. The swaras for Anandabhairavi were rendered in two kalamas very elaborately and had very interesting and unusual combinations. There were lot of usage of sa ga ma in the place of sagarigama which is normally used in plenty in Anandabhairavi. Poorvikalyani alapanas was brief and the alapanas was a lead to the pallavi of the kriti without violin interlude.

In yesterday's concert, I felt like she had decided to explore into new vistas within the structural framework of that raga. I felt this because normally in a concert the length of navaval and swarams will be structured like progressive and then kurai in if it is there and some structured
## Experimental results (II)

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Num. sub-forums</strong></td>
<td>20</td>
</tr>
<tr>
<td><strong>Num. topics</strong></td>
<td>16,595</td>
</tr>
<tr>
<td><strong>Num. posts</strong></td>
<td>192,292</td>
</tr>
<tr>
<td><strong>Posts per topic</strong></td>
<td>$\mu = 11.59$, $\sigma = 34.49$, median = 5</td>
</tr>
<tr>
<td><strong>Num. active topics</strong></td>
<td>1,362 active in the last 12 months</td>
</tr>
<tr>
<td><strong>Num. users</strong></td>
<td>4,332 (with at least one post)</td>
</tr>
<tr>
<td><strong>Num. active users</strong></td>
<td>929 active in the last 12 months</td>
</tr>
</tbody>
</table>

Statistics of rasikas.org as of March 6\(^{th}\), 2012

- Not all the sub-forums are of our interest
  - We selected a subset of 11 sub-forums, 14,309 topics and 172,249 posts
- We generate a network following our proposed methodology
## Experimental results (III)

### Experiment 1: Node betweenness centrality

<table>
<thead>
<tr>
<th>Rank</th>
<th>Raagas</th>
<th>Taalas</th>
<th>Instruments</th>
<th>Performers</th>
<th>Composers</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Nata</td>
<td>Adi</td>
<td>Violin</td>
<td>Chembai</td>
<td>Tyagaraja</td>
</tr>
<tr>
<td>2</td>
<td>Kalyani</td>
<td>Rupakam</td>
<td>Mridangam</td>
<td>Madurai Mani Iyer</td>
<td>Annamacharya</td>
</tr>
<tr>
<td>3</td>
<td>Bhairavi</td>
<td>Chapu</td>
<td>Vocal</td>
<td>Charulatha Mani</td>
<td>Purandara Dasa</td>
</tr>
<tr>
<td>4</td>
<td>Ragamalika</td>
<td>Jhampa</td>
<td>Ghatam</td>
<td>Kalpakam Swaminathan</td>
<td>Swati Tirunal</td>
</tr>
<tr>
<td>5</td>
<td>Kannada</td>
<td>Misram</td>
<td>Morsing</td>
<td>Lalgudi Jayaraman</td>
<td>Papanasam Sivan</td>
</tr>
</tbody>
</table>
Experimental results (IV)

- Experiment 2: Term co-occurrences
  - Frequent co-occurrences: predicting performer/instrument pairs.

<table>
<thead>
<tr>
<th>Parameter configuration</th>
<th>$F = 10, L = 5, \ \rho = 0.01$</th>
<th>$F = 10, L = 10, \ \rho = 0.01$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Num. matched performers</td>
<td>104</td>
<td>114</td>
</tr>
<tr>
<td>Num. matched perf.-instr. pairs</td>
<td>63</td>
<td>70</td>
</tr>
<tr>
<td>Hit %</td>
<td>95.24</td>
<td>80.00</td>
</tr>
<tr>
<td>Mean Reciprocal Rank</td>
<td>95.24</td>
<td>85.48</td>
</tr>
</tbody>
</table>
Experimental results (IV)

- **Experiment 2: Term co-occurrences**
  - Relevant co-occurrences

<table>
<thead>
<tr>
<th>Raaga</th>
<th>Raaga</th>
<th>Relev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kedaram</td>
<td>Gowla</td>
<td>0.121</td>
</tr>
<tr>
<td>Bhavani</td>
<td>Bhavapriya</td>
<td>0.109</td>
</tr>
<tr>
<td>Manavati</td>
<td>Manoranjami</td>
<td>0.092</td>
</tr>
<tr>
<td>Kalavati</td>
<td>Yagapriya</td>
<td>0.088</td>
</tr>
<tr>
<td>Nadamakriya</td>
<td>Punnagavarali</td>
<td>0.081</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Raaga</th>
<th>Composer</th>
<th>Relev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abhang</td>
<td>Tukarama</td>
<td>0.159</td>
</tr>
<tr>
<td>Yaman Kalyani</td>
<td>Vyasa Raya</td>
<td>0.149</td>
</tr>
<tr>
<td>Pharaz</td>
<td>Dharmapuri Subbarayar</td>
<td>0.143</td>
</tr>
<tr>
<td>Reethi Gowlai</td>
<td>Subbaraya Sastri</td>
<td>0.122</td>
</tr>
<tr>
<td>Andolika</td>
<td>Muthu Thandavar</td>
<td>0.108</td>
</tr>
</tbody>
</table>
Experimental results (V)

• Experiment 3: Term semantic relations
  • Relations such as:
    • Musical influence (guru, disciple)
    • Family (father, mother, uncle, son, etc.)
  • From a total of 24 relations, our method correctly infers 14 (58%)
  • Some examples:
    ✓ Msn Murthy – (Husband, Wife) – Pantula Rama
    ✓ Vasundhara Devi – (Mother) – Vyjayanthimala
    ✓ Palghat Mani Iyer – (Guru) – Palghat Raghu
    ✗ Palghat Raghu – (Disciple) – P.S. Nayaranaswamy
    ✗ Karaikudi Mani – (Guru) – G. Harishankar
Conclusions

- A method for extracting musically-meaningful semantic information from online discussion forums.
  - Definition of a dictionary of art music tradition terms
  - Undirected weighted network
    - Nodes: matched dictionary terms + nouns and adjectives
    - Edges: relations of closeness between pairs of terms
  - Network analysis:
    - Node relevance
    - Term co-occurences
    - Term semantic relations
Future work

• Current work in progress:
  • Compare network structure with network of links between Wikipedia articles.
Future work

• Current work in progress:
  • Compare network structure with network of links between Wikipedia articles.
  • Communities of terms/concepts via clustering techniques (e.g., k-means)
Future work

• Current work in progress:
  • Compare network structure with network of links between Wikipedia articles.
  • Communities of terms/concepts via clustering techniques (e.g., k-means)

• Future work:
  • Contextual information (e.g., musical seasons)
  • More sophisticated NLP techniques
  • Capture user opinions
  • Filter forum posts by user relevance
  • More complete dictionaries/ontologies
References

References


Thanks!