Learning Shared Vector Representations of Lyrics and Chords in Music

Timothy Greer, Karan Singla, Benjamin Ma, Shri Narayanan

University of Southern California

Paper ID 1934
ICASSP 2019
May 15th 2019, Brighton, UK
Motivation

- Music is something we all enjoy
- Lyrics and music affect us differently
- Can we capture this joint effect?
- What other musical phenomena can we find?
Short Demonstration

• “Sad” song with “sad” lyrics:

“I hurt myself today, to see if I still feel”

• "Happy” song with “sad” lyrics:

“I’ve been lonely and sad about the things that you do”

ABBA, “Mamma Mia” ABBA, Polar. Spotify, https://open.spotify.com/track/22NN4BS1AlqVbyKIWEExgON?si=9Fsgknt2RIOGtNpg_fSbLQ
Methods

• Word embeddings model (Word2Vec)

• Use two “languages” to find context
  • Chords and lyrics

• Use this model on pilot task

• Make music-theoretical observations
The Model

- Created embeddings similar to Luong, 2015

Luong, H Pham, and C D Manning, “Bilingual word representations with monolingual quality in mind,” in Proc. of the 1st Workshop on Vector Space Modeling for Natural Language Processing, 2015, pp. 151–159
Learning Embeddings

- Monolingual loss term:
  \[ \text{MONO} = \frac{1}{T} \sum_{t=1}^{T} \sum_{-l \leq j \leq l, j \neq 0} \log(p(w_{t+j}|w_t)) \]

- Cross-lingual loss term:
  \[ \text{CROSS}_{WC} = \frac{1}{T} \sum_{t=1}^{T} \sum_{-l \leq j \leq l} \log(p(c_{k+j}|w_t)) \]

- Training objective function:
  \[ \mathcal{L} = \text{MONO}_W + \text{MONO}_C + \text{CROSS}_{CW} + \text{CROSS}_{WC} \]
The Dataset

• Ukutabs.com

• Chords and lyrics are aligned

I've been cheated by you, since I don't know when

So I made up my mind, it must come to an end
The Pilot Task

• Two annotators labeled song segments
• Labeled negative, positive, or neutral
• 929 segments were agreed-upon
• Used embeddings to predict emotion
Results

- Used Logistic Regression classifier
- Chords-and-lyrics model performs best

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baselines</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Majority Class</td>
<td>55.32 %</td>
<td>N/A</td>
</tr>
<tr>
<td>Chord n-grams</td>
<td>57.83 %</td>
<td>.138</td>
</tr>
<tr>
<td>Lyrics n-grams</td>
<td>57.31 %</td>
<td>.194</td>
</tr>
<tr>
<td>Our models</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chords only</td>
<td>59.74 %</td>
<td>.027</td>
</tr>
<tr>
<td>Lyrics only</td>
<td>60.52 %</td>
<td>.012</td>
</tr>
<tr>
<td>Chords &amp; Lyrics</td>
<td>62.28 %</td>
<td>.001</td>
</tr>
</tbody>
</table>
Observations from Embeddings

Embedding with Chords Only

Embedding with Chords and Lyrics
Observations from Embeddings

Bluesy chords are clustered in space
Observations from Embeddings

Enharmonic chords are clustered in space
Observations from Embeddings

Embedding with Chords Only

Embedding with Chords and Lyrics

Chord substitutions are clustered in space
Conclusions/Future Work

• Studying chords and lyrics together may be helpful for music-related studies

• Music theory concepts are supported by our approach

• Is this joint model useful in other areas?

• Can performance be improved with other models?
Thank You

• Special thanks to Morteza Dehghani at USC for encouraging this work

• References:
  • [3] Luong, H Pham, and C D Manning, “Bilingual word representations with monolingual quality in mind,” in Proc. of the 1st Workshop on Vector Space Modeling for Natural Language Processing, 2015, pp. 151–159
The Dataset

- Ukutabs.com
- Chords and lyrics are aligned

I hurt myself today to see if I still feel
I focus on the pain the only thing that's real