Exploring and Monitoring the Social Media Space Using Machine Intelligence

Karl Aberer, EPFL

November 10, 2016
Social Media

• explore perception of a subject, company or product
• identify communities, their opinions and influencers
Example: Migration

- Public perception of the migration issue
- Communication among migrants on their perception of the situation
- Digital tools as enabler for a mobile workforce
State-of-the-art

Social Media Listening

- provides standard business intelligence on basic social media features
  - e.g., how often is “migration” mentioned over time
  - e.g., who are the Twitter users mentioning “migration” with the largest number of followers

No use of machine learning and data mining

- for semantic analysis
- for detecting latent structures

Table 2: Pros and Cons

<table>
<thead>
<tr>
<th></th>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rule-based</td>
<td>• Declarative</td>
<td>• Heuristic</td>
</tr>
<tr>
<td></td>
<td>• Easy to comprehend</td>
<td>• Requires tedious</td>
</tr>
<tr>
<td></td>
<td>• Easy to maintain</td>
<td>manual labor</td>
</tr>
<tr>
<td></td>
<td>• Easy to incorporate domain knowledge</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Easy to trace and fix the cause of errors</td>
<td></td>
</tr>
<tr>
<td>ML-based</td>
<td>• Trainable</td>
<td>• Requires labeled data</td>
</tr>
<tr>
<td></td>
<td>• Adaptable</td>
<td>• Requires retraining</td>
</tr>
<tr>
<td></td>
<td>• Reduces manual effort</td>
<td>for domain adaptation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Requires ML expertise to use or</td>
</tr>
<tr>
<td></td>
<td></td>
<td>maintain</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Opaque</td>
</tr>
</tbody>
</table>

✅ Reasons

- Domain knowledge is important
- Difficult to make experts and machines to work together
Approach

Experts have business and context knowledge and can **choose relevant** structures.

Machines are strong in sifting through masses of data and **detecting hidden** structure.
Challenges

- Enable the (efficient) use of machine learning/data mining tools
- Capture expert domain knowledge
- Filtering of noise
- Coverage of different media and languages
Semantic Analysis
Searching Relevant Data

Every exploration of the Social Space starts with a query, e.g. “diaspora” or “skilled migration”

<table>
<thead>
<tr>
<th>Literals</th>
<th>Name</th>
<th>Categories</th>
<th>Similarity</th>
<th>DF</th>
</tr>
</thead>
<tbody>
<tr>
<td>diaspora</td>
<td>keyword</td>
<td></td>
<td>1.000</td>
<td>483</td>
</tr>
<tr>
<td>african diaspora</td>
<td>keyword</td>
<td></td>
<td>0.776</td>
<td>55</td>
</tr>
<tr>
<td>modernity</td>
<td>keyword</td>
<td></td>
<td>0.709</td>
<td>125</td>
</tr>
<tr>
<td>transnationalism</td>
<td>keyword</td>
<td></td>
<td>0.703</td>
<td>44</td>
</tr>
<tr>
<td>culture of migration</td>
<td>keyword</td>
<td></td>
<td>0.702</td>
<td>45</td>
</tr>
<tr>
<td>human migration</td>
<td>keyword</td>
<td></td>
<td>0.698</td>
<td>132</td>
</tr>
<tr>
<td>cultural heritage</td>
<td>keyword</td>
<td></td>
<td>0.694</td>
<td>62</td>
</tr>
<tr>
<td>global migration</td>
<td>keyword</td>
<td></td>
<td>0.689</td>
<td>107</td>
</tr>
<tr>
<td>contemporary</td>
<td>keyword</td>
<td></td>
<td>0.689</td>
<td>36</td>
</tr>
<tr>
<td>narratives</td>
<td>keyword</td>
<td></td>
<td>0.683</td>
<td>156</td>
</tr>
<tr>
<td>ethnography</td>
<td>keyword</td>
<td></td>
<td>0.681</td>
<td>64</td>
</tr>
<tr>
<td>agsmedics</td>
<td>keyword</td>
<td></td>
<td>0.678</td>
<td>63</td>
</tr>
<tr>
<td>polish migration</td>
<td>keyword</td>
<td></td>
<td>0.677</td>
<td>31</td>
</tr>
<tr>
<td>african migration</td>
<td>keyword</td>
<td></td>
<td>0.677</td>
<td>38</td>
</tr>
<tr>
<td>migration patterns</td>
<td>keyword</td>
<td></td>
<td>0.676</td>
<td>223</td>
</tr>
<tr>
<td>global south</td>
<td>keyword</td>
<td></td>
<td>0.673</td>
<td>105</td>
</tr>
</tbody>
</table>

“Search engine” for related keywords

- NLP processing
- Text mining
- Content clustering
- Deep Learning

✓ Benefits:
  - The system helps to detect variations of the query that you might not have thought about
Syntactic and Semantic Expansion

(near) homonyms

Skilled migration
3k documents

skilled migrant
skilled migrants
highly skilled migrants
high-skilled workers
highly skilled workers
skilled immigration
skilled immigrants
foreign-educated talent
High-Skilled Immigrants
Foreign talent
Talent emigration
Immigrant entrepreneur

Semantically related

<table>
<thead>
<tr>
<th>Literals</th>
<th>Name</th>
<th>Categories</th>
<th>Similarity</th>
<th>DF</th>
</tr>
</thead>
<tbody>
<tr>
<td>points test</td>
<td>keyword</td>
<td></td>
<td>0.939</td>
<td>42</td>
</tr>
<tr>
<td>h-1b employer</td>
<td>keyword</td>
<td></td>
<td>0.938</td>
<td>35</td>
</tr>
<tr>
<td>visa options</td>
<td>keyword</td>
<td></td>
<td>0.929</td>
<td>36</td>
</tr>
<tr>
<td>non-immigrant visa</td>
<td>keyword</td>
<td></td>
<td>0.924</td>
<td>47</td>
</tr>
<tr>
<td>job offers</td>
<td>keyword</td>
<td></td>
<td>0.923</td>
<td>46</td>
</tr>
<tr>
<td>long-term immigration</td>
<td>keyword</td>
<td></td>
<td>0.922</td>
<td>34</td>
</tr>
<tr>
<td>skill shortages</td>
<td>keyword</td>
<td></td>
<td>0.922</td>
<td>34</td>
</tr>
<tr>
<td>entry clearance visas</td>
<td>keyword</td>
<td></td>
<td>0.921</td>
<td>33</td>
</tr>
<tr>
<td>based system</td>
<td>keyword</td>
<td></td>
<td>0.921</td>
<td>37</td>
</tr>
<tr>
<td>foreign-educated talent</td>
<td>keyword</td>
<td></td>
<td>0.938</td>
<td>32</td>
</tr>
</tbody>
</table>

✅ Benefits:
- Larger coverage
- More related topics captured
- Indirect references exploited
Organizing Terminology

Detecting hidden dimensions in the term space

We see

- A clear distinction between positive and negative terms
- Distinction between natural and artificial ingredients

We can embed entities into this space

Proposed analysis: map the main topics related to the migration discussion and link to the countries/actors/media
Detecting Latent Structures
Organizing Documents

The system automatically organizes the document collection into topical collections, e.g., on Bel brands
- Automatic structuring of the collection according to themes
- Elimination of Noise

Benefits:
- Identifying key topics
- Efficient removal of non-relevant content
- Capturing topic related terminology

Actually, this is about data migration
Analyzing Communities

Benefits:

- Identifying communities and their influencers
- Efficient removal of non-relevant content
- Identifying key interests of communities
- Capturing community terminology

Proposed analysis: identify key communities/main drivers such as media/their positions
Identify potentially sites relevant to migrants to identify migrant communities
Analyzing Influencers
The platform: SEMPI
An automated platform to integrate semantic analysis and structure discovery with expert interaction

- relevant social media content
- most important concepts & topics
- community discussions forming around certain topics
- key influencers of those discussions (academics, activists, politicians)
- specific issues (statements) being discussed
- general public perception / sentiment

The platform has been successfully used for projects on public relations, marketing, humanitarian action.
Workflow

Query Generation

Analysis: Terminology & Ontology

Dashboard Exploration

Analysis: Topics and Influencers
Demo: resulting dashboard
Outlook

Discovery of correlations between Social Media data and real-world data

- Politics
- Marketing campaigns
- Health
- Scientific publications

For detailed information and demos contact: karl.aberer@epfl.ch