Cross-Lingual Analyses of Parliamentary Debates

Federico Nanni, Goran Glavaš and Simone Paolo Ponzetto
Text Scaling

Current issues:

1. Topic-based analyses (e.g., extract ideological statements)
2. Cross-lingual analyses
3. Evaluation of unsupervised scaling methods
Text Scaling

Current issues:

1. Topic-based analyses (e.g., extract ideological statements)
2. Cross-lingual analyses
3. Evaluation of unsupervised scaling methods
Cross-Lingual Mapping of Word Embeddings

Linear translation matrix model to map word embeddings across languages.
# Cross-Lingual Scaling (speeches from EP 5th leg)

<table>
<thead>
<tr>
<th></th>
<th>Monolingual</th>
<th></th>
<th></th>
<th>Cross-lingual</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PA</td>
<td>(r_P)</td>
<td>(r_S)</td>
<td>PA</td>
<td>(r_P)</td>
<td>(r_S)</td>
</tr>
<tr>
<td>Random</td>
<td>49.7</td>
<td>-.03</td>
<td>.00</td>
<td>49.7</td>
<td>-.03</td>
<td>.00</td>
</tr>
<tr>
<td>Wordfish</td>
<td>55.0</td>
<td>.21</td>
<td>.20</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>AL-HFLP</td>
<td>61.3</td>
<td>.35</td>
<td>.31</td>
<td>57.3</td>
<td>.20</td>
<td>.25</td>
</tr>
<tr>
<td>AGG-HFLP</td>
<td>67.0</td>
<td>.53</td>
<td>.46</td>
<td>63.3</td>
<td>.34</td>
<td>.39</td>
</tr>
</tbody>
</table>

TopFish: Cross-Lingual Classification and Scaling

1) Mono and Cross-lingual Classification (based on Conv. Neural Net)
2) Mono and Cross-lingual Scaling
3) Wordfish (our python implementation)

https://github.com/codogogo/topfish
Issues

1. Our tool **does not work** when the texts come from national parliaments
Issues

1. Our tool **does not work** when the texts come from national parliaments
2. We did not really approach this project in a **“interdisciplinary”** way
## Issues

1. Our tool **does not work** when the texts come from national parliaments
2. We did not really approach this project in a “interdisciplinary” way
3. Our work has been oriented towards making a **NLP** contribution
Larger Issue

Absence of an interdisciplinary venue, relevant for both communities, for publishing works that involve NLP and political science research.
Thanks

Federico Nanni
Data and Web Science Group
University of Mannheim
federico@informatik.uni-mannheim.de