Machine translation without the Rosetta stone

Yova Kementchedjieva
Overview

• The task: Bilingual Dictionary Induction

• Why: for Machine Translation purposes

• Approaches to it
  • Supervised
  • Semi-supervised
  • Unsupervised

• Augmentations of an existing semi-supervised approach
The task: Bilingual Dictionary Induction (BDI)
Why: for Machine Translation purposes

Phrase-based & Syntax-based translation

Word-based translation
Word vectors
Supervised BDI

Seed dictionary
- cat: gatto
- car: auto
- deep: profondo
- feline: felino

New bigger dictionary:
- cat
- car
- deep
- feline
- recognize
- anchor
- arguing
- gatto
- auto
- profondo
- felino
- reconocen
- anclas
- discutiendo
- ...

Visuals from Lample et al. (2017) with minor modifications
Can we learn better mappings by abstracting to a third latent space?

Can we learn better mappings with multi-lingual, rather than bilingual, training in the (semi-)supervised setting? Yes and Yes.
Procrustes Analysis (PA)

\[
\arg \min_W ||WX' - Z'||^2
\]

\[
Z'TX = U\Sigma V^T
W = VU^T
\]

Generalized Procrustes Analysis (GPA)

\[
\arg \min_{\{w_1, \ldots, w_k\}} \sum_{i<j} ||w_iX_i' - W_jX_j'||^2
\]

\[
G = k^{-1} \sum_{i=1}^k X_i'W_i
\]

\[
G^TX_i' = U\Sigma V^T
W_i = VU^T \text{ for } i \text{ in } 1 \ldots k
\]
## Performance of GPA compared to simple PA

<table>
<thead>
<tr>
<th></th>
<th>GS dictionary</th>
<th>numerals</th>
<th>homographs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Artetxe et al. 2017</td>
<td>39.7</td>
<td>39.4</td>
<td>-</td>
</tr>
<tr>
<td>Smith et al. 2017</td>
<td>43.1</td>
<td>-</td>
<td>34.3</td>
</tr>
<tr>
<td>Artetxe et al. 2018</td>
<td>45.27</td>
<td>39.4</td>
<td>38.33</td>
</tr>
<tr>
<td>Lample et al. 2017 (PA)</td>
<td>44.9</td>
<td>01.13</td>
<td>45.47</td>
</tr>
<tr>
<td>GPA</td>
<td><strong>45.33</strong></td>
<td><strong>45.93</strong></td>
<td><strong>45.80</strong></td>
</tr>
</tbody>
</table>

Table 1: P@1 for English to Italian, using the WaCky monolingual embeddings (Barone, 2016) and Dinu’s et al. (2015) dictionaries.