Multilingual Modal Sense Classification Using a Convolutional Neural Network

Ana Marasović, Anette Frank
Research Training Group AIPHES, Department of Computational Linguistics, Heidelberg University

Overview

- Modal verbs (MVs) are ambiguous between:
  - epistemic sense (possibility)
    - He could be at home.
  - deontic sense (permission/obligation)
    - You can enter now.
  - dynamic sense (capability)
    - Only John can solve this problem.

- MVs are used to implicitly express sentiment
  - Refugees may (de) not (are forbidden to) cross the borders.
  - Writer has negative sentiment towards refugees crossing the borders.

- MSC is special case of word sense disambiguation (WSD)
  - MVs have restricted sense inventory
  - MVs act as operators which take a full proposition as an argument

CNN for Modelling Sentences

![Convolutional neural network (CNN), Kim (2014).](image)

### Experimental Setup

- Corpora
  - MPQA: R&R’s small-scaled manually annotated dataset with strong sense bias
  - EPOSS, EPOSSH: a subset of EuroParl & OpenSubtitles corpora, heuristically tagged via the cross-lingual sense projection method of Z+.
  - CNN-E, CNN-E\textsubscript{D}: 5-fold CV.
  - CNN-G: train and test data from EPOSS

- Baselines
  - random baseline BL\textsubscript{rand}
  - majority baseline BL\textsubscript{maj}
  - MaxEnt classifier from Z+.
  - one-layer neural network NN

### WSD Results

<table>
<thead>
<tr>
<th>Features</th>
<th>Sense</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>past reading of the emb. verb</td>
<td>ep</td>
<td>you must have been out last night</td>
</tr>
<tr>
<td>non-past reading of the emb. verb</td>
<td>de</td>
<td>we must take further efforts</td>
</tr>
<tr>
<td>relative reading of the emb. verb</td>
<td>ep</td>
<td>you must think me a perfect fool</td>
</tr>
<tr>
<td>possible construction</td>
<td>de</td>
<td>actual steps must be taken</td>
</tr>
<tr>
<td>negation</td>
<td>de</td>
<td>we must not fear</td>
</tr>
<tr>
<td>domain specific vocabulary</td>
<td>de</td>
<td>European parliament, present regulation, foreign policy</td>
</tr>
<tr>
<td>telic clauses</td>
<td>de</td>
<td>to address these problems, to prevent both forum</td>
</tr>
<tr>
<td>discourse markers</td>
<td>de</td>
<td>but, and (then)</td>
</tr>
</tbody>
</table>

### MSC Results

<table>
<thead>
<tr>
<th>Feature Detector</th>
<th>Balanced</th>
<th>Unbalanced</th>
</tr>
</thead>
<tbody>
<tr>
<td>BL\textsubscript{rand}</td>
<td>33.33</td>
<td>33.33</td>
</tr>
<tr>
<td>MaxEnt</td>
<td>59.64</td>
<td>61.25</td>
</tr>
<tr>
<td>CNN-E</td>
<td>66.57</td>
<td>70.29</td>
</tr>
<tr>
<td>CNN-G</td>
<td>70.87</td>
<td>66.57</td>
</tr>
</tbody>
</table>

### CNN for WSD

#### Data: SensEval-3 lexical sample dataset

- Baseline: R&S’s sense-specific embeddings
- w: ambiguous word with k senses!
- \( c_1, \ldots, c_k \): centroid — sum of all \( w_2v \) vectors of words in the sentence
- \( S\text{-cosine} \) = \( \cos(c_1, …, c_k, w_1, …, w_n) \)
- \( S\text{-prod} \) = \( \prod_{i=1}^{n} s_i \cdot s_j \)
- \( S\text{-raw} \) = \( \sum_{i=1}^{n} s_i \cdot s_j \)

### References


### Acknowledgements

This research was supported by the German Research Foundation as part of the Research Training Group “Adaptive Preparation of Information from Heterogeneous Sources” (AIPHES) under grant No. GRK 1994/1.