An Iterative Approach for Unsupervised Most Frequent Sense Detection using WordNet and Word Embeddings

Kevin Patel and Pushpak Bhattacharyya
Presented by: Ritesh Panjwani

January 11, 2018
Word Sense Disambiguation (WSD): one of the relatively hard problems in NLP
- Both supervised and unsupervised ML explored in literature

Most Frequent Sense (MFS) baseline: strong baseline for WSD
- Given a WSD problem instance, simply assign the most frequent sense of that word

Ignores context

Really strong results
- Due to skew in sense distribution of data

Computing MFS:
- Trivial for sense-annotated corpora, which is not available in large amounts.
- Need to learn from raw data
Problem Statement

Given a raw corpus, estimate most frequent sense of different words in that corpus.
Problem Statement

Given a raw corpus, estimate most frequent sense of different words in that corpus

- Bhingardive et al. (2015a) showed that pretrained word embeddings can be used to compute most frequent sense
- Our work further strengthens the claim by Bhingardive et al. (2015a) that word embeddings indeed capture most frequent sense
- Our approach outperforms others at the task of MFS extraction
- To compute MFS using our approach:
  1. Train word embeddings on the raw corpus.
  2. Apply our approach on the trained word embeddings.
Intuition

- Strive for consistency in assignment of senses to maintain semantic congruity
- Example:
  - If *cricket* and *bat* co-occur a lot, then *cricket* taking *insect* sense and *bat* taking reptile sense is less likely
Intuition

- Strive for consistency in assignment of senses to maintain semantic congruity
- Example:
  - If *cricket* and *bat* co-occur a lot, then *cricket* taking *insect* sense and *bat* taking reptile sense is less likely
  - If *cricket* and *bat* co-occur a lot, and *cricket*’s MFS is *sports*, then *bat* taking reptile sense is extremely unlikely
- Key point: solve easy words, then use them for difficult words
  - In other words, iterate over degree of polysemy from 2 onward
(Buitelaar and Sacaleanu, 2001) present an approach for domain specific sense assignment.
- Rank GermaNet synsets based on the co-occurrence in domain corpora.

(Lapata and Brew, 2004) acquire predominant sense of verbs.
- Use Levin’s classes as their sense inventory.

(McCarthy et al., 2007) use a thesaurus and the WordNet similarities to find predominant noun senses automatically.

(Bhingardive et al., 2015b) exploit word embeddings trained on untagged corpora to compute the most frequent sense.
Algorithm

Word
Algorithm

- Word
  - Sense 1
  - Sense 2
  - Sense 3

- Tagged Word 1
- Tagged Word 2
- Tagged Word k
Algorithm

Word

Sense 1
Tagged Word 1
w1s1

Sense 2
Tagged Word 2
w2s1
w2s2
w2s3

Sense 3
Tagged Word k
wks1
wks2
wks3
Algorithm

- $w_i s_j$ is vote for $s_j$ due to $w_i$
- Two components
  - Wordnet similarity between $\text{mfs}(w_i)$ and $s_j$
  - Embedding space similarity between $w_i$ and current word
Algorithm

Kevin Patel

Iterative Unsupervised MFS
Parameters

- **K**: The number of nearest neighbors who will vote.
- **WordNet Similarity measure \( (s_i) \)**: Average of normalized Wu Palmer and Lin similarity
- **Vector space similarity measure \( (w_i) \)**: Dot product
Datasets:

- **SemCor**: Sense-annotated corpus, annotated with Princeton WordNet 3.0 senses using WordNet 1.7 to WordNet 3.0 mapping by Rada Mihalcea
- **Senseval 2**: Sense-annotated corpus, annotated with Princeton WordNet 3.0 senses
- **Senseval 3**: Sense-annotated corpus, annotated with Princeton WordNet 3.0 senses

Two setups:

- Evaluating MFS as solution for WSD
- Evaluating MFS as a classification task
### MFS as solution for WSD

#### Accuracy of WSD using MFS (Nouns)

<table>
<thead>
<tr>
<th>Method</th>
<th>Senseval2</th>
<th>Senseval3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bhingardive (reported in Bhingardive et al., 2015b)</td>
<td>52.34</td>
<td>43.28</td>
</tr>
<tr>
<td>Semcor (reported in Bhingardive et al., 2015b)</td>
<td>59.88</td>
<td>65.72</td>
</tr>
<tr>
<td>Bhingardive (optimal)</td>
<td>48.27</td>
<td>36.67</td>
</tr>
<tr>
<td>Iterative</td>
<td>63.2</td>
<td>56.72</td>
</tr>
<tr>
<td>SemCor</td>
<td>67.61</td>
<td>71.06</td>
</tr>
</tbody>
</table>
MFS as solution for WSD (contd.)

<table>
<thead>
<tr>
<th>Method</th>
<th>Senseval2</th>
<th>Senseval3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bhingardive (reported)</td>
<td>37.79</td>
<td>26.79</td>
</tr>
<tr>
<td>Bhingardive (optimal)</td>
<td>43.51</td>
<td>33.78</td>
</tr>
<tr>
<td>Iterative</td>
<td>48.1</td>
<td>40.4</td>
</tr>
<tr>
<td>SemCor</td>
<td>60.03</td>
<td>60.98</td>
</tr>
</tbody>
</table>

Accuracy of WSD using MFS (All Parts of Speech)
### MFS as classification task

<table>
<thead>
<tr>
<th>Method</th>
<th>Nouns</th>
<th>Adjectives</th>
<th>Adverbs</th>
<th>Verbs</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bhingardive</td>
<td>43.93</td>
<td>81.79</td>
<td>46.55</td>
<td>37.84</td>
<td>58.75</td>
</tr>
<tr>
<td>Iterative</td>
<td>48.27</td>
<td>80.77</td>
<td>46.55</td>
<td>44.32</td>
<td>61.07</td>
</tr>
</tbody>
</table>

Percentage match between predicted MFS and WFS
MFS as classification task (contd.)

<table>
<thead>
<tr>
<th></th>
<th>Nouns (49.20)</th>
<th>Verbs (26.44)</th>
<th>Adjectives (19.22)</th>
<th>Adverbs (5.14)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bhingardive</td>
<td>29.18</td>
<td>25.57</td>
<td>26.00</td>
<td>33.50</td>
<td>27.83</td>
</tr>
<tr>
<td>Iterative</td>
<td>35.46</td>
<td>31.90</td>
<td>30.43</td>
<td>47.78</td>
<td>34.19</td>
</tr>
</tbody>
</table>

Percentage match between predicted MFS and true SemCor MFS. Note that numbers in column headers indicate what percent of total words belong to that part of speech.
Better than Bhingardive et al. (2015a); not able to beat SemCor and WFS.

There are words for which WFS doesn’t give proper dominant sense. Consider the following examples:

- **tiger** - an audacious person
- **life** - characteristic state or mode of living (social life, city life, real life)
- **option** - right to buy or sell property at an agreed price
- **flavor** - general atmosphere of place or situation
- **season** - period of year marked by special events

Tagged words ranking very low to make a significant impact. For example:

- While detecting MFS for a bisemous word, the first monosemous neighbour actually ranks 1101
- *i.e.* a 1000 polysemous words are closer than this monosemous word.
- Monosemous word may not be the one who can influence the MFS.
Conclusion and Future Work

- Proposed an iterative approach for unsupervised most frequent sense detection using word embeddings
- Similar trends, yet better overall results from Bhingardive et al. (2015a)
- Strengthens the claim that word embeddings do indeed capture most frequent sense.
- Future Work
  - No language specific restrictions, so apply approach to other languages

References


Thank You

Questions?
For more details, write to: kevin.patel@cse.iitb.ac.in