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

Outline

Introduction

- 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

Problem Statement

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

Problem Statement

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

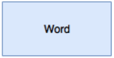
- Strive for consistency in assignment of senses to maintain semantic congruity
- Example:
  - If *cricket* and *bat* co-occur a lot, then *cricket* taking *insect* sense and *bat* taking reptile sense is less likely
  - If *cricket* and *bat* co-occur a lot, and *cricket*'s MFS is *sports*, then *bat*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

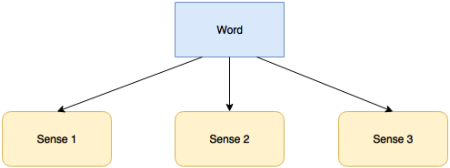
Related Work

- (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.

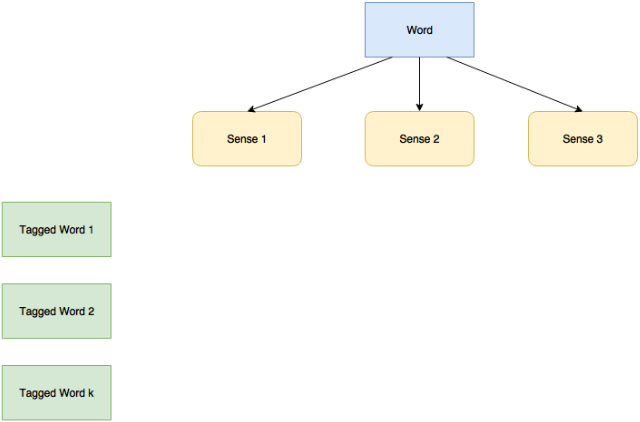
Algorithm



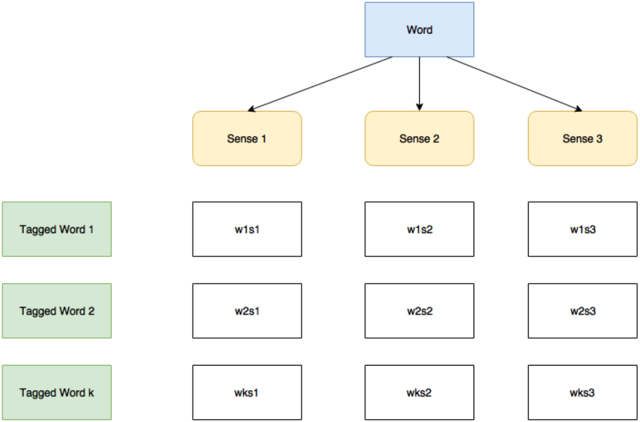
Algorithm



Algorithm


Word
Sense 1
Sense 2
Sense 3
Tagged Word 1
Tagged Word 2
Tagged Word k

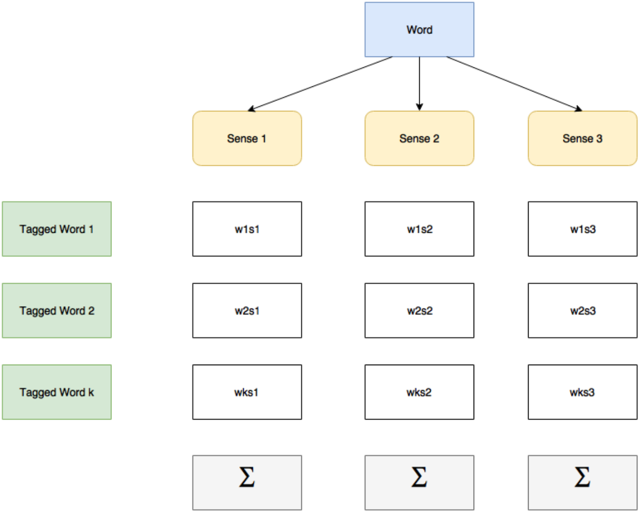
Algorithm


Word
Sense 1
Sense 2
Sense 3
Tagged Word 1
w1s1
w1s2
w1s3
Tagged Word 2
w2s1
w2s2
w2s3
Tagged Word k
wks1
wks2
wks3

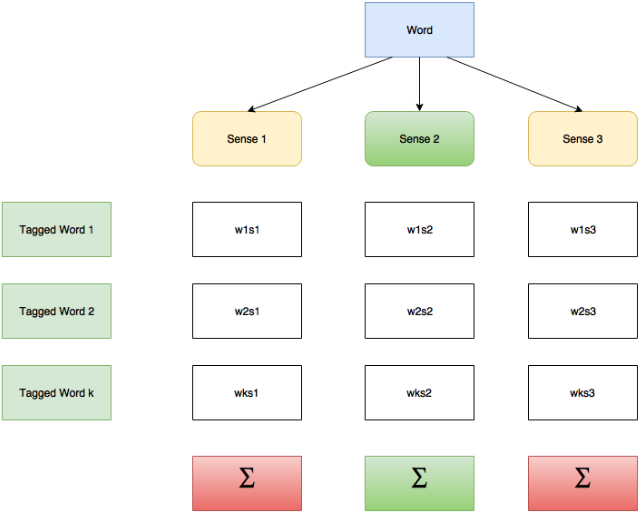
Algorithm

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

Algorithm


Word
Sense 1
Sense 2
Sense 3
Tagged Word 1
Tagged Word 2
Tagged Word k
w1s1
w1s2
w1s3
w2s1
w2s2
w2s3
wks1
wks2
wks3
Σ
Σ
Σ

Algorithm


Word
Sense 1
Sense 2
Sense 3
Tagged Word 1
w1s1
w1s2
w1s3
Tagged Word 2
w2s1
w2s2
w2s3
Tagged Word k
wks1
wks2
wks3
Σ
Σ
Σ

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

Evaluation

- Datasets:
  - SemCor: Sense-annotated corpus, annotated with Princeton WordNet 3.0 senses using WordNet 1.7 to WordNet3.0 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

Method	Senseval2	Senseval3
Bhingardive (reported in (Bhingardive et al., 2015b))	52.34	43.28
Semcor (reported in (Bhingardive et al., 2015b))	59.88	65.72
Bhingardive (optimal)	48.27	36.67
Iterative	63.2	56.72
SemCor	67.61	71.06

Accuracy of WSD using MFS (Nouns)

MFS as solution for WSD (contd.)

Method	Senseval2	Senseval3
Bhingardive(reported)	37.79	26.79
Bhingardive(optimal)	43.51	33.78
Iterative	48.1	40.4
SemCor	60.03	60.98

Accuracy of WSD using MFS (All Parts of Speech)

MFS as classification task

Method	Nouns	Adjectives	Adverbs	Verbs	Total
Bhingardive	43.93	81.79	46.55	37.84	58.75
Iterative	48.27	80.77	46.55	44.32	61.07

Percentage match between predicted MFS and WFS

MFS as classification task (contd.)

	Nouns (49.20)	**Verbs (26.44)**	**Adjectives (19.22)**	**Adverbs (5.14)**	**Total**
Bhingardive	29.18	25.57	26.00	33.50	27.83
Iterative	35.46	31.90	30.43	47.78	34.19

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

Analysis

- 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

Bhingardive, S., Singh, D., V, R., Redkar, H., and Bhattacharyya, P. (2015a). Unsupervised most frequent sense detection using word embeddings. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1238–1243, Denver, Colorado. Association for Computational Linguistics.

Bhingardive, S., Singh, D., V, R., Redkar, H., and Bhattacharyya, P. (2015b). Unsupervised most frequent sense detection using word embeddings. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1238–1243, Denver, Colorado. Association for Computational Linguistics.

References

Buitelaar, P. and Sacaleanu, B. (2001). Ranking and selecting synsets by domain relevance. In *Proceedings of the WordNet and Other Lexical Resources: Applications, Extensions and Customizations. NAACL Workshop*, Pittsburgh. o.A.

Lapata, M. and Brew, C. (2004). Verb class disambiguation using informative priors. *Computational Linguistics*, 30(1):45–73.

McCarthy, D., Koeling, R., Weeds, J., and Carroll, J. (2007). Unsupervised acquisition of predominant word senses. *Computational Linguistics*, 33(4):553–590.

Thank You

Questions?

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