Extending Universal Semantic Tagging

Da Huo, Gerard de Melo

Department of Computer Science. Rutgers, The State University Of New Jersey. New Brunswick. NJ 08854

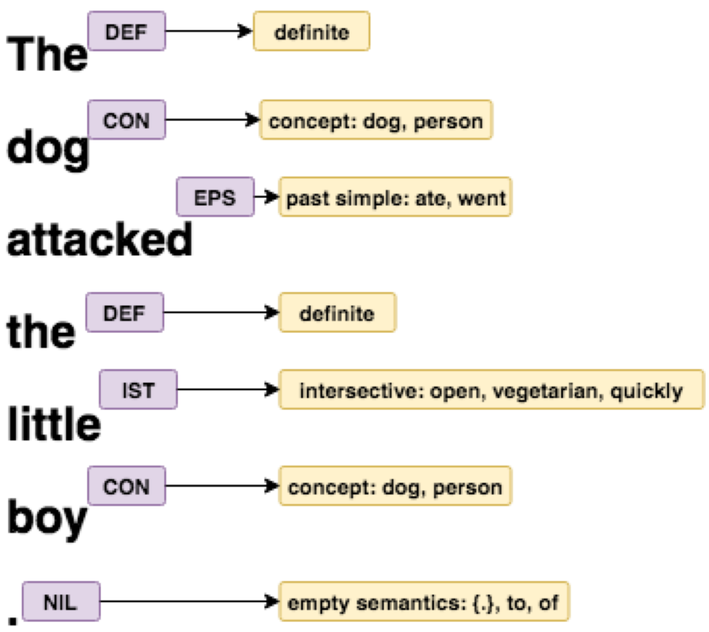
RUTGERS
Department of
Computer Science

Abstract

When it comes to the interaction between humans and computers, natural language plays an increasingly important role, e.g. in personal assistants such as Alexa. In this project, we aim to analyze natural language and infer semantic properties of words. The recently proposed notion of semantic tagging (Bjerva et al. 2016) aims at labeling words with tags that reveal their properties and are useful in semantic tasks. While a training corpus exists, its coverage is limited. We propose a way to automatically extend the coverage by drawing on large-scale word representation data to derive a large new Semantic Tag lexicon. Our experiments show that we can infer semantic tags for unseen words with high accuracy.

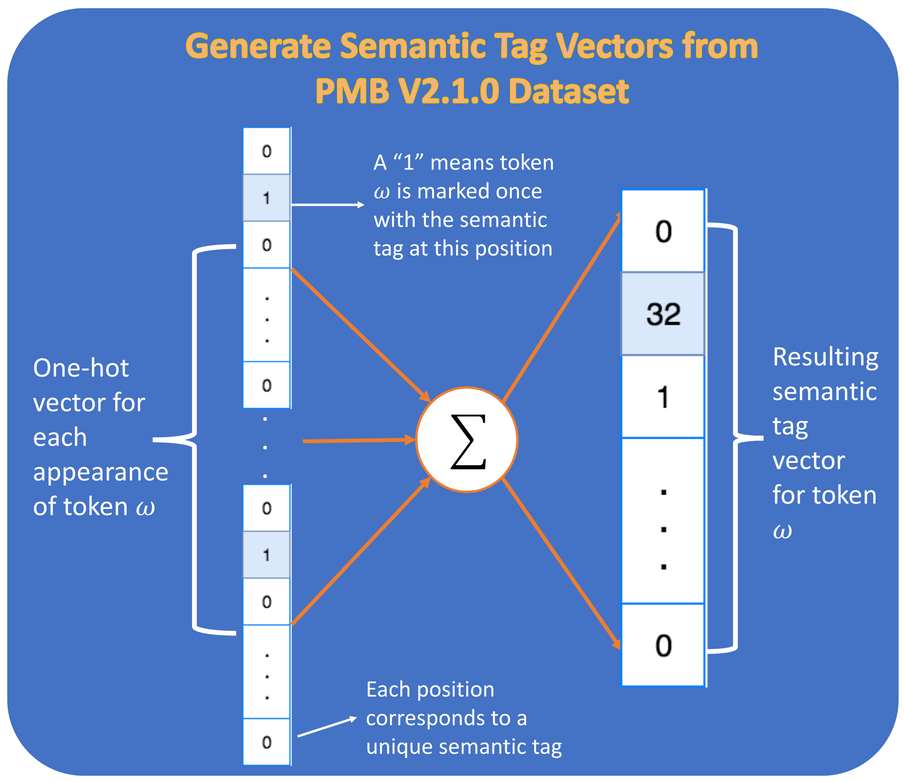
Motivation

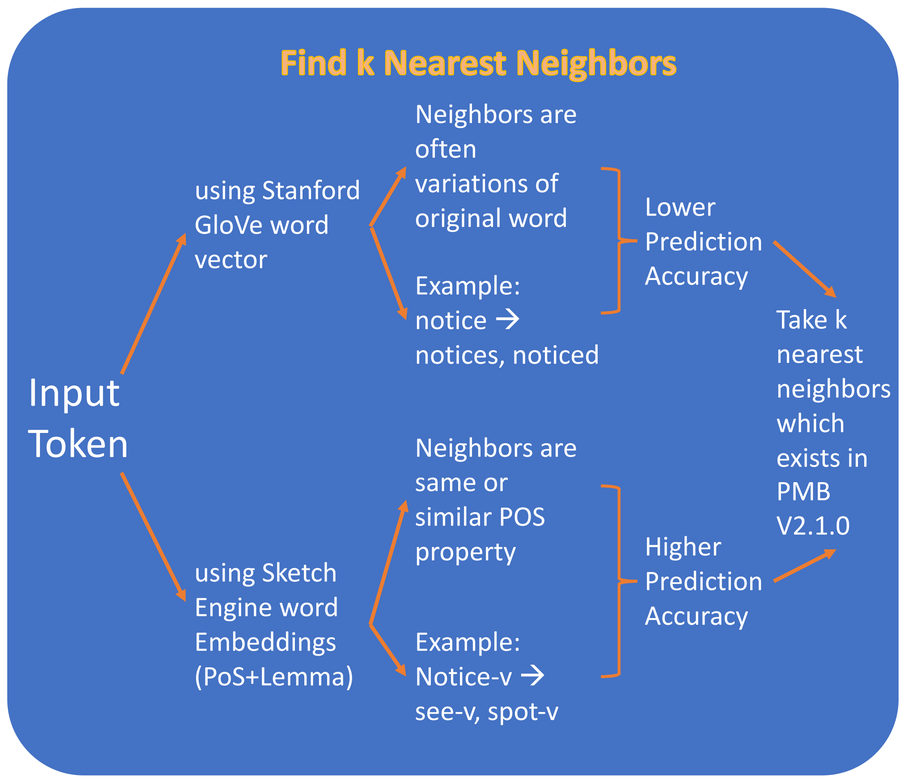
- Previously, Part-of-Speech tagging and Named Entity classes do contribute to determine lexical semantics but they are not sufficiently informative.
- Semantic tagging incorporate semantic virtues of these two tasks and fill gaps in semantic modeling by adding new categories.
- An example of parsing the sentence **The dog attacked the little boy.**

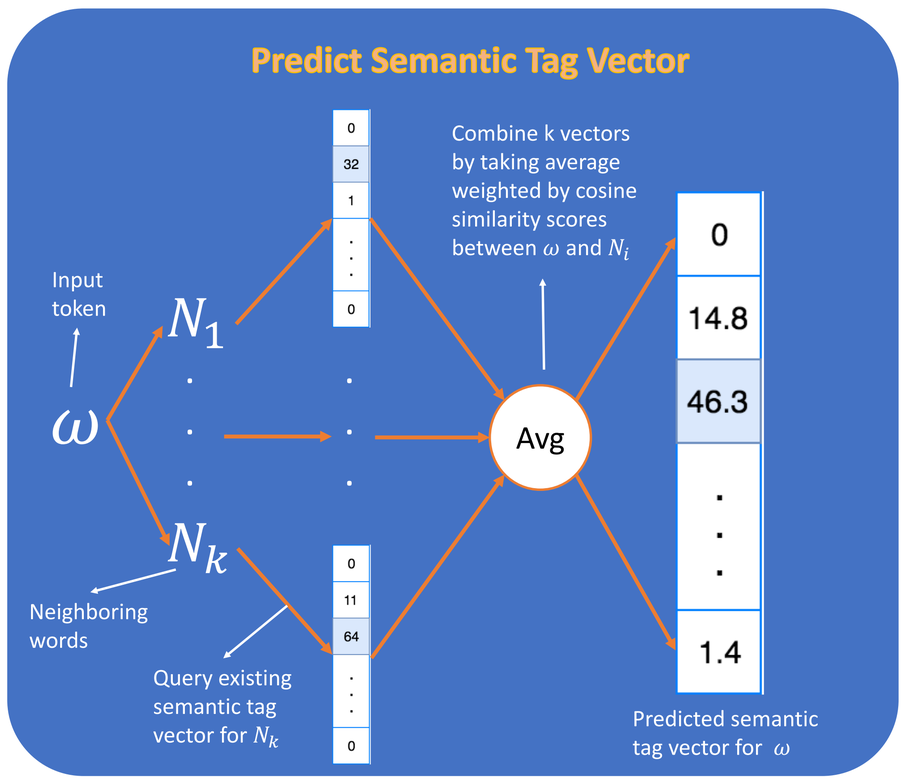


- Given the usefulness of semantic tags, it is helpful to create a large cross-lingual semantic tagging dataset.

Methodology







Results

After extensive testing, our prediction mechanism achieved around 65%-70% accuracy if using Stanford GloVe word vectors. The results is shown below:

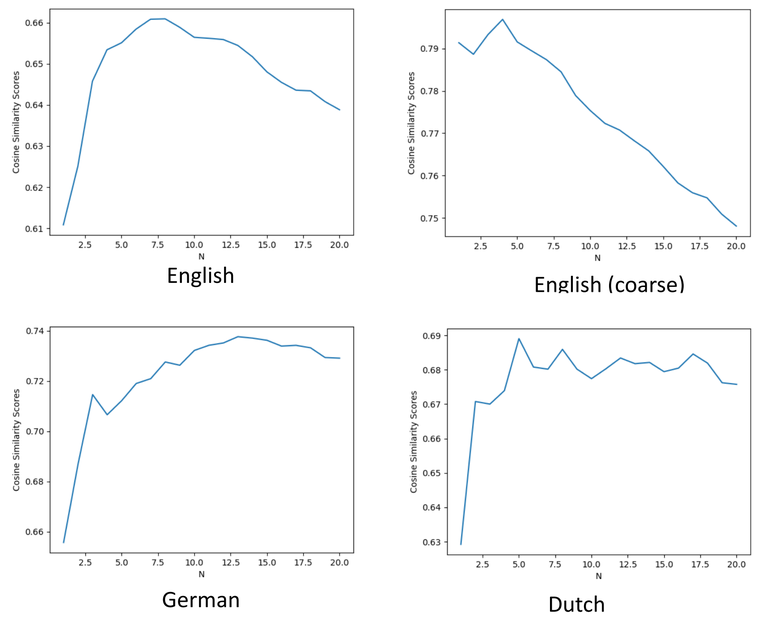


Figure 1: Cosine similarity score vs. k for Form-based Prediction

We achieved more than 78% of prediction accuracy if we take Part-of-Speech into consideration when taking neighbors

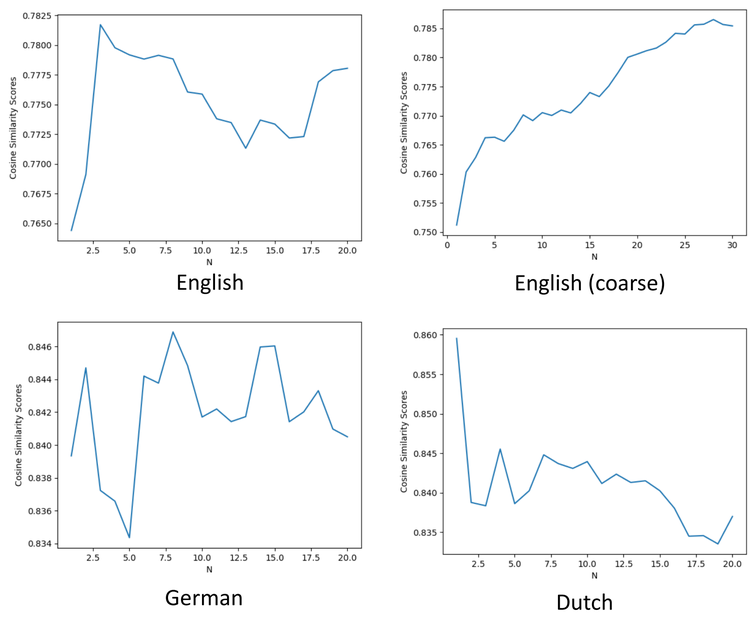


Figure 2: Cosine similarity score vs. k for POS-based Prediction

Future Work

- Generating semantic tag datasets for over 370 languages and make it freely available for download.
- Keep our dataset up-to-date as Parallel Meaning Bank is constantly updating their data

References

[1] Lasha Abzianidze and Johan Bos. 2017. Towards uni- versal semantic tagging. In *IWCS 2017 — 12th Inter- national Conference on Computational Semantics*.

[2] PMB (Parallel Meaning Bank) from: http://pmb.let.rug.nl/

RUTGERS