Expanding Semantic Tag-Based Representation Learning

Da Huo, Gerard de Melo
Rutgers University. New Brunswick, NJ, USA
Contact: gdm@demelo.org

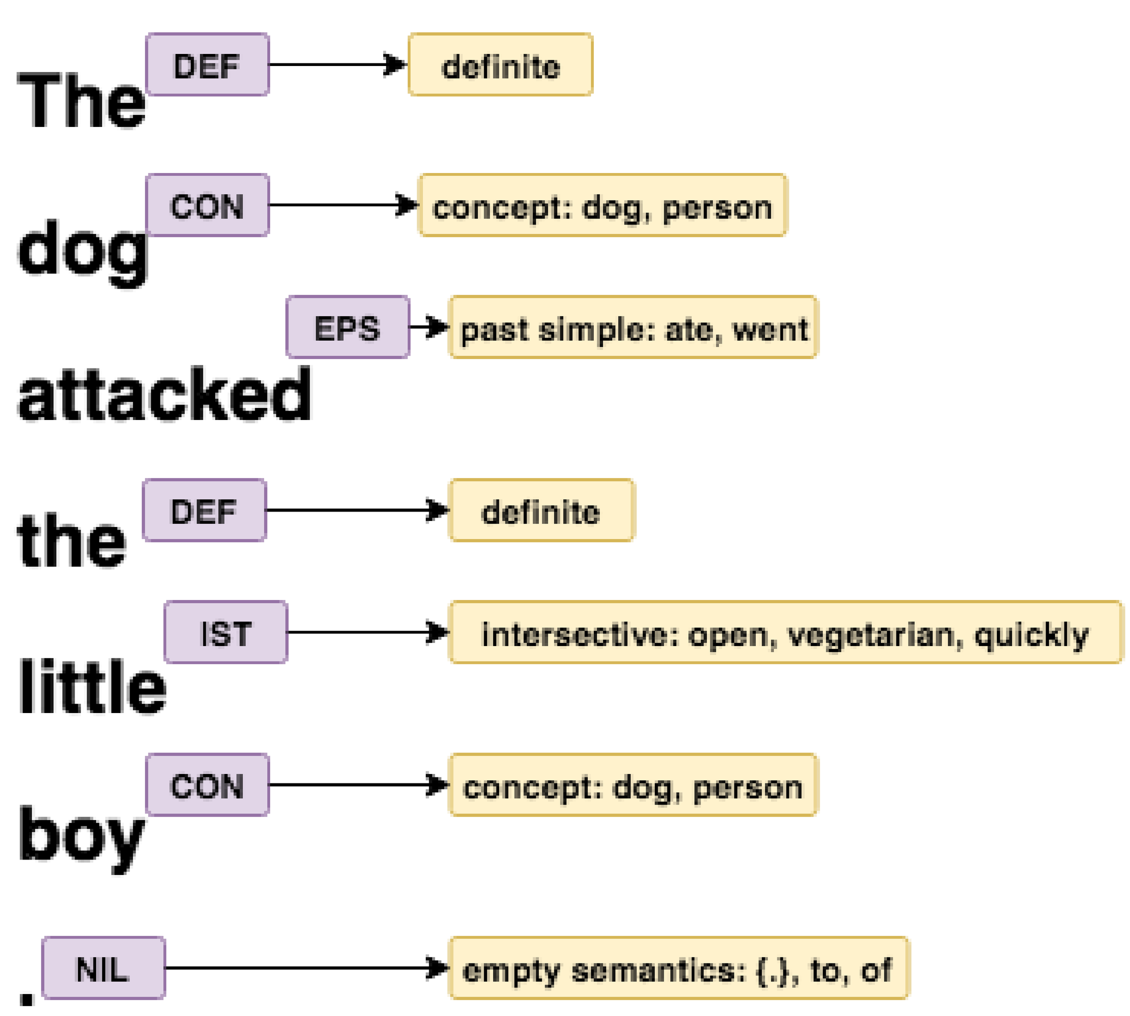
Representation Learning

In machine learning, vector representations are now widely appreciated for their ability to impart background information about items into a learning process. For example, the machine learning algorithm may never have seen the string *Havel* in its training data, but if a vector representation for *Havel* reveals that it is similar to other European rivers such as *Elbe* and *Danube*, the machine learning algorithm may be able to treat it appropriately.

Semantic Tag Vectors

- Most word vector representations contain numbers that are not human-interpretable. We consider vectors that explicitly store information about a word's semantic meaning properties.
- Semantic tagging (Bjerva et al. 2016) is a recently proposed scheme for labeling the properties of words. E.g.:

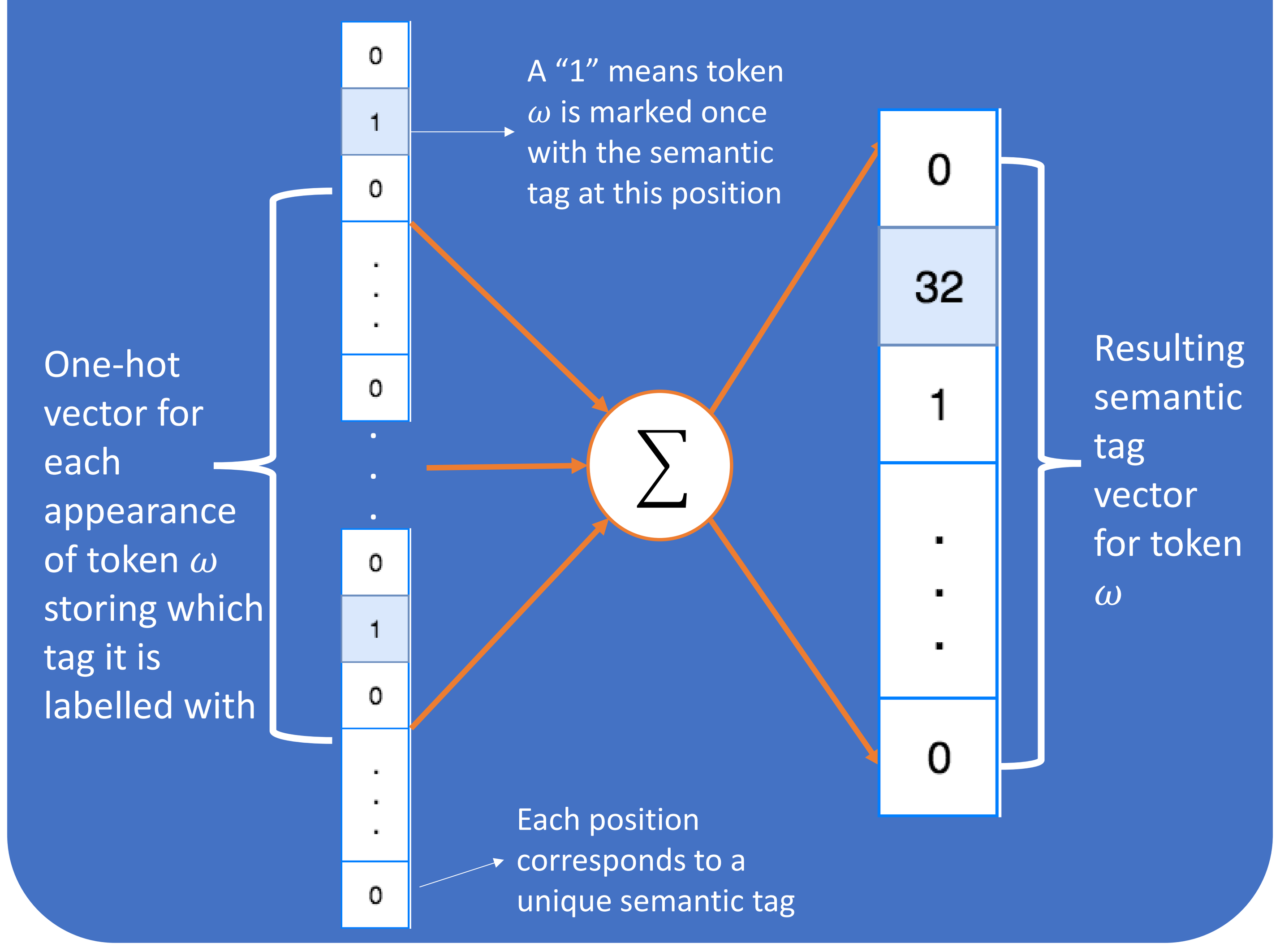
The dog attacked the little boy.



- We create a vector for a word, in which each dimension captures how often we saw it labeled with a specific tag in a human-labeled text collection.
- **Our Goal:** Given these vectors, which we only have for a small set of words for which we have human-provided tags, **automatically create similar semantic tag vector representations for new words.**

Methodology

Generating Semantic Tag Vectors from Labeled Text Collection (PMB Dataset)



Find k Nearest Neighbors

Input Word

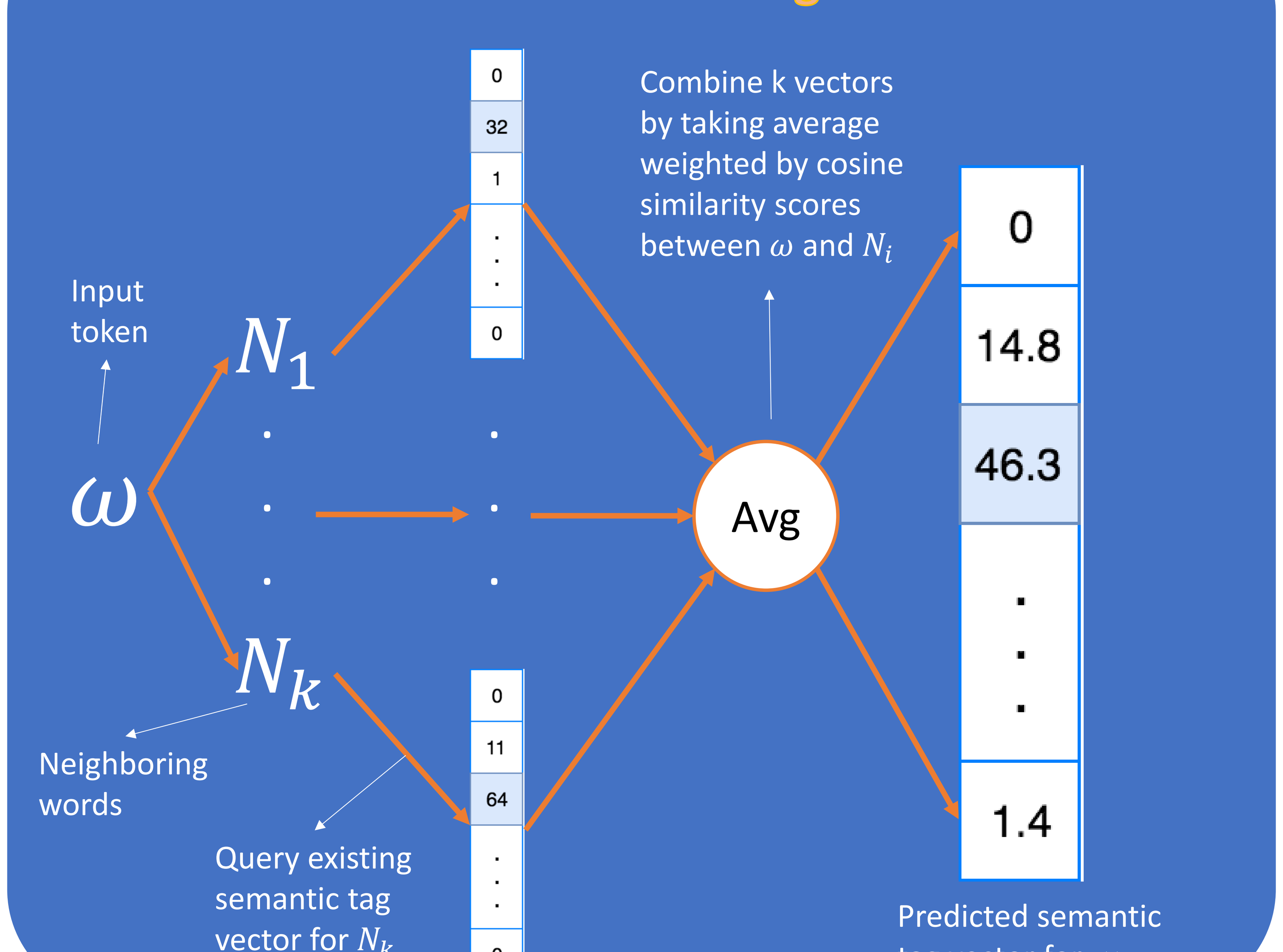
Find most similar words using Sketch Engine word Embeddings.

But importantly, only consider entries with same part-of-speech (e.g. verb vs. noun)

Take k nearest neighbors which we also have Semantic Tags

Example:
Notice-v → see-v, spot-v

Predict Semantic Tag Vector



Note: We have an additional mechanism for propagation across languages (i.e., English to German, etc.)

Results

After extensive testing, our prediction mechanism achieved around 65%-70% accuracy if just using Stanford GloVe word vectors (ignoring part-of-speech). Results:

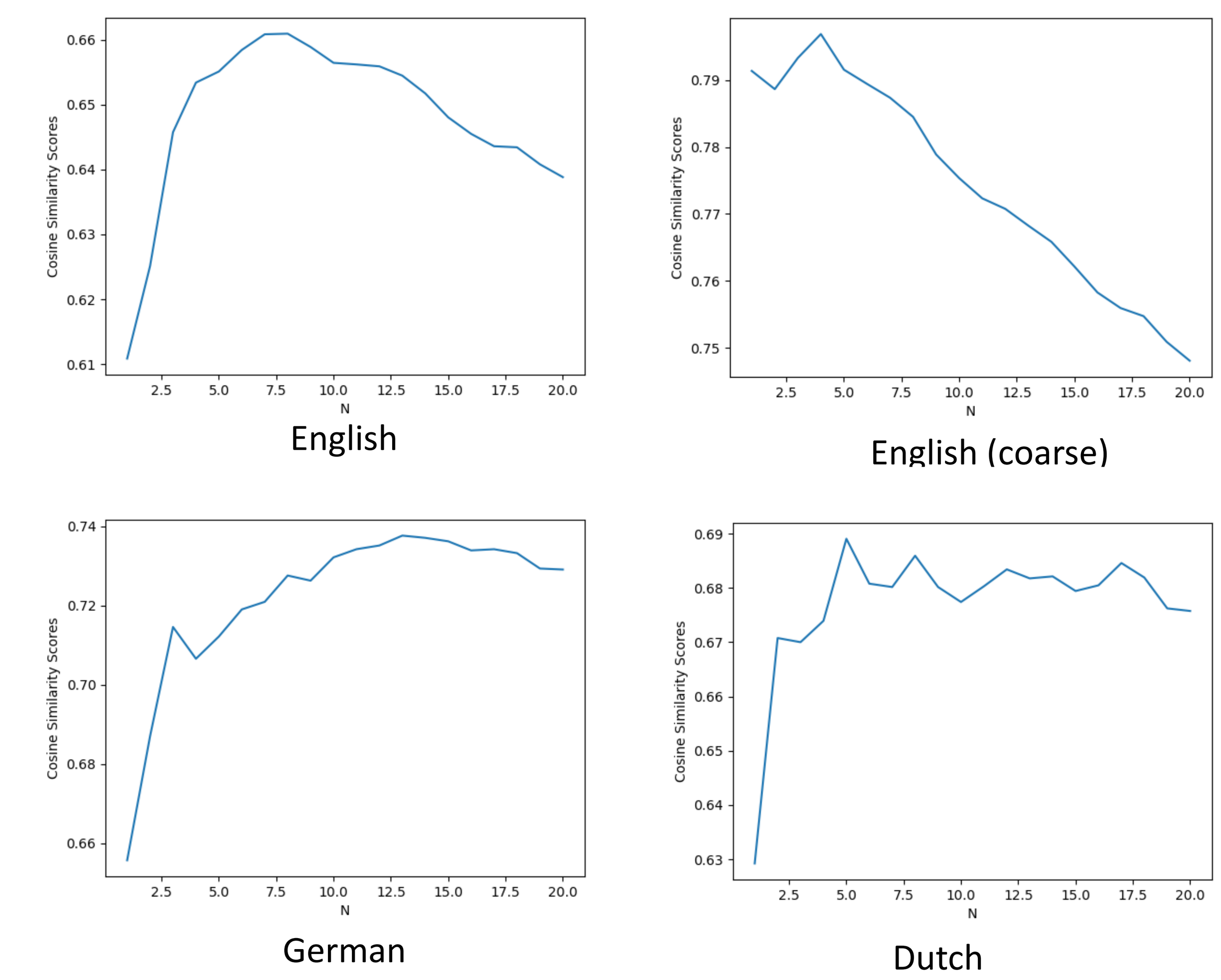


Figure 1: Cosine similarity score vs. k for Form-based Prediction

However, we achieved more than 78% of prediction accuracy if we take part-of-speech into consideration when selecting neighbors

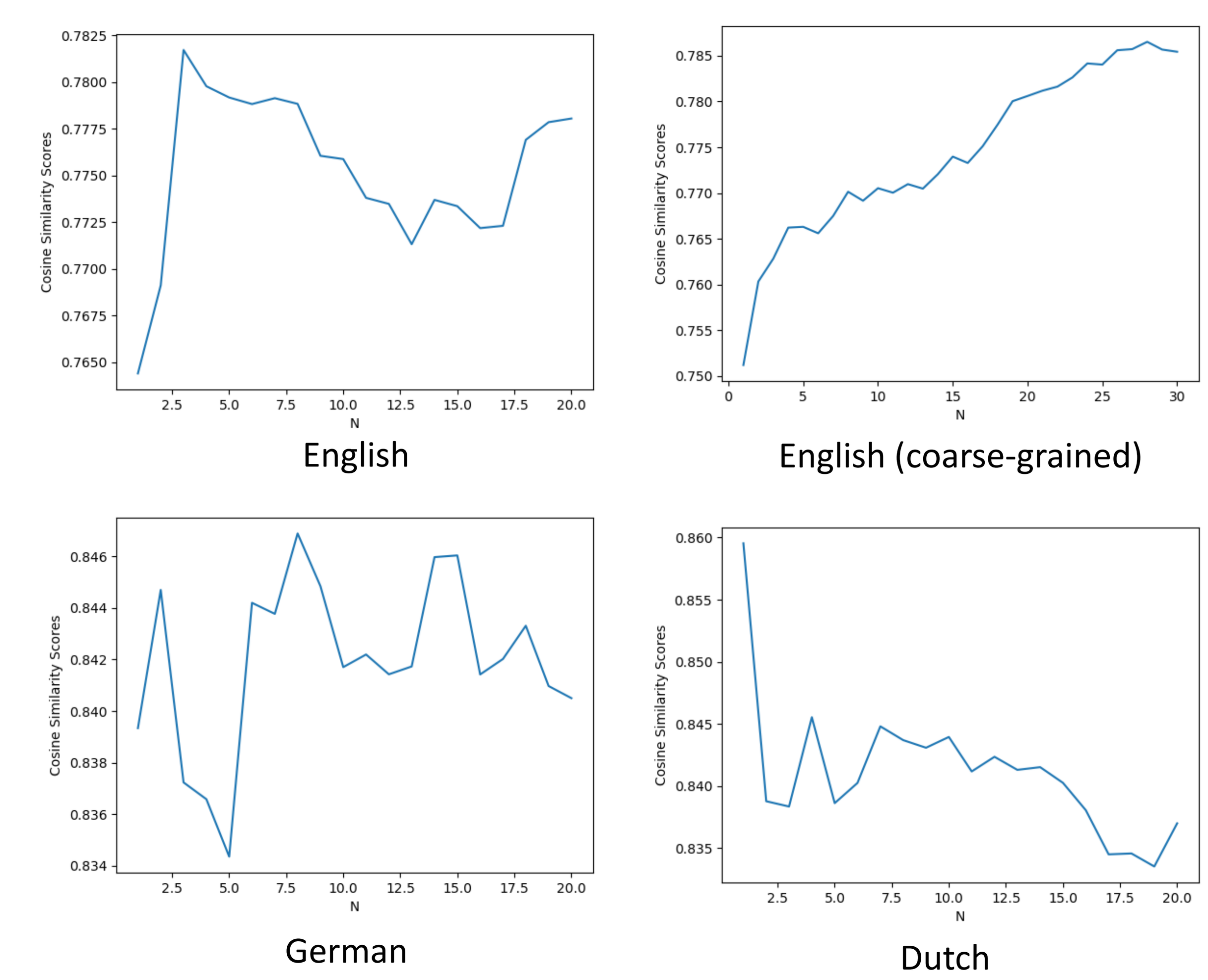


Figure 2: Cosine similarity score vs. k for POS-based Prediction

Future Work

- Scaling up to generate tag vectors for millions of words
- Scaling up to generate tag vectors in over 300 languages
- Releasing data to the public

References

[1] Lasha Abzianidze and Johan Bos. 2017. Towards universal semantic tagging. In *IWCS 2017*
[2] PMB (Parallel Meaning Bank) from: http://pmb.let.rug.nl/

RUTGERS