Towards a dynamic application of distributional semantics

Marijn Schraagen
Digital Humanities Lab
Utrecht University, The Netherlands
M.P.Schraagen@uu.nl

Abstract

Distributional approaches of natural language semantics generally take a static approach, in which a semantic space is computed once for a full corpus in order to quantify term similarity. The current research proceeds towards a more dynamic approach, by measuring differences in semantic space within subsequent sections of a document as well as between documents. Experiments are performed on a set of philosophical documents in 19th century German.

1 Introduction

In distributional semantics, the semantics of a given word are generally modelled as a vector comprised of the word's context. Semantic similarity between words is subsequently estimated by comparing the associated vectors, using well-known vector similarity measures, such as cosine similarity or dot product (Erk, 2012).

In most approaches, word vectors are considered to be static for a given document or corpus. However, within a document or corpus it is expected that word meanings fluctuate, e.g., when different aspects of a concept are discussed, when the topic of a discussion shifts, or when the meaning or the prevalent sense of an expression changes (*semantic shift*). To address these phenomena and allow for the study of consistency of word semantics, a vector representation can be modeled dynamically throughout a document or corpus.

The current work implements a dynamic vector representation by computing semantic spaces for distinct corpus parts, in order to estimate the differences in semantic space between these parts. The hypothesis is that such differences reflect the dynamics of the underlying semantic and pragmatic structure of the documents, which can be investigated in future work. Semantic spaces assigned to each document part are compared in a pair-wise manner, implemented as the total difference in similarity scores for selected term pairs.

2 Experiments

Initial experiments have been performed on a set of German texts written by the Idealist philosopher Georg Hegel in the 19th century, specifically the *Encyclopedia of the Philosophical Sciences* and the *Phenomenology of Spirit*. Paragraph numbers, footnotes, non-latin characters, internal and external references are removed in preprocessing. All text is tokenized, lemmatized and assigned a part-of-speech (POS) tag using the ParZu dependency parser for German (Sennrich et al., 2009). Texts are split into parts of approximately 36,000 tokens. For each part a semantic space is computed using the DISSECT toolkit (Dinu et al., 2013). The spaces are transformed using Positive Point-wise Mutual Information (ppmi). For comparison of semantic spaces, a set of high-frequent terms is selected which are both meaningful in the philosophical domain as well as dispersed throughout the corpus (see Figure 1).

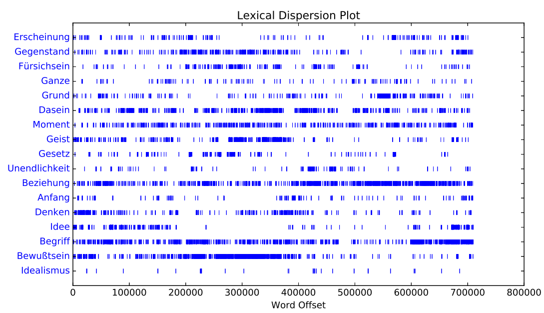

Figure 1: Dispersion of terms used in semantic space similarity computation.

The frequency criterion is necessary for computing valid statistics, dispersion ensures that differences between semantic spaces are caused by changes in semantic relations rather than changes in the occurrence of terms as such, and domain-related terms are hypothesized to reflect the philosophical discourse.

The algorithm to compare two semantic spaces S and S' computes the similarity δ between term pairs in each space. The two resulting lists of similarities are subtracted to obtain a list $\Delta_{S,S'}$ of similarity differences. The sum of $\Delta_{S,S'}$ represents the difference between two semantic spaces in a single number (see Table 1 for an example).

term 1	*term 2*	δ_S	$\delta_{S'}$	$\Delta_{S,S'}$
Dasein	Geist	0.35	0.44	0.09
Dasein	Gesetz	0.68	0.52	0.16
Geist	Gesetz	0.04	0.51	0.47
Sum				*0.72*

Table 1: Example semantic space difference score.

Four different selection methods are used as input for the computation of semantic spaces (Figure 2). First, the classic distributional semantics method is used, which computes a vector for every word in a text using a small window, in this case of length 2 on both sides (*all*). Second, the same method is used, but only (lemmatized) nouns are used as target terms (N_{target}). Third, nouns are used as target terms, and the first two nouns preceding and following the target term are selected. In this case the terms in the window may occur at a distance larger than the window size (N_{all}). Finally, a method is implemented using a large window of size 100 to either side of the target term, and a strong dimensionality reduction using only around 200 terms in the term matrix, selected based on frequency and domain relevance (*domain*). This method is intended to reflect the intuition that domain-specific terms occurring within the same line or paragraph are more likely to be related than terms that may share a neighborhood in the semantic space but do not co-occur in relatively close proximity.

Eine Erklärung, wie sie einer Schrift in einer Vorrede nach der Gewohnheit vorausgeschickt wird über den Zweck [...]

All: einer–Schrift, einer–in, einer–Vorrede, einer–nach
N_{target}: Vorrede–in, Vorrede–einer, Vorrede–nach, Vorrede–der
N_{all}: Vorrede–Erklärung, Vorrede–Schrift, Vorrede–Gewohnheit, Vorrede–Zweck
Domain: Erklärung–Zweck, Schrift–Zweck

Figure 2: Example sentence showing term pairs resulting from selection methods, window size 2.

3 Results

A comparison of semantic space difference using the N_{all} selection method shows a higher similarity for spaces within a document compared to between documents (see Table 2), a result that is expected in terms of semantic consistency. However, the other three selection methods do not show consistent difference scores.

	E_1	E_2	P_1	P_2	P_3	P_4	P_5	P_6
E_1	–	11.18	10.66	11.23	9.90	11.39	10.72	10.26
E_2	11.18	–	10.72	11.15	10.49	12.29	12.12	11.98
P_1	10.66	10.72	–	10.19	8.71	10.61	8.97	8.47
P_2	11.23	11.15	10.19	–	9.08	10.41	9.70	10.63
P_3	9.90	10.49	8.71	9.08	–	8.22	7.70	8.23
P_4	11.39	12.29	10.61	10.41	8.22	–	6.00	8.60
P_5	10.72	12.12	8.97	9.70	7.70	6.00	–	6.55
P_6	10.26	11.98	8.47	10.63	8.23	8.60	6.55	–

Table 2: N_{all} semantic space differences.

4 Discussion

The preliminary experiments suggest that the described method is able to group different parts of the same document based on distributional similarities of shortlisted terms common to all documents. This result is a first step towards a more fine-grained characterization of the development of semantic spaces throughout a document. This in turn is hypothesized to reflect the discourse structure in a text, e.g., shortlisted terms in P_1 and P_6 might be used in a semantically similar way, which needs to be confirmed in future work by domain experts. Use of a more extensive methodology as well a larger dataset are envisioned for further exploration of semantic space dynamics.

References

Georgina Dinu, Nghia The Pham, and Marco Baroni. 2013. DISSECT - DIStributional SEmantics Composition Toolkit. In *Proceedings of the System Demonstrations of ACL 2013*. ACL.

Katrin Erk. 2012. Vector space models of word meaning and phrase meaning: A survey. *Language and Linguistics Compass*, 6/10:635–653.

Rico Sennrich, Gerold Schneider, Martin Volk, and Martin Warin. 2009. A new hybrid dependency parser for German. In *Proceedings of GSCL 2009*, pages 115–124. GSCL.