Words are Malleable: Computing Semantic Shifts in Political and Media Discourse

Azarbonyad, H.; Dehghani, M.; Beelen, K.; Arkut, A.; Marx, M.; Kamps, J.

Publication date
2017

Document Version
Final published version

Words are Malleable: Computing Semantic Shifts in Political and Media Discourse

Hosein Azarbonyad
University of Amsterdam
h.azarbonyad@uva.nl

Mostafa Dehghani
University of Amsterdam
dehghani@uva.nl

Kaspar Beelen
University of Amsterdam
k.beelen@uva.nl

Alexandra Arkut
University of Amsterdam
aeearkut@gmail.com

Maarten Marx
University of Amsterdam
maartenmarx@uva.nl

Jaap Kamps
University of Amsterdam
kamps@uva.nl

ABSTRACT

Recently, researchers started to pay attention to the detection of temporal shifts in the meaning of words. However, most (if not all) of these approaches restricted their efforts to uncovering change over time, thus neglecting other valuable dimensions such as social or political variability. We propose an approach for detecting semantic shifts between different *viewpoints*—broadly defined as a set of texts that share a specific metadata feature, which can be a time-period, but also a social entity such as a political party. For each viewpoint, we learn a semantic space in which each word is represented as a low dimensional neural embedded vector. The challenge is to compare the meaning of a word in one space to its meaning in another space and measure the size of the semantic shifts. We compare the effectiveness of a measure based on optimal transformations between the two spaces with a measure based on the similarity of the neighbors of the word in the respective spaces. Our experiments demonstrate that the combination of these two performs best. We show that the semantic shifts not only occur over time, but also along different viewpoints in a short period of time. This abstract is based on [1].

SUMMARY OF THE RESEARCH

Words are always ‘under construction’, their meaning is unstable and malleable [7]. Semantic fluctuations can result from a concept’s ‘essentially contested’ nature. "What does democracy mean?" or "what values are democratic?". The answer changes according to the ideological perspective or *viewpoint* [3] of the person uttering the term. Equally important is the influence of historic events. The understanding of ‘terrorism’, for example, has significantly changed as a result of the 9/11 attacks [2]. Currently, only a few studies have attempted to compute the ‘malleability of meaning’ and monitor semantic shifts [4–6]. Most (if not all) of these approaches have focused their efforts to uncovering change over time. However, there are other valuable dimensions that can cause semantic shifts such as social or political variability. In this paper, we explore the *semantic stability* of words by computing how contextual factors, such as social background and time, shapes—or, at least reflects—shifts in meaning.

We first use distributional semantics to generate embedding spaces from categorized corpora, where a category can be a certain context (such as speeches given by a political party). Then we propose different approaches to compare the vector representation of words between spaces. In the remainder of this paper, we define each of these categories as *viewpoints*, since they reflect the semantic constellation of terms from a specific social perspective. In this paper we only consider two viewpoints. However, our approaches are easily extendable to multiple (i.e non-binary) viewpoints. The challenging part of this task, and the main contribution of this paper, is to develop techniques that compare vectors across spaces with different dimensionality structures. We consider three methods for comparing meaning across vector spaces. (1) We create a linear mapping between two embedding spaces, project words from one embedding space to the other and measure whether the projected word lands closely to the word in the other space. (2) For each viewpoint, we construct a graph such that the nodes are words and edges are the similarities between them. Then, using graph-based similarity measures we compute how similar the neighbors of a word in two embedding spaces are. (3) We define a measure that combines these two measures.

As stated, in this work our *main research problem* is to study how semantic shifts in words are happening not just over time dimension but also social dimension, quantify the size of shifts, and explore the applications that can benefit from the information about shifts. We evaluate the proposed approaches in three different tasks: measuring semantic shifts, document classification, and contrastive viewpoint summarization. Our main contributions are: (1) We show that semantic shifts not only occur over time, but also across different viewpoints in a short period of time. (2) We improve the linear mapping approach [8] for detecting semantic shifts and propose a graph-based method to measure the size of semantic shifts in the meanings of words. (3) We employ word stability measures in contrastive viewpoint summarization and document classification and extensively evaluate our proposed approach to these tasks. (4) Our analysis shows that the two laws of semantic changes proposed in [4] hold for semantic shifts across viewpoints. Moreover, we introduce a new law of semantic changes which implies that concrete words are less likely to shift meaning while abstract words are more likely to do so. (5) We make the evaluation dataset for detecting semantic shifts and contrastive viewpoint summarization publicly available.[1]

REFERENCES

[1] H. Azarbonyad, M. Dehghani, K. Beelen, A. Arkut, , M. Marx, and J. Kamps. Words are malleable: Computing semantic shifts in political and media discourse. In *CIKM ’17*, 2017.
[2] E. Bleich, H. Nisar, and R. Abdelhamid. The effect of terrorist events on media portrayals of islam and muslims: evidence from new york times headlines, 1985–2013. *Ethnic and Racial Studies*, 39(7):1109–1127, 2016.
[3] W. B. Gallie. Essentially contested concepts. In *Proceedings of the Aristotelian society*, volume 56, pages 167–198, 1955.
[4] W. L. Hamilton, J. Leskovec, and D. Jurafsky. Diachronic word embeddings reveal statistical laws of semantic change. In *ACL ’16*, pages 1489–1501, 2016.
[5] T. Kenter, M. Wevers, P. Huijnen, and M. de Rijke. Ad hoc monitoring of vocabulary shifts over time. In *CIKM ’15*, pages 1191–1200, 2015.
[6] V. Kulkarni, R. Al-Rfou, B. Perozzi, and S. Skiena. Statistically significant detection of linguistic change. In *WWW ’15*, pages 625–635, 2015.
[7] T. Lansdall-Welfare, S. Sudhahar, J. Thompson, J. Lewis, F. N. Team, and N. Cristianini. Content analysis of 150 years of british periodicals. *Proceedings of the National Academy of Sciences*, 114:457–465, 2017.
[8] T. Mikolov, Q. V. Le, and I. Sutskever. Exploiting similarities among languages for machine translation. In *ICLR*, 2013.

[1]The datasets are available here: http://dx.doi.org/10.7910/DVN/BJN7ZI.