Graded Similarity in Context

Tim Lawson

December 10, 2023

1 Introduction

In his *Foundations of Arithmetic*, Frege promises "never to ask for the meaning of a word in isolation, but only in the context of a proposition" (Frege 1980, p. xvii). This 'context principle' is intuitive: words are frequently polysemous, or assume different connotations and emphasis within different expressions.

History of contextualising word representations: word-sense disambiguation, Kintsch (2001), relationship to composition operations (Mitchell and Lapata 2008). The same contextualisation operations can be applied to (static) word embeddings, as opposed to their count-based predecessors. But these efforts have largely been overtaken by models that naturally produce contextual representations, i.e., that are functions of the entire input sequence rather than only individual tokens. When did this start? Recurrent neural networks, sequence-to-sequence models, transformers, etc.

Despite the popularity of these models, there are relatively few like-for-like comparisons of them with respect to simpler operations on static word embeddings. CoSimLex presents such an opportunity: the task is to produce estimates of the different similarities of the same pair of words in different contexts. Therefore the present paper seeks to quantify the performance benefit of using the contextual representations of a transformer model as opposed to the static representations of its input embeddings with simple contextualisation operations, inspired by e.g. Kintsch and Mitchell2008.

Background of CoSimLex, what specifically the task is.

References

Frege, Gottlob (1980). *The Foundations of Arithmetic: A Logico-Mathematical Enquiry Into the Concept of Number*. 2nd rev. ed. Evanston, Ill: Northwestern University Press.

Kintsch, Walter (2001). "Predication". In: *Cognitive Science* 25.2, pp. 173–202.

Mitchell, Jeff and Mirella Lapata (2008). "Vector-Based Models of Semantic Composition". In: *Proceedings of ACL-08: HLT*. Columbus, Ohio: Association for Computational Linguistics, pp. 236–244.