Distributional semantics and vector space models of lexical meaning

Table of Contents

Introduction

Our discussion was mainly to answer the assignment questions and to discuss other points that we do not understand. We summarize conclusions are as follows:

General discussion

Distributional semantic models —will be denoted in the upcoming text as DS— analysis different contexts (corpus) trying to mathematically represent words meaning. The hypothesis that DS is based on is that "Words that occur in similar contexts tend to have similar meanings". [1] The main idea is to model the distances between words by examining different contexts that these words appear in. This modeling is achieved by utilizing linear algebra tools to represent words under study (target words) in terms of other words (contextual words) that appear with or have some connection to the target words. By utilizing linear algebra tools, we can represent each target word by a vector in space of words that its axes are defined the contextual words.

[1] "From Frequency to Meaning: Vector Space Models of Semantics." 4 Mar. 2010, https://arxiv.org/abs/1003.1141. Accessed 15 May. 2020.

The similarity between any two target words could be then examined by calculating the cosine of the angles between their vectors representations. Other semantic relations could also be examined according to the analysis tool that is applied to the words vector.

DS and model-theoretic semantics

Now, the word meaning can be represented in terms of some contextual words in the space vector model. One may see such space modeling as a representation of the human knowledge about the target word or as a reflection of the mental image of the word. While in model-theoretical, the word meaning is represented in terms of an entity that its members refer to a set of objects in the world.

In that sense, the sense of a word in DS is represented by its uses in different contexts. On the other hand, in theoretic model words, the meaning is based on the domain of objects.

DS Capabilities and limitation

DS models can capture semantic similarities such as synonym, priming, and categorization. In addition by utilizing the DS models, we can automatically learn about word polysemy. Based on the natural language compositionality features, we can construct phrases and sentence meaning from the meaning representations of its constituents.

DS is not fully adequate on some of the central aspects of formal semantic approaches: truth conditions, entailment, reference, and certain aspects of compositionality.

Extracting feature norms is among problems for distributional approaches.

Applications of distributional semantics

DS has become the state-of-the-art way on the development of various NLP applications since Mikolov and Dean (2013) proposed word2vec. To list some of the current works:

- Neural information retrieval (Mitra et al., 2016)
- Neural machine translation
- Named-entity recognition (Yadav, 2019)
- Twitter election classification (Yang et al., 2018)
- Sentence classification (Kim 2014)
- Machine translation (Mikolov et al. 2013)
- Syntactic productivity in diachrony (Florent, 2016)
- part-of-speech tagging (Santos et al., 2014)
- sentiment analysis/classification (Liu 2015; Tang et al. 2014)

Advantages of and challenges in using DS

Advantages of using DS model can be summarized as follows:

- Lexical meanings are learned unsupervised way
- DS model benefits can be extended by utilizing supervised machine learning techniques like learning about polysemy without needing to refer to dictionaries
- Word sense disambiguation

Challenges that are faced when using DS:

- Sentence and phrase modeling using natural language compositionality feature
- Computational power needed

Also, One of the main challenges is to find enough digital resources for many of the world languages including corpus as well as the computational resources to process them. There is also a lack of enough data of different domains for languages considered better studied like human dialogue to create data-driven dialogue systems.

Open problems in DS

Some of the open questions can be listed as follows:

- How to encode sentence structures taking word order into consideration?
- How distributional representations can be projected from the lexical to sentence or discourse level?
- How to include additional grammatical information like word-order to get better results in different NLP application development.

DS and dimensionality reduction

It was stated that it's not clear if applying SVD to the term matrix is helpful[2]. In our opinion, we can agree with this statement in the sense that using SVD is not the best method that deals with sparse matrices. Taking into consideration that term-term matrix which is based on word counts or tf-idf is sparse in natural. However, it is crucial to use dimensionality reduction to ensure that the axes of the constructed space vectors are orthogonal to each other. It is worth mentioning that there is no guarantee that the axes which are constructed from the contextual words will be orthogonal.

[2] "Vector Space Models of Lexical Meaning - The Handbook of Contemporary Semantic Theory" 21 Aug. 2015, https://onlinelibrary.wiley.com/doi/10.1002/9781118882139.ch16. Accessed 16 May. 2020.

Statement of Contribution

We had three meetings, each one took about 40 minutes, Charlotte, Tewodros, and Mohamed contributed equally to the discussions. Charlotte initiated the work by extracting the answers to the questions from the recommended articles. Tewodros showed us some new articles after that, with interesting questions and more theory. He also contributed to extracting the answers to the assignment questions and made the first draft of the summary. Mohamad contributed to answering the assignment questions, raised some points during the discussion, prepared the slides for the presentation, presented the work during the class, and compiled all the works produced by Charlotte, Tewodros, and himself into the final summary.

Bibliography

Bahdanau, Dzmitry, Kyunghyun Cho, and Yoshua Bengio. "Neural machine translation by jointly learning to align and translate." arXiv preprint arXiv:1409.0473 (2014).

Yadav, Vikas, and Steven Bethard. "A survey on recent advances in named entity recognition from deep learning models." arXiv preprint arXiv:1910.11470 (2019).

Yang, Xiao, Craig Macdonald, and Iadh Ounis. "Using word embeddings in twitter election classification." Information Retrieval Journal 21.2-3 (2018): 183-207.

Kim, Yoon. "Convolutional neural networks for sentence classification." arXiv preprint arXiv:1408.5882 (2014).

Mikolov, Tomas, Quoc V. Le, and Ilya Sutskever. "Exploiting similarities among languages for machine translation." arXiv preprint arXiv:1309.4168 (2013).

Perek, Florent. "Using distributional semantics to study syntactic productivity in diachrony: A case study." Linguistics 54.1 (2016): 149-188.

Santos, Cicero D., and Bianca Zadrozny. "Learning character-level representations for part-of-speech tagging." Proceedings of the 31st international conference on machine learning (ICML-14). 2014.

Tang, Duyu, Bing Qin, and Ting Liu. "Deep learning for sentiment analysis: successful approaches and future challenges." Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 5.6 (2015): 292-303.

Tang, Duyu, et al. "Learning sentiment-specific word embedding for twitter sentiment classification." Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2014.