Capturing bias in the left-sided and right-sided news corpra

Anonymous ACL submission

Abstract

Many people try to become aware of the important events happening in the world through reading news articles. Many of the big events and decisions, directly or indirectly affects our lives, so it is important to know about them and make reactions if we think it is needed. Unfortunately most of the news sources are politically biased and they convey the news in a way to give the reader an opinion close to themselves. Our work gets the word embedding for each of the sides corpora and aligns the two embeddings. With aligning of the embedding spaces we can compare the vectors and use metrics to show the bias between the left-sided and right-sided news sources.

1 Related Works

Some work have been done focusing on the evolvement of temrs during time. In a work by Garg *et al.*, they integrate word embeddings trained on 100 years of text data with the US Census and develop metrics based on word embeddings to characterize how gender stereotypes and attitudes toward ethnic minorities in the United States evolved during the 20th and 21st centuries starting from 1910 (Garg et al., 2018). They compute the average embedding distance between words that represent womene.g., she, female and a group of gender neutral words like occupations, also compute the average embedding distance between words that represent men and the same occupation words. They have used the intuitive and natural metric for the embedding bias which is the average distance for women minus the average distance for men A group of works concentrate on the evolving of word semantics during time. They have captured interesting biases looking at the metrics in different years. There is no embedding alignment in their work, they get the static word embedding for each year and calculate the metrics for that year and show the gradual change of numbers in plots. The data and code related to their paper are available on GitHub [1]. Using the similar idea for our work we need to get two set of words representing each of political sides and also a set of political neutral interesting words. Finding those sets of words that are also frequent in our dataset is challenging. Another drawback is that calculating euclidean differences in embedding spaces is not a very robust metric.

Some of the related works are focusing on bilingual word embedding which builds semantic embeddings associated across two languages. The work of (Zou et al., 2013) introduces an unsupervised neural model to learn bilingual semantic embedding. The result of this work might not be very interesting for our task because it embeds our two different set of corpus (left and right) in a way that the corresponding words that have the same meanings will end up very close in the vector space. Another disadvantage of this method is its slowness; it took 19 days for their model to train on a 8-core system. This paper is old and they have compared their methods like naive and pruned tf-idf and we don't have comparison of it with contemporary state of the art models.

We want to be able to separately embed the words from the corpora corresponding to each of the right and the left side news sources and then align the vector spaces. The work of (Hamilton et al., 2016) use orthogonal Procrustes in order to align word embeddings across time-periods. This method searches for the best rotational alignment and preserves cosine similarities. They use two measures to evaluate their results: synchronic accuracy (i.e., ability to capture word similarity) and diachronic validity (i.e., ability to quantify semantic changes over time) which they do in two ways:

[1] https:// github.com/nikhgarg/EmbeddingDynamicStereotypes

detecting known shifts and also discovering shifts from data. This method can be applied to our problem because we are trying to find the alignment between embeddings of left-wing news corpora and right-wing news corpora. We also can look at the embeddings of all news corpora during time spans and another interesting question is whether the similarity of the words changes over time in compare to left terms and right terms. A drawback of their method can be that they only look at rotational alignment and don't capture the changes in the cosine similarities between the words. They have their code available on github.[2]

Later than Hamilton's work, there is another work(Yao et al., 2018) that instead of aligning different static embeddings simultaneously learns time-aware embeddings. Previous techniques usually do not consider temporal factors, and assume that the word is static across time. They are interested in computing time-aware embedding of words. They have used qualitative and quantitative methods to evaluate temporal embeddings for evolving word semantics. Their work can be modified for our problem setting to obtain political-aware embeddings.

References

Nikhil Garg, Londa Schiebinger, Dan Jurafsky, and James Zou. 2018. Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences*, 115(16):E3635–E3644.

William L Hamilton, Jure Leskovec, and Dan Jurafsky. 2016. Diachronic word embeddings reveal statistical laws of semantic change. *arXiv preprint arXiv:1605.09096*.

Zijun Yao, Yifan Sun, Weicong Ding, Nikhil Rao, and Hui Xiong. 2018. Dynamic word embeddings for evolving semantic discovery. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, pages 673–681. ACM.

Will Y Zou, Richard Socher, Daniel Cer, and Christopher D Manning. 2013. Bilingual word embeddings for phrase-based machine translation. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1393–1398.

[2] https://github.com/williamleif/histwords