Author: Jacob Fuehne

Date: 10/26/2020

Research Paper Summary

Agree or Disagree: Predicting Judgments on Nuanced Assertions, *Association for Computational Linguistics*, 2020

1. Problem definition and the main ideas of the research

When reading a sentence, one of the most basic reactions is to agree or disagree with it. In this paper, the authors define and work to solve two NLP tasks. They are to predict individual judgements and judgements of whole groups. In this task, they define assertions as textual utterances that are explicit, relevant, and that do not contain multiple positions. The degree to which two assertions are judged to be similar is defined as judgement similarity.

2. Significance of research study (Importance and Challenges of research problem)

The ability to predict the way that individuals and groups of people feel towards an assertion is very beneficial in business, public relations, and research tasks. Businesses, public figures, and other decision makers can judge how the public will react to different perspectives on a controversial topic so that they can make an informed decision. Researchers and general users would also stand to gain from these predictions, just as they would benefit from opinion polls. An issue in determining judgement similarity is that the judgements of other people are not easily available.

3. Main research questions and assumptions

This research seeks to find a method of predicting the judgements of individuals and of groups of people to assertions. It is assumed in this research that by referring to a large dataset of assertions which people have judged, you can measure similarity between previous assertions

and a new assertion and the researchers hypothesize the judgement for a new assertion will be the same as for a highly similar one. Thus, it is assumed that these assertions and judgements are generalizable and that they can be applied to new assertions.

4. Research Methodology

In this paper, the authors use the dataset, Nuanced Assertions on Controversial Issues, which contains annotated assertions on a number of controversial issues. The dataset is organized in a method similar to what one would find in social media, with upvotes and downvotes, however it does not have the same experimental problems as social media, as legal issues and surrounding variables have been controlled for in the collection. Using the dataset, a similarity of assertions matrix is formed, and cosine similarities between each unique pair of judgement vectors for two assertions are calculated. In testing their methodology of judgement similarity, they implement Siamese neural networks (SNN) and SVM, as well as semantic text similarity algorithms such as embedding distance, jaccard, greedy string tiling, and longest common substring.

5. Experiments

Baseline semantic text similarity algorithms, as previously mentioned, are used in calculating the overlap between the surface forms of assertions. In addition, two neural networks (SNN and SVM) are trained by the dataset. These models are then compared against 100,000 judgements on 2000 assertions. The trained models show much higher correlation coefficients than the baselines, demonstrating that they capture judgement similarity separately from semantic similarity. In the final results, it is found that with the proposed judgement similarity methodology, the prediction of individual judgements yield competitive results, and the prediction of group judgements prove to be the best approach by a large margin.

6. Discussion

6.1 Important aspects

- Detailed with clear visualizations

This paper is very clearly written and it properly addresses distinctions in various performance metrics between approaches in the discussions, as well as visualizing these distinctions. I found this particularly useful when they discussed differences between SVM vs SNN in performance and reasoned through it by discussing the structure of the models.

6.2 Limitations of the paper

- Limited testing on baseline methods

While it likely would not have outperformed the neural network models, I think that they could have tested at least one more semantic text similarity algorithm that is more modern/SOTA, as almost all of the STS algorithms tested are nearly 20 years old. Considering that they used imported library code, this was likely done for the sake of saving time, but it is likely that more modern code could have been found/used.

6.3 Questions for presenter

- So, for the models tested, it was found that the ordering of effectiveness (ie: correlation coefficient and test accuracy) was SNN, CNN, and then SVM. Can you explain why SNN is noticeably more effective than CNN and SVM?