ACL Paper Summary

Members: Salmaan Afzal and Shariq Azeem

Introduction

For this assignment, we selected a paper called "Evidence-based Factual Error Correction", from the *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. This paper was written and submitted by James Thorne and Andreas Vlachos in 2021, as part of their research on fact verification, supported by donations from Google Cloud.

At the time of the research, Thorne was a PhD student at the University of Cambridge, supervised by Vlachos and funded by Amazon Alexa Graduate Research Fellowship. Currently, he is an assistant professor at the KAIST Kim Jaechul AI Graduate School. Vlachos is a professor of Natural Language Processing and Machine Learning at the Department of Computer Science and Technology at the University of Cambridge and a fellow of Fitzwilliam College. Similar to Thorne, Vlachos was supported by the ERC grant AVeriTeC.

Problem Addressed

In this paper, they address the topic of fact verification, which is a process of deciding whether a given claim is true or false, using some type of evidence. They suggest that a better process should be considered, which they are calling as *Factual Error Correction,* which will not only assign 'SUPPORTED' or 'REFUTED' labels to the given claims, but it will also rewrite those claims to statements that are better supported by the given or retrieved evidence.

Prior Work

Before Thorne and Vlachos made their model and performed their experiment, they mentioned some previous works, done by other researchers, that addressed a similar topic. One of those works was a Grammatical Error Correction (GEC) model, discussed by twelve researchers in three separate papers. The purpose of this model was to take input sentences from language learners and make meaning-preserving changes, such that there were no more grammatical errors left in them. Since this model only made some surface-level transformations to the given sentence, there was no external information required.

Another work they mentioned utilized a sequence-to-sequence model to remove inconsistencies from artificially generated summary. This model took two inputs: the original source text and its reference summary, generated from abstract summarization;

then the model restored the summary, such that it does not contain any inaccurate information and it is consistent with the source text.

Lastly, one of the most important works they mentioned was done by Darsh J Shah, Tal Schuster, and Regina Barzilay, in 2020. Their work consisted of using short factoid claims to introduce changes to Wikipedia passages. This was done through a two-step process: first, the important tokens in the given claim were masked, using a masker; second, those masks were replaced with tokens from the given evidence, using a corrector. This model gave good results, but it was dependent on two conditions: first, the assumption that the given claim always required meaning-altering change; second, gold standard evidence had to be explicitly provided.

Thorne and Vlachos extend the work of Darsh J Shah, Tal Schuster, and Regina Barzilay, but they try to make a model that does not depend on the conditions mentioned above.

Unique Contributions

The previous work done on factual error correction used to copy portions of the claim rather than correct the claim which human testers looked at corrections as most unintelligible. The human testers thought they the results from their factual error corrector were more intelligible. So, they improved the quality of factual error correction for the future. The purpose of this was to explain reasoning with evidence to see if a claim holds up.

Evaluations

Thorne and Vlachos evaluated how well the factual error correction worked by comparing with results of manual testers. This has been used a fair number of times for human language processing with parts of speech. They used multiple manual testers to evaluate how good was the algorithm in detecting error correction. They used 4 experts to test it that worked in their lab but not on the factual error correction project. They also used the SARI metric to determine how accurate it was. SARI was used for sentence simplification using n-grams to determine to compare information. At the time of the writing a supervised system had a 69% success rate which isn't great but a good start.

Citations and Conclusion

Andreas Vlachos has received a total of 5167 citations on Google Scholar, and most of his work has been about fact extraction, verification, and correction. This is a very important topic, especially in this time of technology and social media, where any post can reach an audience of millions in a small period of time, since this enhanced ability can be used to spread false information.

The paper we chose made some improvements to the existing models, but further research is needed to make a model that can be implemented on different forms of media

and different channels of communication. The main goal is to remove as much corrupt data from those channels as possible and supply the audience with error-free information.