SUMMARY

Factoring Fact-Checks:

Structured Information Extraction from Fact-Checking Articles

By Ayushi Jain, BTech CSE(2023)

This research paper talks about the experiments done to automate summarisation of fact checks into three factors: claim, claimee and the claim verdict. If successful, **this will bring a positive impact on e-journalism, as the fact-checkers will not need to spend large amounts of time using tools like ClaimReview to annotate their fact checks to summarise content for the readers.**

6,216 English fact-check articles with 1,038 words and 24.4 paragraphs on average were used for data exploration.Analysis of data shows that 19% of fact-checkers publish 94% of the fact-checks. Using exact string matching between factors and fact-checks, it was found that 76% verdicts , 80% claimants and only 32% claims can be matched.As claims can be paraphrased in different ways, a fuzzy string matching concept is used where **minimum window substring** algorithm is used to find the paraphrased factors.It is observed that there is a huge difference in the graphs for the **Probability Density Function** of the articles written by well known and not so well known writers for the relative positions of the matched claimant and claimee. This is a unique observation that helped improve the performance of the model. **These points formed the basis of design of the models and have been supported with suitable graphs and statistics.**

Due to the large number of possible linguistic patterns, this problem was classified as a **sequence tagging problem**. **For this, BERT models have been used.** They have a limit of processing 512 characters at a time.A modification of BERT is made where the uniform [CLS] token is replaced with a paragraph position token, to help the BERT model learn better representations for each paragraph based on its location.This has a significant positive effect on the performance of the model. **Another baseline model was developed that uses general observations, and tools like ClaimBuster as a baseline(for claim tagging) to tag the factors.**

The models are evaluated on the basis of how well they tag claims, claimants and verdicts. For evaluation, ROUGE score, the concepts of loose score(measures how correctly the models tag a factor provided the factor is tagged) and tight scores(product of the loose score and the tagged percentage) have been used.

The baseline models always return nonempty predictions. **The BERT model could tag overall 69%-75% claims, 86-90% claimants and 96% - 97% verdicts for well known fact-checkers**. **Overall, the BERT model performs much better than the intuitive baseline model.** Errors like non-tagging, tagging the wrong factor and under-over tagging occur frequently when the journalistic styles are different from well-known fact checks,or if there are confusing contexts. **Mostly, the models tag perfectly (a score of 1) if the writing style of the article follows the majority patterns.**

The performance of tagging claimants and verdicts is improved **after half of the test set is mixed with the training set and the model is retrained** which shows that the performance of the model can keep improving when the model is trained on more fact check patterns. **Once trained to a high accuracy, we can use this as a baseline to develop models to summarise readable resources , extend it to other languages, and add modifications to use it in online**

checking where checking according to the rubric is automated. The paper was well written, organised, and easy to understand. It provided us with a valuable insight of how BERT models work, how one can analyse article based data and the importance of training models on sufficiently large datasets. BERT models are however, computational intensive and will prove to be costly for mass use. This has been described in the research paper "[ALBERT: A Lite BERT for Self-supervised Learning of Language Representations"] where the alternative to this model has been discussed.