**Paper 1**

Summary

The paper presents an automated fact-checking system that consists of two parts: evidence retrieval and claim classification. The author explores various methods for each task and evaluates their performance. The best subsystems are then combined to build a complete system. The evidence retrieval methods include a non-exhaustive approach that filters evidence based on keywords and semantic similarity, and an exhaustive approach that uses clustering to avoid discarding relevant evidence. For claim classification, fine-tuned RoBERTa and XLNet models with a classification layer added on top of them are utilized respectively. The final system is based on the combination of a non-exhaustive evidence retrieval model and fine-tuned XLNet model.

Strengths:

The paper provides a detailed exploration of methods for evidence retrieval and claim classification, combined with a clear awareness of the imbalanced data, demonstrating a comprehensive analysis of the problem.

The author addresses the challenge of processing a large evidence dataset by employing novel and reasonable preprocessing techniques and filtering based on relevant keywords, improving computational efficiency while keeping a high F1 score relative to the computationally expensive exhaustive method. The use of semantic similarity and clustering for evidence retrieval shows a thoughtful approach to capturing nuanced relationships between claims and evidence.

The incorporation of fine-tuned RoBERTa and XLNet models for claim classification, along with the explanation of their effectiveness on similar tasks and possible weaknesses, demonstrates the author's deep comprehension of state-of-the-art techniques.

Weaknesses:

While the paper did a good job of utilizing the tables and the figure to show the comparison, more explanations on how to read the colorful tables would be appreciated, for example, the meaning of the different colors used in Table 5.

The paper could benefit from providing more information about how computationally expensive it is to use BERT to calculate the embedding, such as the time elapsed, since many other students have successfully incorporated this model into their systems and it seems not a problem.

Besides, further discussions on more detailed suggestions for future work would contribute to the paper's completeness and provide insights for other researchers.