Fake News Detection with BERT

David Samtani [1]   **Roger Jin** [2]

Abstract

Fake news has become a prevalent issue in today's society, with the ability to spread rapidly and cause harm through misinformation. In this paper, we propose a fake news detection model using a fine-tuned BERT (Bidirectional Encoder Representations from Transformers) model. BERT is a powerful language model that has been successful in a variety of natural language processing tasks. We fine-tune the BERT model on a dataset of real and fake news articles from Kaggle, and evaluate its performance on a holdout test set. Our results achieve an accuracy of 0.574 on the fake news detection task. Our code is made available here.

1. Introduction

Fake news is a growing issue which gives people a wrong idea of important events happening around the world. The implications can be serious - recently a fake tweet pretending to be an announcement from Eli Lilly that insulin is now free caused their stock to plummet 4.4% (Lee, 2022). The growing popularity and development of language models such as ChatGPT makes it easier than ever to generate fake news.

1.1. Related Work

There have been a number of previous studies on fake news detection, using various techniques such as natural language processing, machine learning, and network analysis (Zellers et al., 2019).These approaches have shown promising results, but there is still room for improvement in terms of detection accuracy.

2. Methods

We fine-tune a BERT model (Devlin et al., 2018) on a dataset of real and fake news articles. BERT is a powerful language model that has been successful in a variety of natural lan-

guage processing tasks (Devlin et al., 2018). We use a two-step process for fine-tuning the model. First, we fine-tune the model on a large dataset of real and fake news articles. Second, we further fine-tune the model on a smaller dataset of fake news articles, in order to improve its performance on the fake news detection task.

3. Results and Discussion

We evaluate the performance of the fine-tuned BERT model on a holdout test set of real and fake news articles. Our results show that the fine-tuned BERT model outperforms baseline models, with an accuracy of 0.574 on a Kaggle fake news detection task (WilliamLifferth, 2018).

4. Conclusion

In this paper, we have presented a fake news detection model using a fine-tuned BERT model. Our results show that the fine-tuned BERT model is effective at detecting fake news, with an accuracy of 0.574 on the task. This is only slightly better than random guessing, but we expended very little effort on hyperparameter tuning, and we remain hopeful in the potential of using fine-tuned BERT models for fake news detection.

References

Devlin, J., Chang, M., Lee, K., and Toutanova, K. BERT: pre-training of deep bidirectional transformers for language understanding. *CoRR*, abs/1810.04805, 2018. URL http://arxiv.org/abs/1810.04805.

Lee, B. Y. Fake Eli Lilly Twitter Account Claims Insulin Is Free, Stock Falls 4.37%. *Forbes*, November 2022. URL https://www.forbes.com/sites/brucelee/2022/11/12/fake-eli-lilly-twitter-account-claims-insulin-is-?sh=42ad068941a3.

WilliamLifferth. Fake news, 2018. URL https://kaggle.com/competitions/fake-news.

Zellers, R., Holtzman, A., Rashkin, H., Bisk, Y., Farhadi, A., Roesner, F., and Choi, Y. Defending Against Neural Fake News. *arXiv*, May 2019. doi: 10.48550/arXiv.1905.12616.

[1]Punahou High School, Honolulu, Hawaii [2]Inspirit Inc.. Correspondence to: David Samtani <dsamtani24@punahou.edu>.