

Fake News Detection

Nitin Kulkarni, Chunming Qiao, and Alina Vereshchaka

Department of Computer Science and Engineering,
State University of New York at Buffalo, Buffalo, USA
{nitinvis, qiao, avereshc}@buffalo.edu

With the world being connected, news can spread around the globe instantly. Unfortunately, this leads to a situation in which news agencies want to get the news out first rather than get it right. Often, people believe what they read and propagate the news within their social circles without fact-checking first. Many governments also use news agencies to spread propaganda.

As an example, Russia is waging not only a classic war against Ukraine and democracy, but also an information war. It uses fake news, post-truth, an army of bots, and black psychological operations to spread lies and sow doubts. Russian news media is only legally allowed to report news that aligns with the government’s version of events.¹ The truth about the war thus becomes harder to find, with people believing what the Kremlin tells them, leading them to support the distorted reality of Russia’s war. Disinformation campaigns conducted by Russia in other countries can directly impact the level of public support for their government’s actions to help Ukraine. Thus, it is paramount to identify these cases of disinformation.

In this work, we investigate fake news detection approaches. To this end, we use a dataset of news articles pertaining to Russia’s war in Ukraine. The fake news dataset was collected from VoxCheck². It consists of 409 unique news articles published by 54 different news agencies in the Italian and German media disseminating Russian propaganda. The dataset contains the article titles, texts, links, and information on the specific disinformation cases presented along with their debunking. Each article contains on average 983 words. We also collected 422 articles from The Washington Post and The New York Times; two trustworthy news sources for our real news dataset. The dataset contains the article titles, texts, links, and author names. Each article contains on average 964 words.

We use the BERT (Bidirectional Encoder Representations from Transformers) language model [1] to detect cases of disinformation. BERT can accurately capture the context representation in sentences and produces state-of-the-art results on several natural language processing tasks. We use two versions of BERT: 1. The original BERT model with 109,531,546 parameters and 2. A smaller DistilBERT model with 66,412,226 parameters. We also compare the results with traditional techniques such as Word2Vec.

The Word2Vec model achieves an accuracy of 88.8% with a Logistic Regression classifier and 91.9% with a Random Forest classifier, the smaller DistilBERT model achieves an accuracy of 92.55% while the full BERT model achieves an accuracy of 93.79%.

References

1. J. Devlin, M.-W. Chang, K. Lee, K. N. Toutanova, [Bert: Pre-training of deep bidirectional transformers for language understanding](https://arxiv.org/abs/1810.04805), 2018.
URL <https://arxiv.org/abs/1810.04805>

¹ <https://www.npr.org/2022/03/15/1086705796/russian-propaganda-war-in-ukraine>

² <https://voxukraine.org/en/voxcheck-presents-propaganda-diary-a-database-of-russian-propaganda-in-the-italian-and-german-media/>