Deep Learning Approaches in Fake News

Kaiwen Hu   Yijia Xue   Yuzhuo Kang   Zhixuan Li

1 Motivation

In the latter part of the twentieth century, the internet revolutionized how people share information, often without stringent editorial standards. Recently, social media has emerged as a significant news source for many individuals. As reported by Statistica[1], approximately 3.6 billion people worldwide are active social media users. Social media offers evident advantages in news dissemination, including immediate access to information, free distribution, no time constraints, and diverse content. However, these platforms lack substantial regulation and oversight [1].

Fake news detection falls under the umbrella of text classification, where it is divided into binary classification (distinguishing between real and fake news) or multi-class classification for higher granularity. We are interested in reviewing the state-of-the-art approaches for labeling fake news detection using various datasets of different languages. The main goal of our project will be to build a mixed dataset with different news domains.

2 Goals

The following scientific observations will be evaluated through this project:

- To discover existing approaches on how fake news is labeled and investigate the possibility of combining some of the datasets together to get a larger dataset with multiple domains.
- To find out whether a knowledge-rich language model could increase the performance of the task.

3 Methodology

The experiments start with data preprocessing and feature extraction. Then the main strategy is categorized as supervised methods and transfer learning.

Data Preprocessing As an essential step for NLP tasks, the datasets must be preprocessed before feeding into a neural network. The preprocessing techniques include text cleaning,

---

[1]https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/

punctuation removal, lemmatization/stemming, etc.

Word embedding Feature extraction is another key point for the model performance. The proposed methods would be either statistical methods (e.g. TF-IDF) or context-sensitive representations (e.g. Transformer or BERT).

Supervised Method - Multi-domain fake news detection For the datasets that have only two labels (i.e., Real or Fake), a unified model will be trained as follows: First, the textual sentences are encoded by a transformer, and a set of word vectors is obtained. The feature will then be fed into a neural network, such as RNN. The classification layer will output the corresponding prediction for the input sentences.

Transfer Learning We intend to take advantage of the pre-trained model to get a better performance of the model prediction. Rather than training the dataset from scratch, we will fine-tune a pre-trained model (such as BERT) with the combined datasets.

4 Milestones

Milestone 1: Perform literature review of three sources. Determine dataset and collect data.

Milestone 2: Complete evaluation script of data, and perform baseline(bi-LSTM) analysis incorporating the reviewed literature.

Milestone 3: Develop extensions based on the baseline model(Robert-BERT). Produce draft presentation.

Milestone 4: Wrap up the project by completing all extensions(exploring in milestone 1) and writing the final report.

References

[1] Raza, S. and Ding, C. (2022). *Fake news detection based on news content and social contexts: a transformer-based approach.* International Journal of Data Science and Analytics, 13:335 – 362.