Literature Review: BERT based classification system for detecting rumours on Twitter

I have recently come across a paper titled "BERT based classification system for detecting rumours on Twitter," which inspired me to explore the application of machine learning and deep learning algorithms for cleaning data in the context of sentiment analysis. The primary motivation behind this idea is to leverage such algorithms to evaluate the contribution of individual words to the sentiment of a text, and subsequently remove the words (stopwords, special character, etc.) that do not have a significant impact on the sentiment. By employing state-of-the-art models like BERT, which can capture intricate language patterns and contextual information, I aim to develop an efficient and effective method for data preprocessing in sentiment analysis tasks, ultimately enhancing the performance of sentiment classification models. (I will also look at RNN alternative.)

1 Introduction

The spread of rumours on social media platforms like Twitter has become a significant concern due to the potential for disseminating misinformation. In their paper, *BERT based classification system for detecting rumours on Twitter*, Rini Anggrainingsih, Ghulam Mubashar Hassan, and Amitava Datta propose using a BERT-based classification system for detecting rumours on Twitter. BERT (Bidirectional Encoder Representations from Transformers) is a pre-trained language model that has demonstrated strong performance in various natural language processing tasks, making it well-suited for rumor detection.

2 Related Work

The authors discuss previous approaches to rumor detection on Twitter, which can be broadly categorized into feature-based and deep learning-based methods. Feature-based methods typically rely on extracting hand-crafted features from the text, such as lexical, syntactic, and semantic features, as well as metadata like user information and social network features. These features are then used to train traditional machine learning classifiers like logistic regression, support vector machines, and random forests.

On the other hand, deep learning-based methods employ neural networks, such as convolutional neural networks (CNNs) and recurrent neural networks

(RNNs), to automatically learn features from the text. Among RNNs, long short-term memory (LSTM) networks have been widely used for rumor detection due to their ability to capture long-range dependencies in the text. However, these methods may not fully exploit the contextual information present in the text.

3 BERT for Rumor Detection

The authors argue that BERT, a transformer-based pre-trained language model, is well-suited for rumor detection tasks due to its ability to capture contextual information and learn complex language patterns. BERT is pre-trained on a large corpus of text using masked language modeling and next sentence prediction tasks, which allows it to learn rich semantic representations.

To adapt BERT for the task of rumor detection, the authors fine-tune the model on a dataset containing tweets labeled as either rumor or non-rumor. The fine-tuning process involves training the model for a few epochs on the labeled dataset with a classification layer added on top of the pre-trained BERT model.

4 Experimental Results

The authors evaluate the performance of the fine-tuned BERT model on the rumor detection task and compare it with other popular machine learning and deep learning models, including logistic regression, support vector machines, random forests, and LSTM networks. The evaluation metrics include accuracy, precision, recall, and F1 score.

The experimental results demonstrate that the BERT-based classification system outperforms other models in detecting rumors on Twitter. The authors attribute this performance improvement to BERT's ability to capture contextual information and leverage pre-training on a large corpus of text.

5 Conclusion

The paper highlights the effectiveness of using a fine-tuned BERT model for rumor detection on Twitter. The BERT-based classification system outperforms traditional feature-based and deep learning-based methods, showcasing its potential for addressing misinformation and rumor detection in social media platforms. Future research could explore the applicability of this approach to other platforms or investigate the integration of additional features, such as user information and social network characteristics, to further improve the model's performance.