Shedding Light on Disinformation: Detecting Bangla Fake News using low rank adaption of large language model

KAZI KAMRUZZAMAN RABBI, BRAC University, Bangladesh
PRANTO DEV, BRAC University, Bangladesh
TASNIM FUYARA CHHOAN, BRAC University, Bangldesh
HUSNE MUBARAK, BRAC University, Bangladesh

The drastic spread of fake news has created a significant challenge in the modern era. In this work, the issue of traditional detection of fake news, especially Bangla fake news has been addressed and the implementation of modern Natural Language Processing models such as BanglaBERT and Low-Rank Approximation (LoRA) has shown promising performance in detecting Bangla fake news effectively. Furthermore, our approach streamlines parameter optimization and enhances efficiency in classification tasks. This study contributes to mitigating the adverse effects of disinformation in low-resource languages and lays a foundation for future research endeavors in this domain.

ACM Reference Format:
Kazi Kamruzzaman Rabbi, Pranto Dev, Tasnim Fuyara Chhoan, and Husne Mubarak. 2018. Shedding Light on Disinformation: Detecting Bangla Fake News using low rank adaption of large language model. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation emai (Conference acronym 'XX).* ACM, New York, NY, USA, 6 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

The rapid spread of information via digital channels has made false news a major social problem in an era where information is shared quickly. Like many other countries, Bangladesh is susceptible to the negative consequences of disinformation, which can cause panic, sway elections, and undermine public confidence in institutions. The population that speaks Bangla is consuming more and more news online, hence it is critical that there be efficient systems in place to identify false news. Because of the grammatical uniqueness of Bangla and the lack of extensive datasets for model training, traditional approaches for fake news identification were frequent failures. With the developments in machine learning and natural language processing (NLP) present viable solutions to this problem. The use of Low-Rank Approximation (LoRA) approaches is one such strategy that is gaining traction; these techniques have shown to be useful in spotting patterns and abnormalities within massive datasets. Large language models have been performing exceptionally well for most of NLP related tasks. These models are very capable in terms of sentiment analysis. Fake news detection is sentiment analysis related work. By leveraging the power LLM using LoRA we can make a better fake news detection model.

Authors' Contact Information: Kazi Kamruzzaman Rabbi, BRAC University , Dhaka, Bangladesh, kazi.kamruzzaman.rabbi@g.bracu.ac.bd; Pranto Dev, BRAC University , Dhaka, Bangladesh, pranto.dev@g.bracu.ac.bd; Tasnim Fuyara Chhoan, BRAC University, Dhaka, Bangldesh, tasnim.fuyara.chhoan@g.bracu.ac.bd; Husne Mubarak, BRAC University , Dhaka, Bangladesh, husne.mubarak@g.bracu.ac.bd.

Manuscript submitted to ACM

2 LITERATURE REVIEW

Peng, Yao, et al. [10] proposed a PLoRA method that enhances personalized human-centered text understanding (HCTU) tasks by combining PEFT-based LoRA and knowledge-injected PKI, introducing few-shot learning, and a Plug and Play (PnP) framework. Its methodology involves mapping user input text to sentiment scores, utilizing LoRA for optimization, and incorporating dropout and mutual information maximization techniques. PLoRA shows promise for personalized AI experiences without data leaving the user's vicinity. However, its performance beyond sentiment analysis needs further evaluation, and limitations include untested performance on complex NLP tasks and concerns regarding information reliability and model hallucinations.

Moreover, Hussain, Md Gulzar, et al. [9] aims to detect fake news in Bangla language using deep learning models including bidirectional GRU, LSTM, 1D CNN, and hybrid architectures. It contributes to creating Bangla fake news detection systems with limited resources, improving accuracy, addressing class inequality, and bridging challenges unique to the Bangla language. Methodology involves using large datasets, WordCloud visualization, and specialized libraries for linguistic content. The proposed model achieves high accuracy, precision, recall, and F1 score, but faces limitations in computational cost and real-life application. To overcome limitations, the study proposes expanding the dataset and collecting more pseudo-events for improved model flexibility and real-world applicability.

The study [5] investigates the detection of fake news in Bangla language, addressing the limited research in this area and the prevalence of fake news on social media. It applies Support Vector Machine (SVM) and Multinomial Naïve Bayes (MNB) classifiers to Bangla social media, contributing to research on fake news detection in the language. The methodology involves using a dataset of Bangla news articles, splitting it for training and testing, and employing feature extraction methods like Count Vectorizer and TF-IDF Vectorizer. However, SVM with a linear kernel outperforms MNB, achieving 96.64% accuracy. Still the study has several limitations including reliance on a specific dataset and the focus on only two classifiers, potentially limiting generalizability and applicability. Future research could explore hybrid classifiers and larger datasets for more robust fake news detection in Bangla.

Another group of researchers [1] addressed the challenge of detecting fake news in low-resource languages like Bangla by combining newly collected data with existing datasets and employing various machine learning (ML), deep learning (DL), and transformer models. Motivated by the scarcity of resources and language processing tools, the study aims to provide a benchmark dataset and explore state-of-the-art methods for fake news detection. The authors contribute by creating a Bangla fake news dataset with 4678 distinct articles, evenly split between real and fake news, and experiment with multiple classifiers, including SVM, LR, CNN, LSTM, and transformer models like mBERT and Bangla-BERT. The best-performing models achieve accuracy scores ranging from 93.3% to 95.9%, with significant improvements in recall compared to previous works. However, limitations include the reliance on specific datasets and the need for further exploration of hybrid models and larger datasets. Future research could focus on expanding the dataset and exploring more advanced models for even better performance in detecting fake news in Bangla.

The paper [6] presents an experimental study on spam detection in Bangla text on social media platforms using Multinomial Naïve Bayes (MNB) classifier, achieving an accuracy of 82.44%. It addresses the growing problem of spamming on platforms like Facebook and YouTube, proposing a novel framework for spam detection at the sentence level. By collecting data from popular Bangla Facebook groups and YouTube channels, preprocessing text data, and applying TF-IDF vectorization, the study demonstrates promising results in detecting spam. While the approach

effectively identifies spam, further improvements such as enriching the corpus, exploring combined classifiers, and integrating stemming techniques are suggested for future research to enhance accuracy and address the challenges of maintaining a healthy online environment.

3 METHODOLOGY

3.1 Dataset

The dataset, named BANFakeNews[8], encompasses both authentic and fake news in Bangla. It comprises approximately 48 thousand authentic Bangla news articles sourced from various Bangla newspapers and news portals, along with over 1 thousand fabricated Bangla news articles. We have merged the fake news dataset and authentic news dataset for our purpose. Hence the dataset we worked on had almost 50 thousand Bangla news.

3.2 Dataset Pre-Processing

The dataset [8] preprocessing involved several steps to prepare the text data for model training and testing. First, the text was tokenized using a basic tokenizer to break it into individual words or tokens. Next, common stopwords and punctuation marks were removed to reduce noise in the data. This cleaning process ensured that the text contained only relevant information for sentiment analysis. After pre-processing, the dataset was split into training and testing sets, where 80% of the data was used for training and 20% for testing. This split ensured that the model could be trained on a sufficient amount of data while still having a separate set for evaluation. Finally, the pre-processed datasets were saved as CSV files for further use.

3.3 BanglaBERT

BanglaBERT [2], a pre-trained large language model for the Bangla language, was trained on a custom dataset comprising nearly 2.2 billion tokens. This dataset, totaling 27.5 GB in size, consisted of 5.25 million documents with an average of 306.66 words per document. The model utilized the ELECTRA[3] architecture due to its superior computational efficiency compared to other models. Specifically, the ELECTRA[3] model featured a 12-layer Transformer encoder with a 768 embedding size, 768 hidden size, 12 attention heads, and a 3072 feed-forward size. Trained with a batch size of 256 for 2.5 million steps, the model demonstrated a generator-to-discriminator ratio of 1:3 and comprised 110 million parameters. The authors established a benchmark for the Bangla language, and BanglaBERT performed commendably across all benchmark tests.

3.4 LoRA

Large language models like GPT-3 and BERT boast billions of parameters, necessitating high-performance computers for loading and fine-tuning due to the computational demands. To address this challenge, the concept of Low-rank Adaptation (LoRA) [4] for Large Language Models (LLMs) was introduced. The LoRA approach involves fine-tuning only a specific matrix of the pre-trained models while keeping other layers frozen. This significantly reduces the number of trainable parameters and alleviates the computational burden of fine-tuning. Despite the substantial reduction in trainable parameters, the proposed method demonstrated robust performance, highlighting its effectiveness in optimizing large language models with limited computational resources.

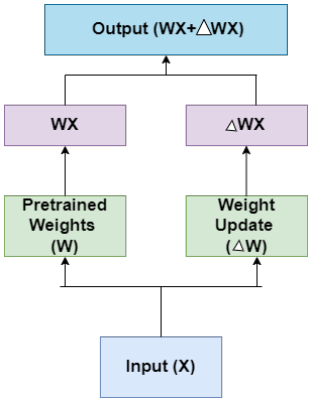

Fig. 1. LoRA architecture

3.5 Our Proposed Method

Given the promising performance of the LoRA[4] method in reducing parameters and training costs for large language models, we sought to leverage this architecture for our classification task using a pre-trained Bangla language model. Specifically, we employed the BanglaBERT model and implemented the LoRA architecture using the Hugging Face PEFT[7] model. By integrating LoRA into the pre-trained BanglaBERT model, we aimed to streamline parameter optimization and enhance efficiency in our classification endeavors.

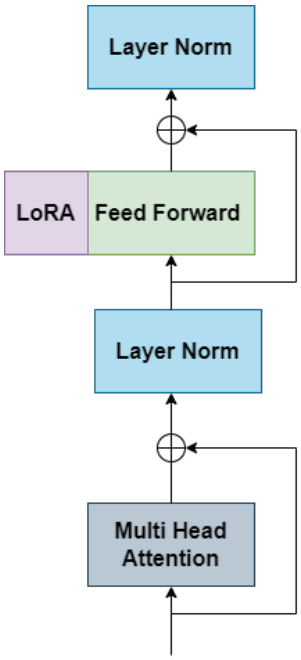

Fig. 2. Proposed Model Architecture with LoRA

For the peft configuration, we used rank dimention of the LoRA injected matrices to 16 and lora_alpha value to 32. For the bert model, our target modules were the query and value matrices. Moreover, we used a dropout of 0.1. After that we converted the banglebert model into a LoRA based banglabert model.

4 RESULT AND ANALYSIS

For the training of our model we divided our dataset into training and test portion where 80% of our data was divided into training set and the rest for test set. As the bert model has large number of trainable parameters it takes a lot of time to train the model even for less amount of epochs. We trained both the banglabert model and the LoRA based banglabert model for 5 epochs. After training the data for 5 epochs, the banglabert model had an accuracy of 95.31% with validation loss of .05% and a training loss of 20.05%, which indicates that our model is underfitting on the data.

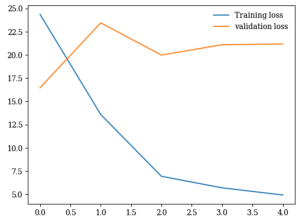

Fig. 3. BanglaBert Model Training and validation loss curve

On the other hand, we trained BanglaBert LoRa based model for 5 epochs and got almost 88% accuracy. Moreover, the training loss was .32% and the validation loss was .31% which shows that our model performed better in lora based model with less number of trainable parameters.

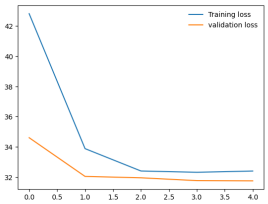

Fig. 4. BanglaBert fine tuned using LoRA Model Training and validation loss curve

The following tables shows the history of the trained models in the last epoch. Moreover, from the graphs we can also say that LoRA based model generalized much better on the data as the banglabert base model was underfitting on our data significantly.

Model	Validation Loss	Training Loss	Accuracy
BanglaBert	.18	.05	96.15%
BanglaBert_LoRA	.32	.31	88%

Table 1. Training History of the trained models

5 CONCLUSION

In conclusion, our study addresses the critical issue of detecting fake news in Bangla language by leveraging advanced machine learning techniques. We utilized the BanglaBERT model and implemented the Low-rank Adaptation (LoRA) architecture to optimize parameter training. Our proposed LoRA-based BanglaBERT model demonstrated promising results, achieving an accuracy of 88% after 5 epochs, outperforming the base BanglaBERT model which achieved 95.31% accuracy but showed signs of underfitting. Moving forward, we plan to compare our model with other existing models and implement more robust approaches. We aim to increase the number of epochs for training to further improve model performance. Additionally, we will explore the potential of hybrid models and larger datasets to enhance the detection of fake news in Bangla. This research contributes to addressing the challenges of disinformation in low-resource languages and provides a benchmark for future endeavors in this area.

REFERENCES

[1] Quazi Adibur Rahman Adib, Md. Humaion Kabir Mehedi, Md. Sadman Sakib, Kabbya Kantam Patwary, Md Sabbir Hossain, and Annajiat Alim Rasel. 2021. A Deep Hybrid Learning Approach to Detect Bangla Fake News. In *2021 5th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*. 442–447. https://doi.org/10.1109/ISMSIT52890.2021.9604712

[2] Abhik Bhattacharjee, Tahmid Hasan, Wasi Ahmad, Kazi Samin Mubasshir, Md Saiful Islam, Anindya Iqbal, M. Sohel Rahman, and Rifat Shahriyar. 2022. BanglaBERT: Language Model Pretraining and Benchmarks for Low-Resource Language Understanding Evaluation in Bangla. In *Findings of the Association for Computational Linguistics: NAACL 2022*. Association for Computational Linguistics, Seattle, United States, 1318–1327. https://aclanthology.org/2022.findings-naacl.98

[3] Kevin Clark, Minh-Thang Luong, Quoc V Le, and Christopher D Manning. 2020. Electra: Pre-training text encoders as discriminators rather than generators. *arXiv preprint arXiv:2003.10555* (2020).

[4] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2021. Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685* (2021).

[5] Md Gulzar Hussain, Md Rashidul Hasan, Mahmuda Rahman, Joy Protim, and Sakib Al Hasan. 2020. Detection of Bangla Fake News using MNB and SVM Classifier. In *2020 International Conference on Computing, Electronics Communications Engineering (iCCECE)*. 81–85. https://doi.org/10.1109/iCCECE49321.2020.9231167

[6] Tanvirul Islam, Subhenur Latif, and Nadim Ahmed. 2019. Using Social Networks to Detect Malicious Bangla Text Content. In *2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)*. 1–4. https://doi.org/10.1109/ICASERT.2019.8934841

[7] Sourab Mangrulkar, Sylvain Gugger, Lysandre Debut, Younes Belkada, Sayak Paul, and Benjamin Bossan. 2022. PEFT: State-of-the-art Parameter-Efficient Fine-Tuning methods. https://github.com/huggingface/peft.

[8] Md Saiful Islam Sudipta Kar Md Zobaer Hossain, Md Ashraful Rahman. 2020. BanFakeNews: A Dataset for Detecting Fake News in Bangla. In *Proceedings of the Twelfth International Conference on Language Resources and Evaluation (LREC 2020)*. European Language Resources Association (ELRA).

[9] Utsha Roy, Mst. Sazia Tahosin, Md. Mahedi Hassan, Taminul Islam, Fahim Imtiaz, Md Rezwane Sadik, Yassine Maleh, Rejwan Bin Sulaiman, and Md. Simul Hasan Talukder. 2024. Enhancing Bangla Fake News Detection Using Bidirectional Gated Recurrent Units and Deep Learning Techniques. (2024). arXiv:2404.01345 [cs.CL]

[10] You Zhang, Jin Wang, Liang-Chih Yu, Dan Xu, and Xuejie Zhang. 2024. Personalized LoRA for Human-Centered Text Understanding. *CoRR* abs/2403.06208 (2024). https://doi.org/10.48550/ARXIV.2403.06208 arXiv:2403.06208