A Comparative Study of Fine-Tuning BERT, BERT+LSTM, and BERT+RCNN for Vietnamese Sentiment Analysis

Hoang Bao Viet, Nguyen Duc Gia Phuc, Nguyen Quang Tuan

The University of Danang, Vietnam - Korea University of Information and Communication Technology

Abstract

Sentiment analysis is a fundamental task in Natural Language Processing (NLP) that focuses on identifying and evaluating opinions expressed in text. This study explores the application of BERT-based models for sentiment analysis of Vietnamese user comments. We investigate and compare three fine-tuning approaches: the standard BERT model using the [CLS] token as input to a feed-forward neural network, a hybrid BERT+LSTM model, and a novel BERT+RCNN architecture that utilizes the full sequence of output vectors for classification. Experiments conducted on two Vietnamese sentiment datasets demonstrate that models leveraging BERT significantly outperform traditional deep learning baselines. Notably, our proposed BERT+RCNN approach consistently achieves the best performance, surpassing the standard BERT fine-tuning method in both accuracy and robustness.

Keywords: BERT, LSTM, RCNN, Vietnamese, Sentiment Analysis

1   Introduction

In today's digital era, millions of user-generated comments are shared daily on social networks and e-commerce platforms. These comments provide valuable feedback for both consumers and service providers. However, due to the vast amount of data, it becomes impractical for humans to manually process and analyze such information. Therefore, there is a growing need for automated systems capable of identifying and evaluating users' opinions.

Sentiment analysis is a fundamental task in Natural Language Processing (NLP) that aims to determine the emotional tone behind textual content. A sentiment classification system can either predict fine-grained sentiment scores (e.g., from 1 to 5) or perform binary classification to detect whether the content expresses a *positive* or *negative* sentiment. In this study, we focus on the latter task.

In 2018, Devlin et al. [1] introduced BERT (Bidirectional Encoder Representations from Transformers), a pre-trained language representation model that has significantly advanced the state of the art across various NLP tasks. BERT captures the context of words by considering both left and right contexts simultaneously, making it particularly effective for understanding the sentiment expressed in text.

This paper investigates the application of BERT-based models to sentiment analysis in Vietnamese. We present and compare two approaches. The first is the original fine-tuning strategy introduced by Devlin et al., which uses the representation of the `[CLS]` token as input to a classification layer. The second approach, proposed in this paper, extends BERT by combining its contextual embeddings with additional neural architectures, including Long Short-Term Memory (LSTM) and Recurrent Convolutional Neural Network (RCNN) layers.

Our objective is to evaluate the effectiveness of these combined models and to demonstrate how different fine-tuning strategies can improve sentiment classification performance on Vietnamese datasets.

The remainder of this paper is organized as follows:

- Section 3 reviews related work in sentiment analysis.

- Section 4 introduces word embeddings, language models, and the BERT architecture.

- Section 5 describes our two BERT-based fine-tuning methods.

- Section 6 presents the experimental setup and results.

- Section 7 concludes the paper.

2 Related Work

3 Related Work

Sentiment analysis is a subtask of text classification in which the goal is to categorize textual data into sentiment classes, typically *positive* or *negative.* Early research in this field primarily relied on traditional machine learning techniques. One of the pioneering studies was conducted in 2002, where reviews were classified into positive and negative categories [2]. This work utilized supervised learning models such as Support Vector Machines (SVM) [3] and Naive Bayes classifiers [4] to perform sentiment classification. Another common approach employed sentiment lexicons, which use predefined dictionaries of emotionally charged words annotated with sentiment polarity and intensity [?].

With the rise of deep learning, sentiment analysis has seen significant advancements through improved word and context representations. Kim [?] proposed the use of Convolutional Neural Networks (CNNs) for sentence classification, treating text as character-level input sequences. Mikolov et al. [?] introduced the Paragraph Vector model (also known as Doc2Vec), an unsupervised method for learning fixed-length representations of variable-length texts. Unlike traditional bag-of-words models, this approach learns document-level embeddings that capture semantic meaning [?].

In the context of Vietnamese language, Duyen et al. [?] applied traditional classifiers such as Naïve Bayes, Maximum Entropy, and SVM to review classification tasks on Agoda, a hotel booking platform. Their results indicated that the SVM model outperformed other approaches. On the deep learning front, Quan et al. [?] proposed a hybrid architecture combining Long Short-Term Memory (LSTM) and CNN, named Multi-Channel LSTM-CNN, for Vietnamese sentiment analysis. This model outperformed standalone CNN and LSTM models. A similar approach was introduced in [?], where word vectors were first processed by a CNN layer, and the resulting feature maps were passed to an LSTM network for final sentiment classification. This pipeline demonstrated strong performance, especially for handling negative comments on social media platforms.

These studies demonstrate the evolution of sentiment analysis methods—from traditional machine learning and lexicon-based approaches to more powerful neural network architectures. However, few works have explored the integration of pre-trained language models like BERT for Vietnamese sentiment classification, which motivates our study.

4   Background

5   Methodology

We fine-tuned all models on the same Vietnamese sentiment dataset. BERT's contextual embeddings were passed directly to classification layers or further processed by LSTM and RCNN architectures.

6   Experiments

Table 1: Performance Comparison on Sentiment Dataset

Model	Precision	Recall	F1 Score
BERT	88%	91%	**90%**
BERT + LSTM	88%	90%	89%
BERT + RCNN	87%	83%	85%

7   Conclusion

While all models performed well, the original BERT model yielded the best F1 score. Adding LSTM or RCNN layers did not significantly improve performance and in some cases led to worse results.

References

[1] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: pre-training of deep bidirectional transformers for language understanding," *CoRR*, vol. abs/1810.04805, 2018. [Online]. Available: http://arxiv.org/abs/1810.04805

[2] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? sentiment classification using machine learning techniques," in *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP 2002)*. Association for Computational Linguistics, Jul. 2002, pp. 79–86. [Online]. Available: https://aclanthology.org/W02-1011/

[3] B. M. and V. B., "Sentiment analysis using support vector machine based on feature selection and semantic analysis," *International Journal of Computer Applications*, vol. 146, pp. 26–30, 07 2016.

[4] L. Dey, S. Chakraborty, A. Biswas, B. Bose, and S. Tiwari, "Sentiment analysis of review datasets using naive bayes and k-nn classifier," *arXiv preprint arXiv:1610.09982*, 2016.