Fine Tuning BERT On Sentimental Analysis

INFO-H519 NLP with Deep Learning SP 2024 - Project Report

Hari Shivani Gudi | Hasararanga Jayathilake |Hymavathi Gummudala
Luddy School Of Informatics, Computing, And Engineering | Indiana University Indianapolis | Prof. Ming Jiang

Abstract

This research delves into sentiment analysis using BERT models, emphasizing the significant language-processing advancements for English and Vietnamese. It builds upon the groundwork laid by previous studies, outlining a focused objective to refine sentiment detection and explore multilingual capabilities. The study employs fine-tuning BERT models, notably PhoBERT for Vietnamese and mBERT for English, to assess their adaptability and efficiency. The methodology includes systematic hyperparameter adjustments and integrating LSTM to enhance model performance across diverse linguistic datasets. The research utilizes datasets like IMDB for English and a local Vietnamese dataset for sentiment analysis. Experiments are meticulously designed with a balanced approach to training and testing, ensuring rigorous evaluation of model effectiveness against set benchmarks. PhoBERT outperforms mBERT, particularly when embedding with LSTM, demonstrating higher recall and precision across languages. The findings advocate for targeted models in handling language-specific nuances, with broader implications for enhancing automated sentiment analysis tools in various applications. Future work will expand on refining these models and exploring their scalability across additional languages with utilization of PhoBERT model as a machine translation purposes.

Table of Contents

Table of Figures

Introduction

Customer reactions and reviews are the one of the key indicators for the organizations to understand on how good or bad their present offerings to the market (Allegrino et al., 2019; Chiang, 2021). Most importantly reviews are helping to shape the organizations to align their future market directions according to the customer taste which may help them to avoid the unnecessary spending on R&D[1]s that are not attractive for the gain momentum in the future market or setting up new R&D projects align with the future market demands to expand their customer base with first mover advantage (Abakouy et al., 2019; Bleier et al., 2020). And even, reviews share the present market offering modifications that leads to gain more customer attraction (Bleier et al., 2020). Therefore, in present dynamic VUCA[2] market structure, customer reviews are very important role is playing for the organizations to secure their sustainable competitive advantage over the competitors in the market (Bleier et al., 2020; Braverman, 2015; Rosário & Dias, 2023).

Traditional to understand the customer reactions, needed to conduct market research by using one on one in person surveys, but with the help of the technology advancement (Braverman, 2015). presently, widely utilizing sentiment analysis mechanism under the NLP[3] techniques to understand and monitor the customer's reviews on the products by the companies, which save time for the companies to be more productive on their offerings to represent the exact customer needs and wants(Chiang, 2021).

According to Srivastava et al, (2024), sentiment analysis deals with the computational study of people's opinions, sentiments, attitudes, and emotions expressed in written text. The primary goal of sentiment analysis is to classify text data into different sentiment categories, such as positive, negative, or neutral (Loc et al., 2023).

Moreover, classification tasks under the sentiment analysis can be performed at different levels, including document-level, sentence-level, or aspect-level (Van Thin et al., 2023b). Even more, apart from the level of the classification, based on the approaches also can be dived the sentimental analysis into 2 categorises as rule-based approach where uses predefined rules and lexicons to classify text as positive, negative, or neutral, and machine learning approach which utilizes algorithms to automatically learn and classify sentiment from labelled training data (Cho et al., 2014; Van Thin et al., 2022, 2023c).

Therefore, in this research, the primary focus and objective has been set as **on utilizing a machine learning approach embedded with sentence-level sentiment analysis classification**. Since machine learning models are not as rigid as traditional approaches, they have the ability to capture contextual information and understand the nuances of language in a holistic

[1] R&D – Research and Development

[2] VUCA - Volatility, Uncertainty, Complexity, And Ambiguity

[3] NLP – Natural Language Processing

manner, given that most words contain multiple meanings, unlike the predefined rules in traditional approaches (Devlin et al., 2018; Q. N. Nguyen et al., 2024).

Beyond technical aspects, the importance of language has also been considered in selecting languages for this research. The focus is primarily on Vietnamese and English. Vietnamese is widely used by the residents of Vietnam and is considered the 21st most spoken language globally, with about 90 million speakers, mainly in Vietnam (T.-N. N. Le & Trofimovich, 2023).

However, due to Vietnam's rapidly growing economy and its appealing climate, which has made it an affordable hub for industries such as footwear and electronics manufacturing and has drawn millions of tourists in recent years, English has become essential (Tan Phat, 2015). English is the primary language used in international trade and tourism worldwide (Selvi et al., 2023). Therefore, the ability to translate and the level of English understanding among the Vietnamese are crucial factors (Revilla Diez, 2016).

In recent years, research has expanded, and Vietnamese language models have been developed to improve machine-level comprehension of the Vietnamese language (Van Thin et al., 2023a). This advancement may foster an ecosystem that leverages the vast amount of Vietnamese data for better decision-making and the creation of products more tailored to Vietnamese consumers (D. Q. Nguyen & Nguyen, 2020; Van Thin et al., 2023c). Consequently, this research focuses on sentiment analysis, which is crucial for utilizing feedback from diverse linguistic markets to maintain a competitive advantage (Van Thin et al., 2023b).

This study aims to refine sentiment analysis models for both English and Vietnamese, recognizing the growing importance of sentiment analysis in data-driven markets. It tackles the challenges associated with linguistic subtleties and the need for extensive training data while exploring opportunities in multilingual model development and cross-lingual transfer learning.

Prior Related Work (Literature Review)

Development of BERT and Its Impact: Devlin et al., (2018) (Google AI researchers) have introduced the Bidirectional Encoder Representations from Transformers (BERT) model which pre-training language representations using bidirectionally on large text corpus to enhance the sentence level understanding from words within the sentence. Moreover in 2018, Google AI has introduced the BERT Multilingual (mBERT) model which able to generate responses in 104 languages, and it has included the low resource langue of Vietnamese(Van Thin et al., 2023d). Furthermore, mBERT model is capable to perform tasked with zero-shot learning mechanism whereas model finetunes on the specific task on one language and able to perform the same task on another language without task training on new language (Van Thin et al., 2022). Therefore, mBERT model widely utilized for the multi-language task solution development aspects, especially, from the sentimental analysis-based application development segment, where mBERT enabled to cross-lingual transfer leaning to enhance the accuracy of the application while simplifying the development process with scalability to relate to the development process for global applications.

Moreover, Nguyen & Nguyen, (2020) released the pre-trained Vietnamese language model which developed using RoBERTa architecture to optimize mBERT model to enhance the RoBERTa output on the Vietnamese language representation applications. And it becomes the one of the dominate Vietnamese language model in present context (Van Thin et al., 2023a, 2023b).

Transfer learning on BERT model: Even, in terms of the Vietnamese language, mBERT is widely utilized as its original form and finetune manner as PhoBERT on aspect category and aspect category detection type sentiment analysis works (Q. N. Nguyen et al., 2024; Van Thin et al., 2023a). Importantly, Le et al., (2020) have been shown that, effectiveness and capability of utilization of mBERT to aspect category sentiment understanding on Vietnamese datasets in that work and gave the guidelines on how to combine the mBERT in effectively with different preprocessing structures for the Vietnamese datasets. Moreover, Dang et al., (2022) illustrate the on impact of utilizing PhoBERT model for aspect category sentiment analysis the same dataset as Le et al., (2020) proposed.

Multilingual and Cross-Lingual Challenges: recent literature has identified that, mBERT model is having low F1 scores on the sentimental analysis applications when it utilized on the low resource language representation applications (D. Q. Nguyen & Nguyen, 2020; Van Thin et al., 2023c). B. H. Le et al., (2022) have identified that main drawback of the mBERT, which has developed using high resource languages predominately that leads to reduce performance of the model on the low resources' languages like Vietnamese.

Integration of BERT with Other Models: Si et al., (2021) demonstrated that integrating BERT with deep learning architectures like Long Short-Term Memory (LSTM) networks can significantly enhance text classification capabilities for sentiment analysis tasks. Additionally, combining BERT with other deep learning models, such as Convolutional Neural Networks (CNNs), has been shown to elevate performance in text analysis (Loc et al., 2023).

Effectiveness of Fine-Tuning on Domain-Specific Datasets: Furthermore, D. Q. Nguyen & Nguyen, (2020 and Van Thin et al., (2023c) have identified that utilization of the domain specific dataset on the training purpose would enhance the task relevant optimized output from the mBERT and PhoBERT models. Q. N. Nguyen et al., (2024) shown that utilization android application permission dataset for train the sentiments on same type of application representations where are shown that able to achieve better accuracies compare to non-relevant dataset training mechanisms. Furthermore, fine-tuning BERT on domain-specific datasets has been found to significantly boost performance by aligning the model's focus with contextual nuances overcome the cross-lingual challenges (Loc et al., 2023; D. Q. Nguyen & Nguyen, 2020; Si et al., 2021).

Research Gap and Novelty of the Research: Despite these advancements, there are limitations and gaps in existing research. For instance, there is a lack of direct comparisons and comprehensive evaluations of models specialized for the Vietnamese language, such as PhoBERT (Nguyen et al., 2020). Few studies have analysed how different models handle the nuances of the Vietnamese language, and there is a need for a broad assessment across a range of metrics for Vietnamese datasets. Additionally, limited comparisons with other languages, such as English, restrict the measurement of model generalizability, and a lack of cross-language insights hinders the understanding of model adaptability and robustness to utilize the PhoBERT model as machine translation model for the Vietnamese langue with any other languages.

Table 1: Literature Review Summary on main articles utilized for the study.

Papers	Methods Used
Le et al., (2020)	Transfer learning on mBERT model. (mBERT – F1 Score – 67.74%
Dang et al., (2022)	Transfer learning on PhoBERT model. (PhoBERT – F1 Score - 83.92%)
B. H. Le et al., (2022)	Multilingual and Cross-Lingual Challenges (mBERT – F1 Score – 40.64% and PhoBERT – F1 Score – 78.76% for Sentimental Detection Tasks.)
Si et al., (2021)	Integration of BERT with Other Models (mBERT + BiLSTM+ CNN - F1 Score - 87.05%
Loc et al., (2023)	Integration of BERT with Other Models (PhoBERT + CNN - F1 Score - 91.43%)
Q. N. Nguyen et al., (2024)	Fine-Tuning on Domain-Specific Datasets (Finetuned without Domain specific data - mBERT – F1 Score – 81.59% and PhoBERT – F1 Score – 83.29% and Finetuned with domain specific on PhoBERT - F1 Score - 85.10%)

The literature review underscores the impact of BERT in NLP and its evolution through multilingual support. Previous studies highlight the effectiveness of BERT in multiple languages and its integration with deep learning architectures like LSTM to improve sentiment analysis. However, there exists a gap in direct comparisons of Vietnamese-specific models and broad assessment across metrics with using high resource language (English) sentimental analysis applications.

Approach (Methodology)

The method involved several essential steps to ensure strong performance of the model. The process started with the careful selection of pre-trained models as the base for the sentiment analysis system. Models like PhoBERT, known for its effectiveness with the Vietnamese language, and BERT base multilingual model, which supports multiple languages including Vietnamese, were chosen(Dang et al., 2022; Devlin et al., 2018; Q. N. Nguyen et al., 2024; Van Thin et al., 2023a). Preparing the dataset was handled with great care to maintain data quality by removing repeated entries and dealing with missing values (B. H. Le et al., 2022). The data was then split into training and testing groups, 80% for training and 20% for testing, to help in solid training and assessing the model (Van Thin et al., 2023c). The understanding of text by the models was enhanced by using different embedding methods like BERT for contextual understanding, GloVe for tracking word relationships, and FastText for efficient word representation (D. Q. Nguyen & Nguyen, 2020; Van Thin et al., 2023c). The strategy for training and testing the model included repeated improvements, beginning with tests on a small part of the dataset to set initial performance levels and adjust the system and settings (Van Thin et al., 2023b). When training on the full dataset, issues like system instability and limited computing power were managed by using resources like University Big Red and effective problem-solving. The final step was to apply the trained models to new data to predict sentiments, demonstrating their usefulness and confirming the success of the method with practical examples. Below table illustrates summary of approaches utilized for research.

Table 2: Summary of Approaches Utilized For Research

Description	Activities Performed
Selecting a Pre-trained Model	Selected mBERT[4], mBERT+LSTM[5], PhoBERT[6]
Preprocess and splitting the dataset	No duplicates found, dropped null values. And the data was split into an 80-20 ratio.
Embeddings	Glove, Fasttext, and BERT contextualized embedding were used
Types of Models	Recreated the models as per the original work.
Train & Evaluating Model	Based on 2 datasets training initiated first. Followed by the testing and evaluation outputs of the models respectively.
Predict the sentiment label	Test the model on understanding using samples.

Data

Two distinct datasets are employed to develop and assess sentiment analysis models. The IMDB movie review dataset, widely recognized as a benchmark in this field, contains 50,000 movie reviews. Each review is labeled for sentiment polarity, with 0 indicating negative and 1 positive sentiments. Its binary classification system simplifies crafting and testing algorithms (Rashid, 2022).

Additionally, a dataset designed for sentiment analysis tasks in the Vietnamese language is utilized, focusing on hotel and food reviews. It comes pre-divided into training and testing subsets, ensuring efficient model training and evaluation. The dataset includes reviews in Vietnamese, with sentiment polarity marked similarly to the IMDB dataset, addressing the need for language-specific data in model development (Linh, 2020).

Table 3: Summary of Datasets used in the study

<table>
<tr><th>Descriptions</th><th>English IMDB Dataset</th><th colspan="2">Vietnamese Dataset</th></tr>
<tr><td>Content</td><td>Movie Review descriptions</td><td colspan="2">Hotel Food Review Descriptions</td></tr>
<tr><td rowspan="2"># of Rows of Data</td><td rowspan="2">50000 Rows of Text Data
(Used 80:20 rule on split data</td><td>Train Dataset</td><td>Test Dataset</td></tr>
<tr><td>40761 Rows</td><td>10000 Rows</td></tr>
<tr><td>Sentimental category</td><td colspan="3">Binary (2 Labels)</td></tr>
<tr><td>Label: 0 Definition</td><td colspan="3">Negative</td></tr>
<tr><td>Label: 1 Definition</td><td colspan="3">Positive</td></tr>
</table>

Data distribution has been done as per the below manner for in terms of the 2 datasets.

[4] mBERT – source - https://huggingface.co/google-bert/bert-base-multilingual-cased

[5] mBERT+LSTM – source - https://arxiv.org/abs/2011.10426

[6] PhoBERT – source - https://huggingface.co/vinai/phobert-base

Table 4: Data Distribution of The Selected Dataset for The Study

Dataset	Train		Test	
	Positive	Negative	Positive	Negative
IMDB	19961	20039	5039	4961
Vietnamese	20493	20268	5000	5000

Experiments

Robust training protocols involving multiple epochs and cross-validation were established to ensure model reliability (N. C. Le et al., 2020). Both standard and fine-tuned versions of BERT were evaluated, assessing the impact of fine-tuning on sentiment analysis performance (Van Thin et al., 2023a, 2023b).

LSTM layers were integrated into models to effectively capture sequential patterns in text data (Si et al., 2021). Architectural adjustments optimized configurations for sentiment analysis tasks. Comprehensive evaluation metrics, including accuracy and F1-score, assessed model performance, offering insights into the efficacy of methodologies (N. C. Le et al., 2020; Loc et al., 2023; D. Q. Nguyen & Nguyen, 2020; Van Thin et al., 2023c).

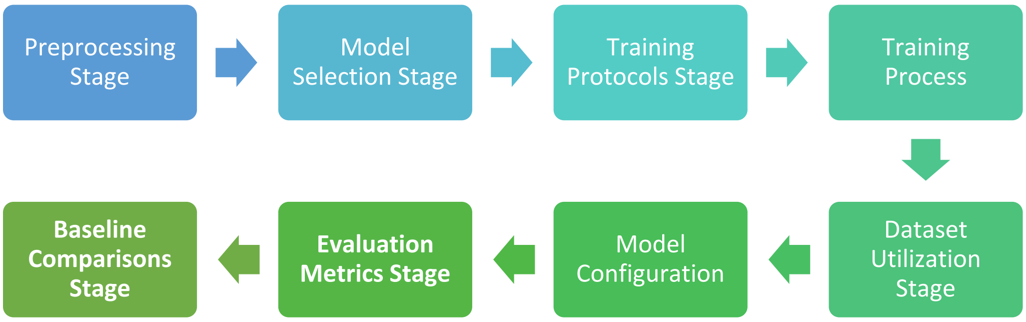


Figure 1: Workflow of the experiment of the research.

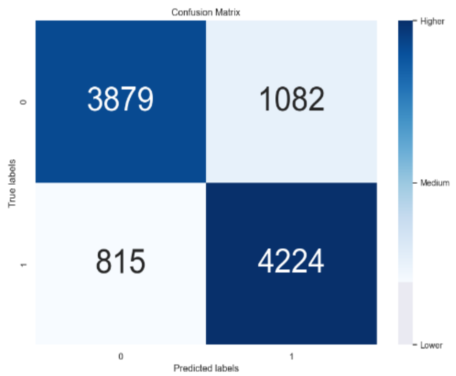
Results

In the research, 5 epoch has been run on each model to receive the below results from each model. In evaluating sentiment analysis models across Vietnamese and English datasets, distinct patterns emerged. The PhoBERT model, when combined with LSTM, achieved the highest accuracy at 93.24% for Vietnamese. This configuration also excelled in precision and recall, marking a significant improvement over the base models. For English, however, the results were more moderate, with the Fine-Tune mBERT-Base + LSTM model leading in recall at 89.90% and the PhoBERT + LSTM model achieving the top F1 score of 83.19%. The use of LSTM layers generally resulted in performance enhancements, suggesting their effectiveness in capturing linguistic nuances across languages.

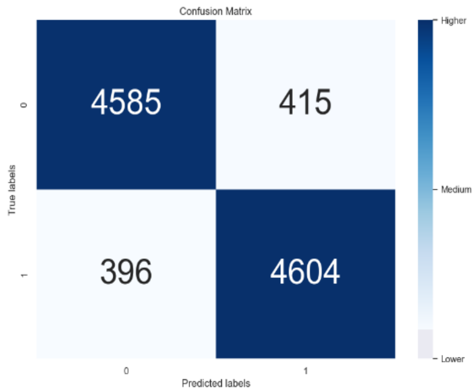
Table 5: Result Summary of The Study

Models (epoch = 5)	Datasets							
	Vietnamese				English			
	Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall	F1
mBERT-Base	0.9113	0.8744	0.9080	0.8875	0.8103	0.7961	0.8382	0.8162
mBERT-Base + LSTM	0.7564	0.7459	0.7806	0.7629	0.6613	0.6301	0.7735	0.6945
Fine-Tune mBERT-Base	0.8296	0.8097	0.8618	0.8349	0.8103	0.7961	0.8383	0.8166
Fine-Tune mBERT-Base + LSTM	0.8515	0.7737	0.8547	0.8122	0.8590	0.8525	0.8990	0.8757
PhoBERT	0.9189	0.9208	0.9171	0.9189	0.8072	0.8360	0.8091	0.8223
PhoBERT + LSTM	0.9324	0.8959	0.9532	0.9236	0.8508	0.8360	0.8277	0.8319

mBERT-Base Confusion Matrix



Fine-Tune mBERT Confusion Matrix



PhoBERT Confusion Matrix

The study presents confusion matrices for Vietnamese language sentiment analysis using various models. The mBERT-Base model demonstrated balanced classification, while the Fine-Tune mBERT model showed a higher misclassification rate. PhoBERT outperformed both, with the least misclassifications, indicating its superior capability in understanding Vietnamese text. This comparative analysis highlights PhoBERT's robustness and potential as a reliable tool for Vietnamese language processing tasks.

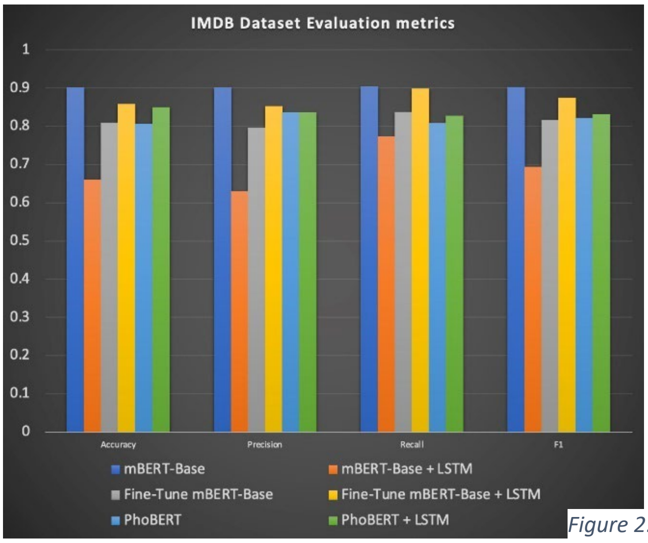


Figure 2: English Dataset metrics results.

For the IMDB dataset, the bar chart illustrates that PhoBERT and its LSTM-enhanced variant deliver superior performance across all metrics. PhoBERT with LSTM, in particular, achieves the highest marks, underscoring the effectiveness of integrating LSTM for sentiment analysis in English.

Turning to the Vietnamese dataset, the graphical comparison reveals that PhoBERT, both in its base form and when augmented with LSTM, consistently outperforms the mBERT variants. The addition of LSTM to PhoBERT notably enhances recall, indicating its proficiency in capturing the sentiment of Vietnamese text.

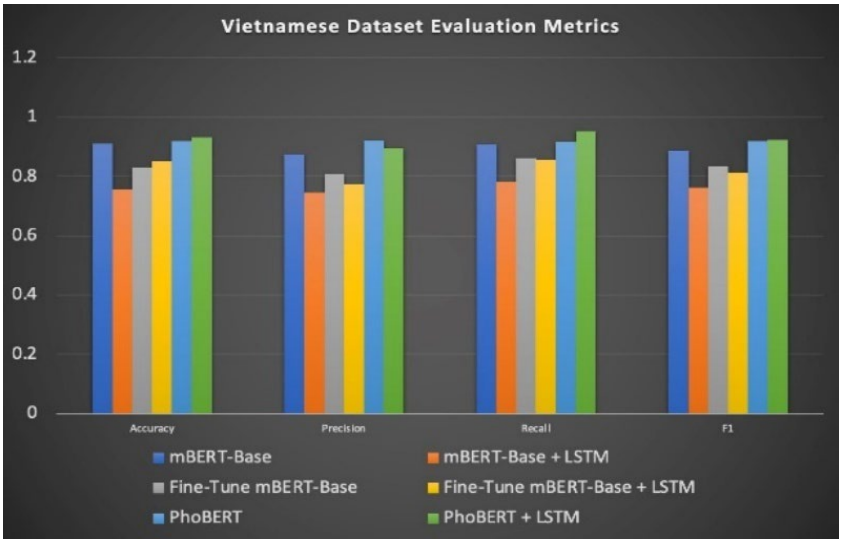


Figure 3: Vietnamese Dataset Metrics Results.

Analysis

PhoBERT Tops BERT: In sentiment analysis tasks for English and Vietnamese, PhoBERT consistently outperformed Google's BERT, demonstrating higher accuracy and F1 scores. The data suggests PhoBERT's design better captures the linguistic features of both languages, resulting in more accurate sentiment predictions.

LSTM Enhances Recall: Adding LSTM layers to models generally resulted in improved recall, meaning the models became better at identifying positive sentiments. However, this benefit was sometimes at the cost of precision, where models might mistakenly label negative sentiments as positive.

PhoBERT LSTM Excels: PhoBERT combined with LSTM stood out, showing the strongest performance overall. Its F1 score, which balances precision and recall, was notably higher than others. This indicates a well-rounded model that accurately identifies sentiments without sacrificing the ability to discern between positive and negative .

Model Effectiveness

PhoBERT's Cross-Lingual Effectiveness: PhoBERT's high performance across both Vietnamese and English datasets highlights its effectiveness, outshining its peers in cross-lingual sentiment analysis tasks and proving its versatility (Si et al., 2021).

LSTM Boosts Sentiment Identification: The BERT models, especially those enhanced with LSTM, show a strong ability to identify relevant sentiments, which is critical for accurate sentiment analysis across different contexts (D. Q. Nguyen & Nguyen, 2020).

LSTM Enhances Sentiment Detection: Incorporating LSTM into the models has universally improved their ability to detect sentiment nuances, demonstrating its value as an enhancement to sentiment analysis models in varied linguistic datasets (Van Thin et al., 2023c).

Conclusion

Challenges and Limitations

PhoBERT Beats mBERT: In tasks tailored to specific languages, it was observed that PhoBERT surpassed multilingual BERT in effectiveness, signifying that models dedicated to a single language better capture its nuances (N. C. Le et al., 2020; Van Thin et al., 2023b).

LSTM Yields Mixed Results due to potential overfitting: The integration of LSTM led to mixed performance outcomes, indicating a potential for overfitting or bias in sentiment analysis, especially when the number of training epochs was lowered (Van Thin et al., 2022).

Computational Limits Impact: Computational resource limitations necessitated a reduction in epochs as below table manner, restricting the scope for extensive fine-tuning and thus impacting the models' depth of training and optimization capabilities (Q. N. Nguyen et al., 2024).

Table 6: Epochs Limitations on Operationalizing The Study

Model Name	Original Optimization Parameters			In this Project Optimization Activities		
	Epochs	Max Length	#params	Epochs	Max Length	#params
mBERT	40	512	179M	5	512	179M
mBERT + LSTM	10	256	179M	5	256	179M
PhoBERT	40	256	135M	5	256	135M

Theoretical Implications

PhoBERT Model Superiority: The study identified Vietnamese-specific models outperforming the BERT multi-model in sentiment analysis accuracy, notably with limited training cycles. Wit just using five epochs sufficed to exhibit superior language understanding, highlighting the effectiveness of tailored model architecture in natural language processing tasks (Dang et al., 2022).

Language-Targeted Model Importance: Results from this analysis reinforce the significance of creating language-specific models for sentiment analysis. These models have proven to capture linguistic details and sentiments more accurately than their multilingual counterparts, which is pivotal for nuanced language understanding (Van Thin et al., 2022).

LSTM's Role in Sentiment Analysis: The effectiveness of LSTM in the study implies that processing sequential data is a critical component of sentiment analysis. Its ability to retain information from previous inputs is instrumental in understanding and predicting the sentiment of textual data accurately (Devlin et al., 2018; Q. N. Nguyen et al., 2024).

Contextual Embeddings in NLP: PhoBERT's strong performance underscores the value of contextualized embeddings within the NLP domain. This suggests that the model's nuanced understanding of context and language greatly enhances its ability to discern sentiment, leading to more accurate sentiment analysis results.

Utilizing PhoBERT in Translation Tasks: Given PhoBERT's demonstrated high accuracy, its potential utility in machine translation between English and Vietnamese is promising (L. T. Nguyen & Dien, 2017). The results suggest that PhoBERT's advanced understanding of context and language nuances could significantly benefit translation accuracy. Further research is recommended to explore PhoBERT's capabilities within parallel corpora for machine translation, aiming to enhance the fidelity of translated content and improve cross-lingual communication (Q. N. Nguyen et al., 2024).

Practical Applications

PhoBERT's Translation Potential: Given PhoBERT’s exceptional performance in sentiment analysis, its application to machine translation is promising. The high accuracy suggests it could significantly enhance translation quality between English and Vietnamese, meriting further investigation (Van Thin et al., 2022).

Enhanced Sentiment Analysis: PhoBERT’s success indicates its capability to deeply understand context, making it ideal for sentiment analysis applications where nuanced language comprehension is key (Q. N. Nguyen et al., 2024).

Advanced Language Processing: The results imply that PhoBERT can serve as a powerful tool for complex language processing tasks, potentially improving the accuracy of automated translation systems and natural language understanding applications (Devlin et al., 2018).

Future Work and Improvements

Optimizing Hyperparameter Tuning: Future studies could focus on refining hyperparameters to potentially boost the accuracy of models like PhoBERT, especially in domain-specific sentiment analysis tasks (Dang et al., 2022; Van Thin et al., 2023a).

Expanding Linguistic Model Research: There's a need to test the current models across more varied languages and domains, to truly assess their versatility and performance in broader NLP applications (Van Thin et al., 2023a, 2023b).

Creating More Accessible Models: Developing streamlined models could mitigate computational barriers, thus making advanced NLP tools more accessible for research and practical applications in diverse environments (Q. N. Nguyen et al., 2024; Van Thin et al., 2022).

Investigating Vietnamese Model Efficacy: Intensive experimentation with higher iteration counts is warranted to better understand the performance gap between Vietnamese-trained models and multilingual models in sentiment analysis (Van Thin et al., 2023e).

To access the codes and the data set that utilized for this work use GitHub Link: Click Here.

Reference

Abakouy, R., En-Naimi, E. M., El Haddadi, A., & Lotfi, E. (2019). Data-driven marketing: How machine learning will improve decision-making for marketers. *ACM International Conference Proceeding Series*. https://doi.org/10.1145/3368756.3369024

Allegrino, F., Gabellini, P., Di Bello, L., Contigiani, M., & Placidi, V. (2019). The Vending Shopper Science Lab: Deep Learning for Consumer Research. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11808 LNCS*, 307–317. https://doi.org/10.1007/978-3-030-30754-7_31

Bleier, A., Goldfarb, A., & Tucker, C. (2020). Consumer privacy and the future of data-based innovation and marketing. *International Journal of Research in Marketing*, *37*(3), 466–480. https://doi.org/10.1016/j.ijresmar.2020.03.006

Braverman, S. (2015). Global review of data-driven marketing and advertising. *Journal of Direct, Data and Digital Marketing Practice*, *16*(3), 181–183. https://doi.org/10.1057/DDDMP.2015.7

Chiang, W. Y. (2021). Using a data-driven marketing strategy on customer relationship management: an empirical case of urban coffee shops in Taiwan. *British Food Journal*, *123*(4), 1610–1625. https://doi.org/10.1108/BFJ-06-2020-0523

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. *EMNLP 2014 - 2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, 1724–1734. https://doi.org/10.3115/v1/d14-1179

Dang, H. Q., Nguyen, D. D. A., & Do, T. H. (2022). Multi-task Solution for Aspect Category Sentiment Analysis on Vietnamese Datasets. *Proceedings - 2022 IEEE International Conference on Cybernetics and Computational Intelligence, CyberneticsCom 2022*, 404–409. https://doi.org/10.1109/CYBERNETICSCOM55287.2022.9865479

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, *1*, 4171–4186. https://arxiv.org/abs/1810.04805v2

Le, B. H., Nguyen, H. M., Nguyen, N. K. P., & Nguyen, B. T. (2022). A New Approach for Vietnamese Aspect-Based Sentiment Analysis. *Proceedings - International Conference on Knowledge and Systems Engineering, KSE*, *2022-October*. https://doi.org/10.1109/KSE56063.2022.9953759

Le, N. C., The Lam, N., Nguyen, S. H., & Thanh Nguyen, D. (2020). On Vietnamese Sentiment Analysis: A Transfer Learning Method. *Proceedings - 2020 RIVF International Conference* *on Computing and Communication Technologies, RIVF 2020*. https://doi.org/10.1109/RIVF48685.2020.9140757

Le, T.-N. N., & Trofimovich, P. (2023). Exploring Socio-Political Dimensions of Heritage Language Maintenance: The Case of Vietnamese Speakers in Montréal. *Https://Doi.Org/10.3138/Cmlr-2022-0078*. https://doi.org/10.3138/CMLR-2022-0078

Linh, T. (2020). *Sentiment-Analysis-using-BERT/Data/NTC_SV at master · thoailinh/Sentiment-Analysis-using-BERT*. https://github.com/thoailinh/Sentiment-Analysis-using-BERT/tree/master/Data/NTC_SV

Loc, C. V., Viet, T. X., Viet, T. H., Thao, L. H., & Viet, N. H. (2023). Pre-Trained Language Model-Based Deep Learning for Sentiment Classification of Vietnamese Feedback. *International Journal of Computational Intelligence and Applications*, *22*(3). https://doi.org/10.1142/S1469026823500165

Nguyen, D. Q., & Nguyen, A. T. (2020). PhoBERT: Pre-trained language models for Vietnamese. *Findings of the Association for Computational Linguistics Findings of ACL: EMNLP 2020*, 1037–1042. https://doi.org/10.18653/V1/2020.FINDINGS-EMNLP.92

Nguyen, L. T., & Dien, D. (2017). English-Vietnamese cross-language paraphrase identification method. *ACM International Conference Proceeding Series*, *2017-December*, 42–49. https://doi.org/10.1145/3155133.3155187

Nguyen, Q. N., Cam, N. T., & Van Nguyen, K. (2024). XLMR4MD: New Vietnamese dataset and framework for detecting the consistency of description and permission in Android applications using large language models. *Computers & Security*, *140*, 103814. https://doi.org/10.1016/J.COSE.2024.103814

Rashid, Z. Bin. (2022). *zbrpucstd/Sentiment-Analysis-on-IMDB-Movie-Reviews: In this project, we use machine learning to analyze the sentiment of movie reviews from IMDB. The goal is to classify reviews as positive or negative in order to understand how people feel about a movie.* https://github.com/zbrpucstd/Sentiment-Analysis-on-IMDB-Movie-Reviews

Revilla Diez, J. (2016). Vietnam 30 years after Doi Moi: Achievements and challenges. *Zeitschrift Fur Wirtschaftsgeographie*, *60*(3), 121–133. https://doi.org/10.1515/ZFW-2016-0035

Rosário, A. T., & Dias, J. C. (2023). How has data-driven marketing evolved: Challenges and opportunities with emerging technologies. *International Journal of Information Management Data Insights*, *3*(2), 100203. https://doi.org/10.1016/J.JJIMEI.2023.100203

Selvi, A. F., Galloway, N., & Rose, H. (2023). Teaching English as an International Language. *Elements in Language Teaching*. https://doi.org/10.1017/9781108902755

Si, L., Van, D., Luu-Thuy, N., & Quoc, S. (2021). A multi-filter BiLSTM-CNN architecture for vietnamese sentiment analysis. In H. Marcin, W. Krystian, & S. Edward (Eds.), *In Advances in Computational Collective Intelligence* (pp. 752–763).

Tan Phat, N. (2015). Structural Transformation to Take Off Economy in Developing Nations: Research on Theory and Practice in Vietnam. *Journal of World Economic Research*, *4*(1), 1. https://doi.org/10.11648/J.JWER.20150401.11

Van Thin, D., Hao, D. N., Hoang, V. X., & Nguyen, N. L. T. (2022). Investigating Monolingual and Multilingual BERT Models for Vietnamese Aspect Category Detection. *Proceedings - 2022 RIVF International Conference on Computing and Communication Technologies, RIVF 2022*, 130–135. https://doi.org/10.1109/RIVF55975.2022.10013792

Van Thin, D., Hao, D. N., & Nguyen, N. L. T. (2023a). A Systematic Literature Review on Vietnamese Aspect-based Sentiment Analysis. *ACM Transactions on Asian and Low-Resource Language Information Processing*, *22*(8). https://doi.org/10.1145/3610226

Van Thin, D., Hao, D. N., & Nguyen, N. L. T. (2023b). Vietnamese Sentiment Analysis: An Overview and Comparative Study of Fine-tuning Pretrained Language Models. *ACM Transactions on Asian and Low-Resource Language Information Processing*, *22*(6). https://doi.org/10.1145/3589131

Van Thin, D., Hao, D. N., & Nguyen, N. L. T. (2023c). Vietnamese Sentiment Analysis: An Overview and Comparative Study of Fine-tuning Pretrained Language Models. *ACM Transactions on Asian and Low-Resource Language Information Processing*, *22*(6). https://doi.org/10.1145/3589131

Van Thin, D., Hao, D. N., & Nguyen, N. L. T. (2023d). Vietnamese Sentiment Analysis: An Overview and Comparative Study of Fine-tuning Pretrained Language Models. *ACM Transactions on Asian and Low-Resource Language Information Processing*, *22*(6). https://doi.org/10.1145/3589131

Van Thin, D., Hao, D. N., & Nguyen, N. L. T. (2023e). Vietnamese Sentiment Analysis: An Overview and Comparative Study of Fine-tuning Pretrained Language Models. *ACM Transactions on Asian and Low-Resource Language Information Processing*, *22*(6). https://doi.org/10.1145/3589131