A Novel Framework for Vietnamese Sentiment Classification

Cuong V. Nguyen [a], Khiem H. Le [a], Anh M. Tran [a], Binh T. Nguyen [a,b,c,1],

[a] *AISIA Research Lab*

[b] *University of Science, Ho Chi Minh City, Vietnam*

[c] *Vietnam National University in Ho Chi Minh City, Vietnam*

Abstract

With the booming development of E-commerce platforms in many counties, there is a massive amount of customers' review data in different products and services. Understanding customers' feedbacks in both current and new products can give online retailers the possibility to improve the product quality, meet customers' expectations, and increase the corresponding revenue. In this paper, we investigate the Vietnamese sentiment classification problem on two datasets containing Vietnamese customers' reviews. We propose eight different approaches, including Bi-LSTM, Bi-LSTM + Attention, Bi-GRU, Bi-GRU + Attention, Recurrent CNN, Residual CNN, Transformer, and PhoBERT, and conduct all experiments on two datasets, AIVIVN 2019 and our dataset self-collected from multiple Vietnamese e-commerce websites. The experimental results show that all our proposed methods outperform the winning solution of the competition"AIVIVN 2019 Sentiment Champion" with a significant margin. Especially, Recurrent CNN has the best performance in comparison with other algorithms in terms of both AUC (98.48%) and F1-score (93.42%) in this competition dataset and also surpasses other techniques in our dataset collected. Finally, we aim to publish our codes, and these two datasets later to contribute to the current research community related to the field of sentiment analysis.

Keywords. Bi-LSTM/GRU, Attention, Recurrent CNN, Residual CNN, Transformer, Transfer Learning

1. Introduction

Nowadays, sentiment analysis, which is also known as opinion mining, has become one of the fundamental and essential tasks in multiple research fields, including natural language processing (NLP), data mining (DM), and information retrieval (IR) [32]. Also, it has had a lot of applications in social networks and e-commerce [4]. Many companies (e.g., Facebook, Google, Atlassian, KFC) are continuously monitoring users' feedbacks to extract valuable information from consumers' comments, detect sentiments, wishes, and recommendations regarding the product (new features, bugs, and services) in general and its specific elements. These comments or reviews can be beneficial to provide each company with another perspective on the weak and strong points of each product.

[1] Corresponding Author; E-mail: ngtbinh@hcmus.edu.vn

For example, using sentiment analysis to analyze thousands of reviews about one product automatically could help a company find out whether its customers are happy about a new pricing plan and the current customer service. Subsequently, it can help to investigate errors or bugs, enhance the quality quickly, and then increase the satisfaction of users. Typically, sentiment analysis can be studied in three different directions: document-level, sentence-level, and aspect-level [1, 15].

There have been a lot of investigations related to the sentiment classification, especially for English and Chinese. Zhang and co-workers [38] apply Word2Vec [16] for word segmentation to compute semantic features from comment texts in the selected domain and Chinese language. For training a sentiment classification model, they use the SVMperf [9] and handle experiments on the data set of Chinese comments on clothing products. The experimental results show that the proposed method outperforms other techniques in terms of precision, recall, and F1-score. Jianqiang and colleagues use Glove [6] to extract semantic features from tweets and deep convolution neural networks to train the corresponding sentiment classification model on five different Twitter data sets. This approach achieves a better performance than others for twitter sentiment classification on both accuracy and F1-score. Also, convolutional neural networks (CNNs) and recurrent neural networks (RNNs) are commonly used to extract latent semantic features and make a significant contribution to many NLP tasks, especially for text classification, such as, e.g., deep character-level CNNs [39], shallow word-level CNNs [25], Deep Pyramid CNN [10] and LSTM [26]. Besides, Tang et al. [27] investigate the document-level sentiment classification by using gated recurrent neural networks to learn vector-based document representation and obtain semantics of sentences and the associated relations from one document. The experimental results show that the proposed approach hugely outperforms other state-of-the-art (SOTA) methods on four large-scale review datasets from IMDB and Yelp Dataset Challenge. Yang and co-workers [36] present a hierarchical attention network for document classification, which incorporates attention mechanism into the existing GRU [2] and then outperforms other methods on six large-scale text classification tasks by a substantial margin. Xu et al. [35] propose a Cached Long Short-Term Memory neural networks (CLSTM) to capture the overall semantic information in long texts and conduct experiments on three publicly available document-level sentiment analysis datasets. The experimental results show a better performance in comparison with the state-of-the-art models on these datasets. There are other approaches by employing the transfer learning technique by applying pre-trained models such as BERT [5] for the sentiment classification problem and obtain the state-of-the-art results [24,34], especially for the Arabic language [7].

Up to now, there have been several works related to the Vietnamese sentiment analysis. Phu and colleagues [20] present an approach by applying both traditional and deep learning techniques for sentiment classification and topic classification on a Vietnamese Students' Feedback Corpus, collected from a university in Vietnam. Phan and co-workers [23] use a Skip-gram based model to investigate the sentiment analysis on unstructured documents for Vietnamese text comments about locations. By using the Bag-of-Structure technique as well as traditional machine learning methods, Hung et al. [30] study the topic classification and sentiment analysis for a Vietnamese education survey system. Son and the team [28] study the sentiment analysis for the Vietnamese language on one dataset containing Facebook comments. They apply a lexicon-based approach for computing the semantic features and use Support Vector Machine (SVM) to train a machine learning model to identify the corresponding emotion in a given user message with good performance. Recently, Tuyen at el. [22] present a Vietnamese sentiment analysis approach by exploring different classifiers and lexicon features to improve the performance of the sentiment polarities task. Other works can be found at [14, 19, 21].

In this paper, we present an efficient framework for the Vietnamese sentiment classification by exploring the most recent techniques in natural language processing, including Bi-LSTM/GRU, attention models, Recurrent CNN, Residual CNN, Transformer methods and Transfer Learning. We conduct our experiments on two different datasets, one public dataset, namely AIVIVN, which contains users' reviews from the Vietnamese e-commerce websites used in the Vietnam Sentiment Analysis Challenge 2019. We collect another dataset from Vietnamese sites, which consists of 459,442 customers' reviews. The experimental results show that our proposed methods significantly outperform the winning solution of the competition AIVIVN 2019. Among the proposed techniques, the Recurrent CNN surpasses others in both two datasets. We aim at later publishing our methods and the new dataset for giving an additional contribution to the research community on the Vietnamese sentiment classification field.

2. Our proposed methods

In this section, we present our approaches for the Vietnamese sentiment classification task on two datasets selected. We first describe the main problem and propose an overall framework for the sentiment classification problem. Subsequently, we explain the data processing step for each review and introduce all predictive models using in this paper.

2.1. Problem Formulation

Sentiment analysis is the text analysis method of identifying the polarity within a text, whether a whole document, paragraph, sentence, or clause. Typically, the current sentiment analysis methods focus not only on polarity (positive, negative, neutral) but also on feelings and emotions (angry, happy, sad, disgust, and surprise), and even on intentions (interested or not interested).

In this paper, we aim at investigating a binary classification model for the Vietnamese sentiment classification problem, where for each review in Vietnamese, the model can determine whether it has the polarity as positive or negative. We use two datasets, AIVIVN 2019[2] and our collected dataset, for training and measuring the performance of the proposed models. For instance, are several examples of Vietnamese reviews are shown below:

- Positive: *sản phẩm tốt tiki giao hàng nhanh tôi rất hài lòng* (The product is nice, the delivery in Tiki is speedy, so I am pleased).
- Negative: *mới mua máy được 1 hôm dùng thử lần đầu thì lỗi luôn lỗi e2 lại phải gửi máy đi sửa chữa ở tỉnh thì hết sức bất tiện khá nản* (The new machine was broken right after one-day usage as well as having errors for the first trial. It is super inconvenient to send and repair this machine in another province. So discouraged!)

[2] https://www.aivivn.com/contests/1

Assume that w is a review from a given customer. We denote "0" as the label for positive sentiment and "1" as the label for negative polarity. The main problem is to construct a binary classifier F to predict the corresponding label for w. The overall framework for a Vietnamese sentiment classification system can be visualized in Figure 1.

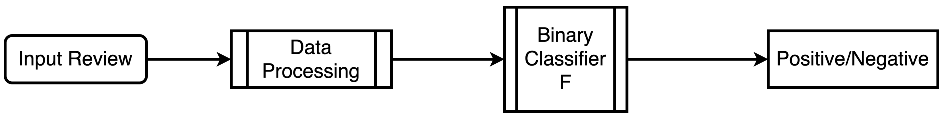

Figure 1. Our proposed framework for a Vietnamese sentiment classification system.

2.2. Data Processing

Data processing is one of the essential steps in natural language processing, especially for text data collected from social networks and e-commerce websites. In our data collected, we observe that there exist many informal texts and words that do not conform to the usual standard of the Vietnamese language. For this reason, data processing can help us to remove noisy data and transform human language contents into such a form that machine learning models can efficiently learn and perform well. In this study, from a given review, we apply various preprocessing steps for the corresponding text before training or testing an appropriate model.

First, we lowercase all characters and correct all elongated words (such as, e.g., ***đẹppppp quá***) to the correct format (***đẹp quá*** (so beautiful)). After that, we remove punctuation marks, special characters, icons such as ***! @ ? ( )***. Also, we filter all numbers out, which likely do not contribute to the sentiment of a sentence. Although all reviews collected from Vietnamese e-commerce websites, there still exist a small ratio of reviews written in other languages, including Korean, Chinese, and English. For this reason, during the data processing step, these such reviews are excluded as well. Finally, we observe that there exist many freestyle letters and acronyms in Vietnamese reviews, and these letters represent words heavily contributing to the sentiment of a sentence. As a result, it is worth replacing these letters with the correct ones. For instance, we substitute ***kp*** by ***không phải*** (which means "not" in Vietnamese) and ***ô kê*** by ***ok***.

2.3. Predictive Models

In this section, we describe our proposed methods for the Vietnamese review sentiment problem. Typically, there are two approaches to this work. First, we utilize a pre-trained word embedding method (e.g., Fasttext in our work) and train an appropriate deep neural network to learn the semantic features as well as the text representation. We explore five different architectures, including Bi-LSTM/GRU, Bi-LSTM/GRU + Attention, Recurrent CNN, Residual CNN, and Transformer. In the second approach, we choose a transfer-learning method by employing a pre-trained model, PhoBERT, for the Vietnamese language and fine-tuning for the current sentiment classification problem. To the best of our knowledge, all these techniques have never been explored in the Vietnamese sentiment analysis.

2.3.1. Bi-LSTM/GRU

Bi-Recurrent Neural Networks (Bi-RNNs) are ubiquitously used in many problems in natural language processing. Its particular structure allows us to capture both backward and forward information related to a sequence at every time step. Two variants of RNNs, Long Short-Term Memory (LSTM) [26] and Gated Recurrent Unit (GRU) [2], are proven to capture potential long-term dependencies in a sequence efficiently. In our work, we utilize both Bi-LSTM and Bi-GRU as baseline methods for the Vietnamese sentiment classification problem.

Assume that after the preprocessing step, we obtain n words $\{w_1, w_2, \ldots, w_n\}$ from a given review w. We denote $\{x_1, x_2, x_3, ..., x_n\}$ as the corresponding word embedding of n words w_i respectively and $\{h_1, h_2, h_3, ..., h_n\}$ represents hidden vectors computed by the network. Typically, one can easily find the corresponding formulae to compute all these hidden vectors h_i in both LSTM and GRU architectures at [2, 26]. Generally, Bi-LSTM/GRU can acquire information from both forward and backward directions of a sentence. At each time, the forward network computes the forward hidden vector fh_t based on the previous ones $\{fh_{t-1}, fh_{t-2}, \ldots\}$ and the backward network computes the backward hidden vector bh_t based on the previous backward hidden vectors $bh_{t-1}, bh_{t-2}, \ldots\}$. Then, the final hidden vector h_t of a Bi-LSTM/GRU at time t can be obtained by concatenating both the forward hidden vector and the backward hidden vector as: $h_t = [fh_t, bh_t]$. One can see more details at Figure 2(a).

Unlike previous RNN-based methods [20, 27] for text classification, which only using the last hidden vector to feed into a fully connected layer to generate the sentiment prediction. In this work, we propose a new approach by utilizing all hidden vectors $\{h_1, h_2, h_3, ..., h_n\}$ to apply the Max Pooling 1D layer and the Average Pooling 1D layer consecutively for obtaining two new hidden vectors h_{max} and h_{avg}. After that, we concatenate all hidden vectors $[h_{max}, h_{avg}, h_n]$ as the final hidden vector to feed it into the last linear layer to predict the corresponding sentiment for the input review w. Our architecture is depicted at Figure 2(b).

2.3.2. Bi-LSTM/GRU + Attention

The structure of our proposed Bi-LSTM/GRU + Attention architecture can be visualized as Figure 2(c). The only difference from the Bi-LSTM/GRU above is instead of concatenating all hidden vectors $[h_{max}, h_{avg}, h_n]$ as the input vector for the last linear layer, we use all hidden vectors $h_1, h_2, h_3, ..., h_n$ as the input data for the General Global Attention layer [13], to compute the attention hidden vector h_{att}. Subsequently, we combine all three vectors $[h_{max}, h_{avg}, h_{att}]$ into a vector and feed it into the last fully connected layer to predict the final sentiment.

2.3.3. Recurrent CNN

In another approach, we apply the recurrent CNN (RCNN) architecture [12] for the Vietnamese sentiment analysis, as shown in Figure 2(d). In this architecture, we use Bi-GRU to extract three types of hidden vectors, including the forward vectors fh_t, the backward vectors bh_t, and their concatenated vectors h_t (similar to two techniques above). However, at each time t, we define a new representation of the word w_t as the concatenation of the forward context, the word embedding, and the backward context $[fh_t, x_t, bh_t]$. After obtaining the new representation of the word w_t, we put it to go through two fully connected

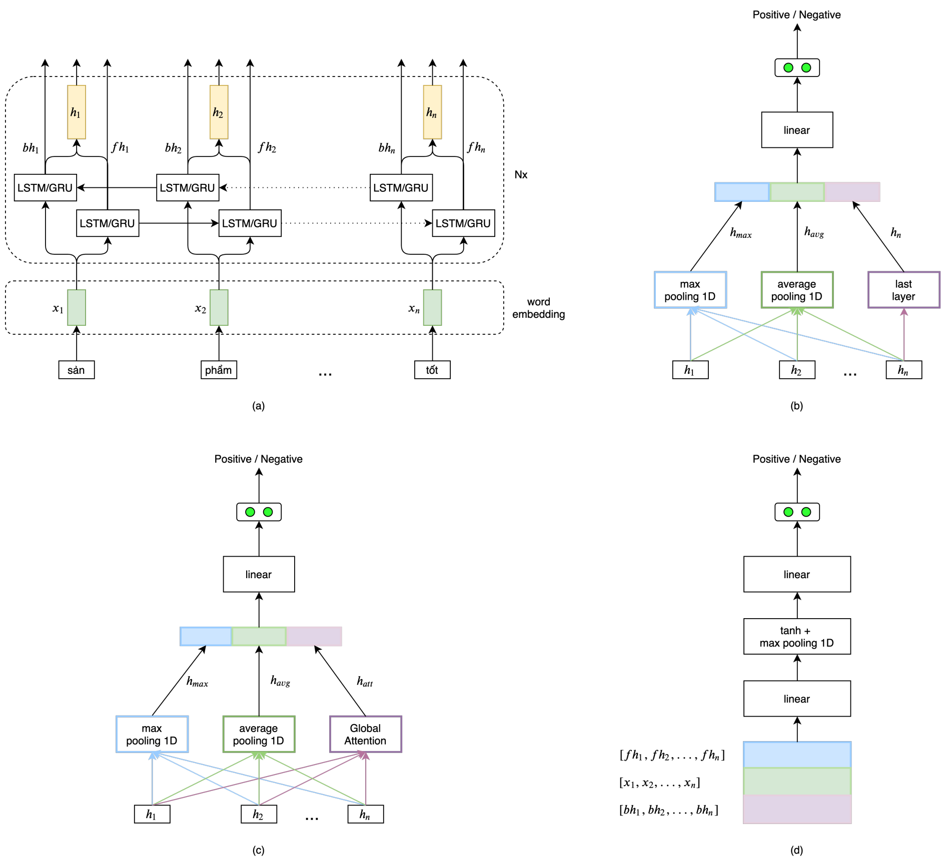

Figure 2. Our proposed architectures using Bi-LSTM/GRU, ATTENTION, Recurrent CNN for Vietnamese review sentiment. (a) The baseline backbone of Bi-LSTM/GRU with the outputs of hidden vectors: forward hidden fh_i, backward hidden bh_i, concate hidden h_i. Where Nx is number of layers. (b) The detail of our proposed Bi-LSTM/GRU architecture. (c) The detail of our proposed Bi-LSTM/GRU + Attention architecture. (d) The detail of our proposed Recurrent CNN architecture.

layers. Here, the first linear layer aims to reduce the dimension of the new representation embedding followed by a Max Pooling 1D, and the last linear layer aims to generate the predicted sentiment of the initial review w.

2.3.4. Residual CNN

Most of the conventional approaches for text classification are heavily related to the recurrent network, which can capture the dependencies of the input sequence. Recently, the only CNN-based architectures with Residual blocks are explored [10, 31, 33], and achieve promising results. Deep Pyramid CNN (DPCNN) [10] is one of the most successful methods among these and get state-of-the-art benchmarks. In this paper, we utilize DPCNN as a Residual CNN architecture for constructing the sentiment classification model. It is important to note that such a technique has never been explored in the Vietnamese sentiment classification problem. The detail of our Residual CNN architecture can be introduced in Figure 3.

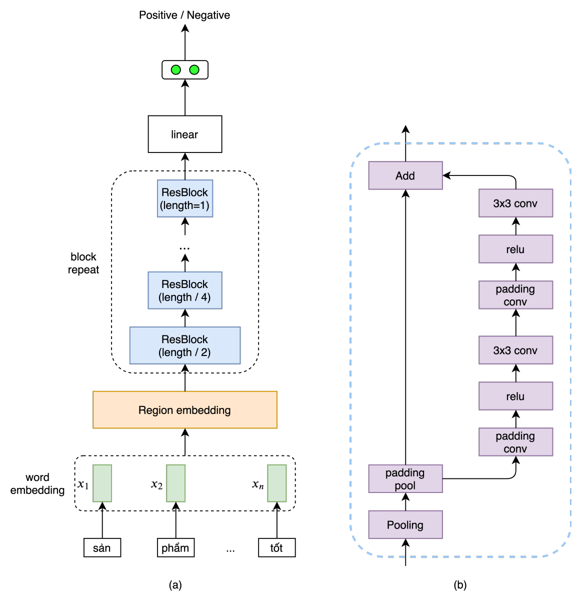

Figure 3. Our proposed Residual CNN architecture for Vietnamese review sentiment. (a) A typical architectural layout. (b) The detail of residual block.

2.3.5. Transformer

Recently, Transformer [29] has emerged as one of the most advanced architectures contributing to the latest SOTA results in natural language processing. The Transformer is also integrated into the different SOTA models, such as BERT [5] and GPT-2 [24]. Typically, a Transformer architecture has two components, Encoder, and Decoder. The details of these components can be found at [29]. By taking advantage of the architecture of the Transformer, we use the Encoder component to create a new architecture for the Vietnamese sentiment classification problem (Figure 4). Remarkably, in our architecture, we do not use the Decoder component, and the encoder component has six layers of Multi-Head Attention and Feed Forward. Finally, the output vector from the last Encoder is fed into the last linear layer to create the predicted output.

2.3.6. Transfer Learning

Transfer learning is a common concept in machine learning by taking advantage of the knowledge gained from one problem and apply it to a similar problem. Practically, it involves that taking a pre-trained model which has been trained on a considerable amount of data and then retraining or fine-tuning that model on a domain-specific data. This technique is quite useful as we do not need to have a massive amount of data and computational power to train a vast architecture with million or even billion parameters from scratch [5]. For the Vietnamese sentiment classification problem, we apply a transfer learning approach by using PhoBERT [17] to extract semantic features from a given review and then train an appropriate model. This approach can also be considered as a baseline model for this study.

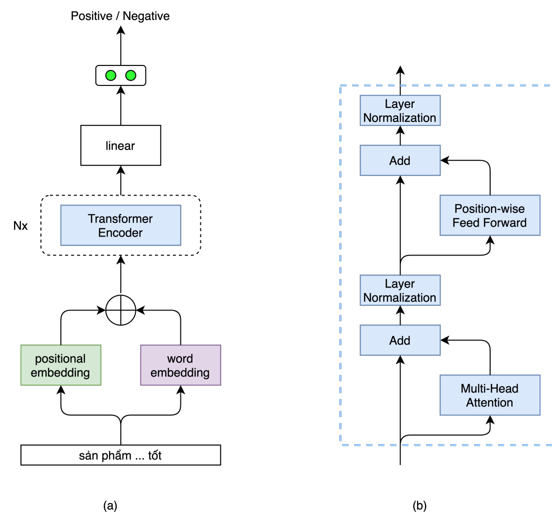

Figure 4. Our architecture utilizing Transformer Encoder for Vietnamese review sentiment. (a) An overall architecture. (b) The typical architecture of Transformer Encoder.

3. Experiments

In this section, we describe our datasets, the implementation of the proposed techniques, and experimental results in detail. We conduct all experiments on a workstation with Intel Core i9-7900X CPU, 128GB RAM, and two GPUs RTX-2080Ti.

3.1. Datasets

To demonstrate the effectiveness of the proposed methods for Vietnamese sentiment analysis, we first use a public dataset AIVIVN including user reviews of Vietnamese e-commerce pages, which was used for the Vietnam Sentiment Analysis Challenge 2019[3]. The first dataset has 16,073 samples in the training set and 10,981 in the testing dataset. As all labels in the testing dataset are not available (which are kept private from the competition organizers), we decide to carefully label all these testing samples by ourselves and double-check multiple times by our team members to measure the proposed methods. Later, we will publish this dataset labeled to the research community for further investigation.

Besides, we collect another dataset by crawling more customer reviews in different Vietnamese e-commerce pages to create a more challenging dataset for our proposed methods. Our dataset collected contains 358,743 positive reviews and 100,699 negative reviews. It is important to note that after crawling raw reviews, it took a lot of time for us to manually label all these reviews into two categories: positive and negative. This dataset can be considered as another contribution to the research community in the Vietnamese sentiment classification topic. In both two datasets, we divide it into training and testing datasets with a ratio of 7 : 3. Again, all reviews are preprocessed carefully by our data processing techniques before putting in training and testing models. The detail of the two datasets is given in Table 1.

[3] https://www.aivivn.com/contests/1

AIVIVN	Positive	Negative	Our Dataset	Positive	Negative
Train	8690	7383	Train	251120	70489
Test	5767	5214	Test	107623	30210

Table 1. AIVIVN 2019 Sentiment Challenge Dataset and Our Sentiment Dataset.

3.2. Implementation

In experiments, we implement all proposed methods by using Pytorch. Except for the pretrained model PhoBERT for the Vietnamese language, all other model architectures are designed and implemented from scratch. To use PhoBert for our problem, we have to apply the Vietnamese word segmenter RDRsegmenter [18] to process raw data and generate segmented words before feeding them into PhoBert. The Fasttext embedding of Vietnamese version [4], whose embedding dimension is 300, is used as the embedding layer for the non-Bert methods. All models are trained with Adam optimizer with a learning rate of 0.001, on two GPUs by 30 epochs. Dropout is not used in our architectures and set by 0 as default. The batch size is placed by 32 for the Transformer, and by 512 for other models. All model architectures and hyper-parameters respectively are kept the same while training on the two datasets. We measure the performance of different models by using AUC, accuracy, and the F1 score.

3.3. Experimental results

Tables 2 and 3 illustrate the performance of all proposed techniques on two datasets in terms of AUC, accuracy, and F1-score. First, one can interesting see that for the AIVIVN 2019 dataset, all our proposed methods remarkably outperform than the winning solution of the contest "AIVIVN 2019 Sentiment Champion" (which has the best F1-score as 0.90012). It is worth remarking that the winning solution of the competition is the weighted ensemble of TextCNN [11], VDCNN [3], HARNN [36], and Self-Attention RNN [29]. One can see that our best-proposed technique on this dataset is Recurrent CNN, with the F1-score as 0.93416. It can demonstrate the absolute efficiency of our proposed framework for the Vietnamese sentiment classification task on the dataset AIVIVN 2019. Also, among recurrent networks, Bi-GRU provides better and more efficient results than Bi-LSTM. Bi-LSTM + Attention leads to a bit of improvement in comparison to Bi-LSTM/GRU. Finally, using PhoBERT can give a better performance than the winning solution in the competition. Still, it achieves the worst performance compared to the remaining of our proposed methods in terms of F1-score.

On our dataset collected, Recurrent CNN again outperforms all other techniques in terms of both AUC and F1-scores, while Bi-GRU + Attention achieves the highest accuracy. Different from the first dataset, using the General Global Attention layer in Bi-LSTM and Bi-GRU architectures can help to improve the performance. More detailed, both Bi-LSTM + Attention and Bi-GRU + Attention have a better F1-score than Bi-LSTM and Bi-GRU, respectively. Interestingly, both the Transformer and PhoBERT have the lowest performance among all proposed methods in all two datasets.

[4]https://fasttext.cc

Methods	AUC	Accuracy	F1
AIVIVN 2019 Sentiment Champion	-	-	0.90012
Bi-LSTM	0.97722	0.9198	0.91599
Bi-LSTM + Attention	0.97749	0.92182	0.91962
Bi-GRU	0.97914	0.92463	0.92093
Bi-GRU + Attention	0.97886	0.92071	0.91535
Recurrent CNN	**0.98475**	**0.93651**	**0.93416**
Residual CNN	0.98253	0.93037	0.92621
Transformer	0.97729	0.92091	0.9147
PhoBERT	0.98172	0.9154	0.91264

Table 2. The results of the winning solution in the competition "AIVIVN 2019 Sentiment Champion" and our eight proposed methods on the dataset AIVIVN 2019.

Methods	AUC	Accuracy	F1
Bi-LSTM	0.95942	0.9154	0.80129
Bi-LSTM + Attention	0.95944	0.91551	0.80361
Bi-GRU	0.96035	0.91556	0.80634
Bi-GRU + Attention	0.95984	**0.91618**	0.80838
Recurrent CNN	**0.96077**	0.91608	**0.80923**
Residual CNN	0.95712	0.91207	0.78404
Transformer	0.92678	0.89718	0.71807
PhoBERT	0.95004	0.90477	0.75189

Table 3. The result of our proposed methods for our Vietnamese reviews dataset.

4. Conclusion

We have investigated the Vietnamese sentiment classification problem by presenting eight different methods, including Bi-LSTM, Bi-LSTM + Attention, Bi-GRU, Bi-GRU + Attention, Recurrent CNN, Residual CNN, Transformer, and PhoBERT. We have conducted all experiments on two datasets, AIVIVN 2019 and our dataset self-collected from multiple Vietnamese e-commerce websites, and compare the performance of proposed techniques by using AUC, accuracy, and F1-score. The experimental results show that our proposed methods all outperform the winning solution of the competition"AIVIVN 2019 Sentiment Champion" with a significant margin. Among those proposed techniques, Recurrent CNN has the best performance in comparison with other algorithms; meanwhile, both Transformer and PhoBERT have the lowest F1-scores on both two datasets. We aim at publishing all implementing codes and datasets later to give an additional contribution to the research community.

In future work, we plan to extend our research for multiple polarities and use aspect extraction to enhance the performance of the Vietnamese sentiment classification problem.

References

[1] Priya Chakriswaran, Durai Raj Vincent, Kathiravan Srinivasan, Vishal Sharma, Chuan-Yu Chang, and Daniel Gutiérrez Reina. Emotion ai-driven sentiment analysis: A survey, future research directions, and open issues. *Applied Sciences*, 9(24), 2019.

[2] Kyunghyun Cho, Bart Van Merrienboer, Caglar Gul-cehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation. 2014.

[3] Alexis Conneau, Holger Schwenk, Loïc Barrault, and Yann Lecun. Very deep convolutional networks for text classification. 2016.

[4] Felipe G. Contratres, Solange N. Alves-Souza, Lucia Vilela Leite Filgueiras, and Luiz S. DeSouza. Sentiment analysis of social network data for cold-start relief in recommender systems. In Álvaro Rocha, Hojjat Adeli, Luís Paulo Reis, and Sandra Costanzo, editors, *Trends and Advances in Information Systems and Technologies*, pages 122–132, Cham, 2018. Springer International Publishing.

[5] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. 2019.

[6] effrey Pennington, Richard Socher, and Christopher Manning. Glove: Global vectors for word representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)s*, page 1532–1543, 10 2014.

[7] Obeida ElJundi, Wissam Antoun, Nour El Droubi, Hazem Hajj, Wassim El-Hajj, and Khaled Shaban. hulmona: The universal language model in arabic. *Proceedings of the Fourth Arabic Natural Language Processing Workshop*, page 68–77, 8 2019.

[8] Feifan Fan, Yansong Feng, and Dongyan Zhao. Multi-grained attention network for aspect-level sentiment classification. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, page 3433–3442, 11 2018.

[9] Thorsten Joachims. A support vector method for multivariate performance measures. In *Proceedings of the 22nd International Conference on Machine Learning*, ICML '05, page 377–384, New York, NY, USA, 2005. Association for Computing Machinery.

[10] Rie Johnson and Tong Zhang. Deep pyramid convolutional neural networks for text categorization. *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, 1:562–570, 07 2017.

[11] Yoon Kim. Convolutional neural networks for sentence classification. 2014.

[12] Siwei Lai and Liheng Xu andKang Liu andJun Zhao. Recurrent convolutional neural networks for text classification. *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, page 2267–2273, 01 2015.

[13] Minh-Thang Luong, Hieu Pham, and Christopher D. Manning. Effective approaches to attention-based neural machine translation. 2015.

[14] Long Mai and Bac Le. *Aspect-Based Sentiment Analysis of Vietnamese Texts with Deep Learning*, pages 149–158. 01 2018.

[15] Walaa Medhat, Ahmed Hassan, and Hoda Korashy. Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, 5(4):1093 – 1113, 2014.

[16] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. Distributed representations of words and phrases and their compositionality. In *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2*, NIPS'13, page 3111–3119, Red Hook, NY, USA, 2013. Curran Associates Inc.

[17] Dat Quoc Nguyen and Anh Tuan Nguyen. Phobert: Pre-trained language models for vietnamese. 2020.

[18] Dat Quoc Nguyen, Dai Quoc Nguyen, Thanh Vu, Mark Dras, and Mark Johnson. A Fast and Accurate Vietnamese Word Segmenter. In *Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC 2018)*, pages 2582–2587, 2018.

[19] Hoang-Quan Nguyen and Quang-Uy Nguyen. An ensemble of shallow and deep learning algorithms for vietnamese sentiment analysis. pages 165–170, 11 2018.

[20] Phu X. V. Nguyen, Tham T. T. Hong, Kiet Van Nguyen, and Ngan Luu-Thuy Nguyen. Deep learning versus traditional classifiers on vietnamese students' feedback corpus. *2018 5th NAFOSTED Conference on Information and Computer Science (NICS)*, 2018.

[21] Thuy Nguyen-Thanh and Giang Tran. Vietnamese sentiment analysis for hotel review based on overfitting training and ensemble learning. 12 2019.

[22] Bich-Tuyen Nguyen-Thi and Huu-Thanh Duong. *A Vietnamese Sentiment Analysis System Based on Multiple Classifiers with Enhancing Lexicon Features*, pages 240–249. 08 2019.

[23] Dang-Hung Phan and Tuan-Dung Cao. Applying skip-gram word estimation and svm-based classification for opinion mining vietnamese food places text reviews. In *Proceedings of the Fifth Symposium on Information and Communication Technology*, SoICT '14, page 232–239, New York, NY, USA, 2014.

Association for Computing Machinery.

[24] Alec Radford, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. Language models are unsupervised multitask learners. 2019.

[25] Tong Zhang Rie Johnson. Convolutional neural networks for text categorization: Shallow word-level vs. deep character-level. 2016.

[26] Hochreiter S. and Schmidhuber J. Long short-term memory. *Neural Computing*, 9:1735–1780, 11 1997.

[27] Duyu Tang, Bing Qin, and Ting Liu. Document modeling with gated recurrent neural network for sentiment classification. *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, page 1422–1432, 9 2015.

[28] Son Trinh, Luu Nguyen, Minh Vo, and Phuc Do. *Lexicon-Based Sentiment Analysis of Facebook Comments in Vietnamese Language*, volume 642, pages 263–276. 02 2016.

[29] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. 2017.

[30] Hung Vo, Hai Lam, Duc Dung Nguyen, and Nguyen Tuong. Topic classification and sentiment analysis for vietnamese education survey system. *Asian Journal of Computer Science and Information Technology*, 6:27–34, 05 2016.

[31] Baoxin Wang. Disconnected recurrent neural networks for text categorization. *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, page 2311–2320, 7 2018.

[32] Gang Wang, Jianshan Sun, Jian Ma, Kaiquan Xu, and Jibao Gu. Sentiment classification: The contribution of ensemble learning. *Decision Support Systems*, 57:77 – 93, 2014.

[33] Yiren Wang and Fei Tian. Recurrent residual learning for sequence classification. *In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, page 938–943, 2016.

[34] Qizhe Xie, Zihang Dai, Eduard Hovy, Minh-Thang Luong, and Quoc V. Le. Unsupervised data augmentation for consistency training. 2019.

[35] Jiacheng Xu, Danlu Chen, Xipeng Qiu, and Xuanjing Huang. Cached long short-term memory neural networks for document-level sentiment classifi- cation. *In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, page 1660–1669, 2016.

[36] Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alex Smola, and Eduard Hovy. Hierarchical attention networks for document classification. *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, page 1480–1489, 06 2016.

[37] Biqing Zeng, Heng Yang, Ruyang Xu, Wu Zhou, and Xuli Han. Lcf: A local context focus mechanism for aspect-based sentiment classification. *Applied Sciences*, 9(16), August 2019.

[38] Dongwen Zhang, Hua Xu, Zengcai Su, and Yunfeng Xu. Chinese comments sentiment classification based on word2vec and svmperf. *Expert Systems with Applications*, 42(4):1857 – 1863, 2015.

[39] Xiang Zhang, Junbo Zhao, and Yann LeCun. Character-level convolutional networks for text classification. In *Advances in neural information processing systems*, pages 649–657, 2015.