Proposing System to Recognize Emotions in Public Network Using Phobert Deep Learning Model

Phat Nguyen Huu[1], Tuan Nguyen Anh[1], Long Hoang Phi[1], Dinh Dang Dang[1],
Chau Nguyen Le Bao[2], and Quang Tran Minh[3,4]
[1] School of Electrical and Electronic Engineering, Hanoi University of Science and Technology, Hanoi, Vietnam
[2] Hanoi-Amsterdam Highschool for the Gifted, Hanoi, Vietnam
[3] Department of Information Systems, Faculty of Computer Science and Engineering,
Ho Chi Minh City University of Technology, Hochiminh, Vietnam
[4]Vietnam National University HoChiMinh City, Hochiminh, Vietnam
Email: phat.nguyenhuu@hust.edu.vn; tuan.na172900@sis.hust.edu.vn; long.hp182651@sis.hust.edu.vn;
dinh.dd200152@sis.hust.edu.vn; chau.nguyenlebao2007@gmail.com; quangtran@hcmut.edu.vn

***Abstract*—People can receive information faster especially in the 4.0 revolution with the continuous development of revolutions. The information can affect our emotional, psychological and spiritual well-being, especially in the recent high school graduation exam across the country. Therefore, we propose to build a user emotion analysis system in a public network with the PhoBert training model in the paper. Besides, we built our dataset aggregated from social networks, articles, blogs, etc. We next use the PhoBert model to solve the processing data problems. The simulation results have shown an accuracy of 86.5% on the training and 81.32% on the validation dataset with a training time of 3 hours (about 180 minutes). The results also show that we can build a warning system to avoid health and psychological effects with great emotion.**

***Index Terms*—NLP, PhoBert, emotional classification, social networks, deep learning.**

I. Introduction

In recent years, Internet has been a popular means of communication and entertainment that is widely used and preferred by many users, especially on major social networks such as Facebook, Instagram, Twitter, and major newspapers (Vnexpress, Thanhnien, etc.). Users in Vietnam in 2019 are spending an average of 2.32 hours per day according to a report [1]. However, this number has increased to nearly 7 hours per day in 2020 [2]–[6], and over 60% of them are using the Internet.

The Internet brings many positive benefits to the lives of people such as providing services of storing, searching, sharing, and exchanging information with each other, buying goods, and posting news and images. These are services where people can express their personal views, share knowledge and life experiences, or reflect on the negative sides and illegal acts of organizations. However, there are still many negative aspects affecting users, especially from 11 to 13-year-old women and from 14 to 15-year-old men [2]–[6] who are two vulnerable groups. Persons and information distribute violent articles and images that harm psychological and mental health. Although this violence occurs on Internet, it will have real consequences even leading to suicide. Therefore, we propose to build a user emotion analysis system in a public network to be able to warn about content that may affect psychological and mental health, especially among users aged between 11 and 15 years old in this article.

Based on the user emotion analysis system [7], we propose the system with two main points as follows.

- Firstly, we create our dataset that is aggregated from social networks, articles, and blogs and then label emotions.
- Secondly, we use the PhoBert model for training.

The rest of the paper includes four parts and is organized as follows. Section II will discuss the related work. In Section III, we present the proposal system. Section IV will evaluate the proposed system and analyze the results. In the final section, we give conclusions and future research directions.

II. Related work

Recently, there have been many studies related to user emotions on Internet, especially on social networks. In Vietnam [7]–[9], the authors have developed a dataset related to six emotions (interest, anger, disgust, sad, fear, and surprise). The authors used a lot of models for training. They have separated the dataset into two cases. In the first case, the dataset does not have the emotional label and the highest accuracy that the authors have is 59.74% in the CNN+word2vec model. In the second case, the authors add the emotion label, and the accuracy has been improved up to 66.48%. In [8], the authors built their dataset, the Vietnamese open-domain complaint detection dataset (UIT-ViOCD). This dataset includes 5485 user reviews of products on e-commerce sites. They then used the models in which the PhoBert model achieved the highest results with an accuracy of up to 92.16%. However, only two main labels have complaints and no complaints with the dataset. In [9], the authors also used a dataset with two labels positive and negative. The final result is that the model achieved the highest accuracy of 84.54%.

In [10], the author uses machine learning models to recognize six basic emotions including anger, disgust, fear, happiness, sad, and unexpected. The authors have shown that the sequential minimal optimization (SMO) algorithm is optimized for the support vector machines (SVM) method

978-1-6654-5422-3/22/$31.00 ©2022 IEEE

to achieve better performance than other methods. The best result is 81.16% and 57.81% for the training dataset and the testing dataset, respectively. In [11], the authors review sentiment analysis and emotion detection from text across different articles since they could be compared objectively. In this article, the authors have many documents with good results, including comparative Analyses of BERT, RoBERTa, DistilBERT, and XLNet for text-based emotion recognition with several labels up to seven. The total number of data is 7666 sentences. The best results that they achieved with the RoBerta, XLNet, BERT, and DistilBERT models were 74.31%, 72.99%, 70.09%, and 66.93%, respectively. In [12], the authors use a dataset of labels of love, joy, surprise, anger, sad, fear, and gratitude with a total of 1,387,787 sentences. The highest results are 63.2%.

As analyzed above, it can be seen that there are many datasets created that have been applied to the other problems. The above-surveyed algorithms and data only have accuracy from 57% to 85%. Besides, there are research articles that have an accuracy of up to 92%. However, the dataset has only two types, namely positive and negative labels. Therefore, we propose to build a dataset with many types and use the PhoBert model to solve the problems of analyzing user emotions on the Internet.

III. Proposal system

A. Overview

The overview of the proposed system is shown in Fig. 1.

The system consists of two main steps, namely training and prediction.

- In the first training part, we will create a suitable labeling dataset for each data and preprocess that data to put it into the training model.
- In the prediction part, the results of the training model will be applied to determine the emotions of users on the Internet.

Our main contribution to this paper is to create a dataset of user posts as shown in Fig. 1.

B. Dataset

We have collected public posts on social networks, articles, blogs, and forums, especially on major social networks such as Facebook and Instagram.

- Labeling principle

  The labeling principle is based on the analysis of basic human emotions [13]–[16]. We build a guide with descriptions for each emotion as follows.

  •Disgust: It often occurs in response to situations that make you uncomfortable or unwanted. It can make us feel disliked, disapproved, offended, and terrified.

  •Pleasure: It has a feeling of contentment.

  •Sad: It defines as a transient emotional state characterized by feelings of frustration, grief, despair, disinterest, and depressed mood.

  •Fear: A powerful emotion can play an important role in survival. When we face several kinds of danger or threat it can be harmful whether the danger is physical or psychological.

  •Anger: A particularly strong emotion characterized by feelings of hostility, incitement to frustration, and resistance to others. Depending on the intensity, the level changes from slightly uncomfortable to angry and indignant.

  •Other is different emotions or no expression at all.
- **Labeling data**

  We perform to label the dataset with the rules for each emotion type. Table I illustrates several the labeled data.
- **Amount of dataset**

  Tab. II shows the number of data for each label and the percentage of the total dataset.

C. Training

The training process includes the following steps:

- Pre-processing:

  Data preprocessing is a very important step in solving any problem in the field of natural language processing (NLP). Most of the datasets used in problems related to machine learning in general and language processing in particular need to be processed, cleaned and transformed before training them. This process consists of steps, namely word separation, and stopword.

  Word separation is a processing process to determine the boundaries of words in a sentence. It is the process of identifying single words or compound words in a sentence. It is necessary to determine which word is in the sentence for language processing. This problem seems simple to humans. However, this is very difficult to solve for computers. In the paper, we use the tool Vitokenizer () [17] to separate words as shown in Fig. 2.

  Stopwords are words that do not add meaning to a sentence. They can be omitted without losing the meaning of the sentence. In several search engines, they are short-function words, such as, ‘ah’, ‘less’, ‘like’, etc. In this case, they can cause problems when searching for terms that include them. Currently, there are many ways to help us separate these words. In the paper, we use the Vietnamese-stopword dictionary [18] to remove unnecessary words.
- Training:

  Currently, there are many training models such as recurrent neural network (RNN) and long short-term memory (LSTM) for language processing and emotion classification. However, these models have not brought high efficiency to the problem. Therefore, we choose the model PhoBert [19].

  PhoBert model has several outstanding features as follows.

  •Firstly, it is a single-language pre-trained for the Vietnamese language based on RoBERTa of Facebook and its architecture was introduced in mid-2019. RoBERTa is more advanced than the BERT model

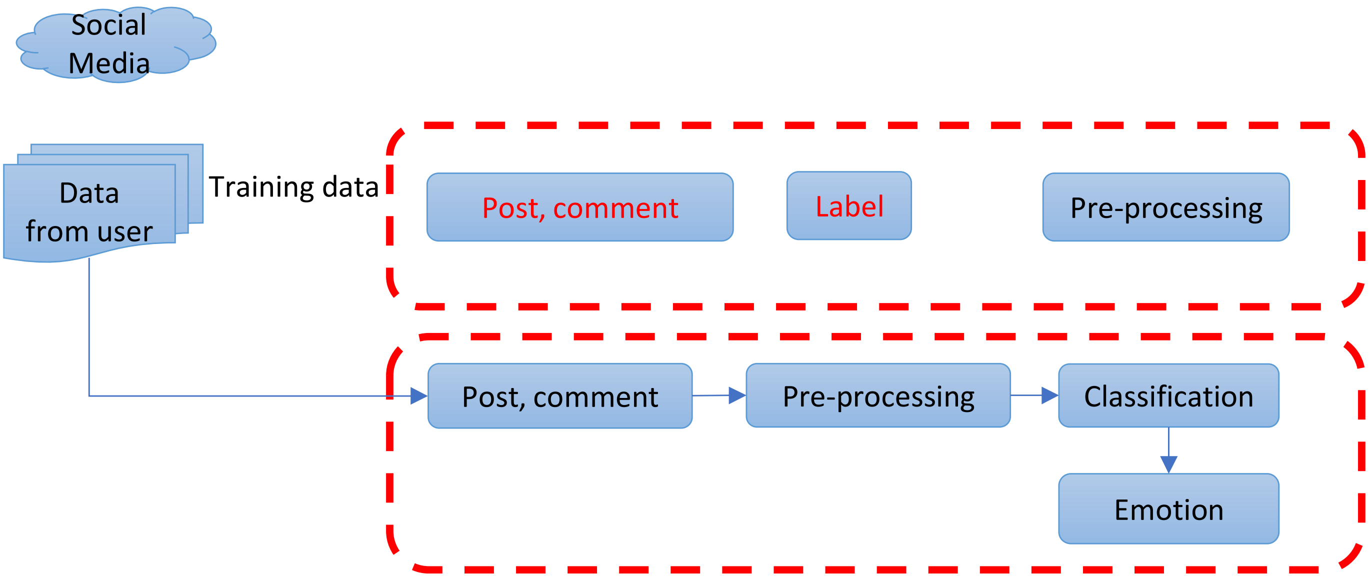


Fig. 1. The proposed system model.

TABLE I
ILLUSTRATING DATA LABELING PATTERN.

No.	Dataset	Label (English)	Label (Vietnamese)
1	Meaning of existence of lovers? What does love mean?	Disgust	Chan ghet
2	I promise from now on I will work hard to lose more weight	Pleasure	Thich thu
3	Haven't confessed yet :))	Sad	Buon ba
4	Letting the children come into contact with the swarm is too much :(	Fear	So hai
5	Look at this girl, Be careful	Anger	Gian du
6	Hello everyone living in the Ice Age!	Other	Khac

TABLE II
ANALYZING THE DATASET.

No.	Label	Label (Vietnamese)	Quantity	Rate (%)
1	Disgust	Chan ghet	1162	16.05
2	Pleasure	Thich thu	1748	24.14
3	Sad	Buon ba	1647	22.74
4	Fear	So hai	376	5.19
5	Anger	Gian du	463	6.39
6	Other	Khac	1846	25.49

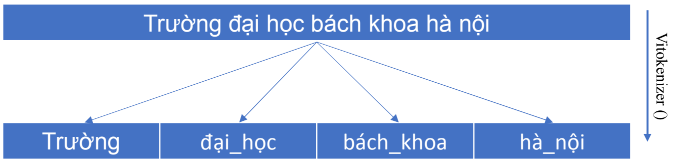


Fig. 2. Result of Vietnamese word separation.

[20]–[23].

•Secondly, there are two sessions, namely PhoBERT_base with 12 transformers block and PhoBERT_large with 24 transformers block.

•PhoBERT was trained on about 20GB of data including 1GB of Vietnamese and the remaining 19GB taken from Vietnamese news. This is a pretty good amount of data to train a model with architecture like BERT.
•The BERT model [20]–[23] uses a transformer as an attention model that learns the correlation between words in a text. The transformer consists of two main parts,

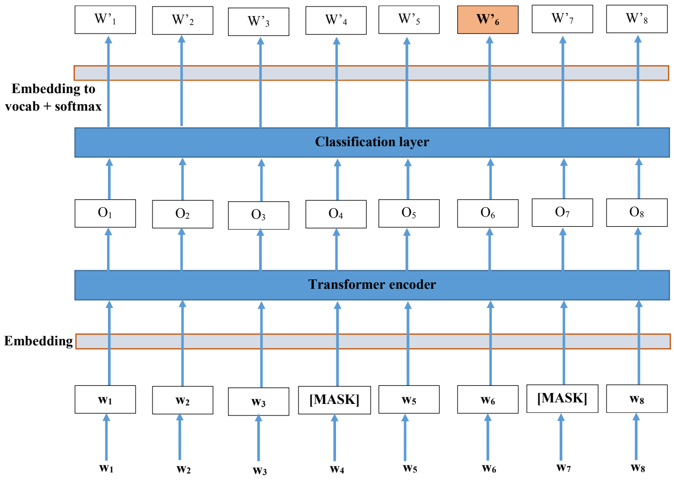


Fig. 3. Masked LM model BERT.

namely the encoder and decoder. The encoder reads input data and the decoder makes predictions. BERT only uses the encoder of the transformer model.

Unlike directional models (models that only read data in a single direction left ↦ right, right ↦ left), the encoder reads all data which makes BERT the ability to train data in both directions since the model can learn the context of the word better by using the words around it (right and left).

15% of the words in the string is replaced by the token [MASK] before entering BERT. The model will then predict the word to be replaced by [MASK] with the context that is not replaced. The process includes the following processing steps as shown in Fig. 3:

- Step 1: Adding a classification layer with input as the output of the encoder.
- Step 2: Multiplying the output vectors with the embedding matrix to return them to the vocabulary dimensional.
- Step 3: Calculating the probability of each word in the vocabulary using the softmax function.

In the next sentence prediction (NSP) step, the model uses a pair of sentences as input and predicts whether the second is the text of the first one or not. During the training process, 50% of the input data is a pair of sentences where the 2^{nd} sentence is a follow-up to the 1^{st}. The remaining data is the 2^{nd} sentence randomly selected from the dataset. Several guidelines are given while handling data as follows:

- Inserting the token $[CLS]$ before the first sentence and $[SEP]$ at the end of each sentence.
- The tokens in each sentence are marked as A or B.
- Inserting an embedding vector representing the position of the token in the sentence as shown in Fig. 4.

The processing steps in NSP include:

- The entire input sentence is fed into the transformer.
- Converting the output vector of $[CLS]$ to 2×1 size by a classification layer.
- Calculating the next sequence probability using softmax.

The algorithm flow chart is shown in Fig. 5. In Fig. 5, we train input data that have been labeled with corresponding emotions. We then move on to the preprocessing step. We then will be stripped of special characters, separated from the selected method, and refined acronyms. We next unset the dataset to 20% and 80% for the testing and training datasets, respectively. In the next step, we use the PhoBert model for training. We will test the model with the testing dataset and finish the training process. More details of the results will be presented in the next section.

IV. Result

A. Evaluation criteria

To evaluate the effectiveness of the model, we use

$$Accuracy = \frac{1}{n} \sum_{i=1}^{n-1} (\hat{y}_i = y_i) \, , \qquad (1)$$

where n is the number of classes, $\hat{y}_i$ is the value predicted by the model, and y_i is the corresponding actual result value.

We use the GoogleColab Pro device with the configuration of Intel Xeon Processor 2 cores, 2.20 GHz, and 13 GB of RAM while evaluating the accuracy.

B. Result

The results are shown in Figs. 6 and 7.

In Fig. 6 (a), it can be seen that the accuracy of training and testing data reached the highest value of 85.8% and 80.9%, respectively. Our accuracy results have been improved by 5% to 15% compared to the results in [7]–[9] for the Vietnamese and [10]–[12] for English data, respectively.

In Fig. 6 (b), we can see that the accuracy of the LSTM model has been overfitted since it reaches the maximum of 51%. The results show that the accuracy of model is 5% lower in [7]–[9] for Vietnamese and 20% [10]–[12] for English, respectively. Therefore, we can see that the PhoBert model is bringing better results compared to the related studies.

Figure 7 shows the accuracy of PhoBert and LSTM models after nearly 180 and 30 minutes, respectively. Although this time of model is much shorter than that of the PhoBert, its accuracy is lower. Therefore, we choose the PhoBert model to perform the simulation on the website platform.

V. Conclusion

In this article, we create a dataset of user posts on the Internet and use the Phobert model to train that achieve positive results. However, the dataset is still quite small and the amount of data between unbalanced labels (fear is 5.19%). Therefore, the model results are not optimized. In the next direction, we will add the amount of data to balance among labels and use newer models for training to improve the accuracy.

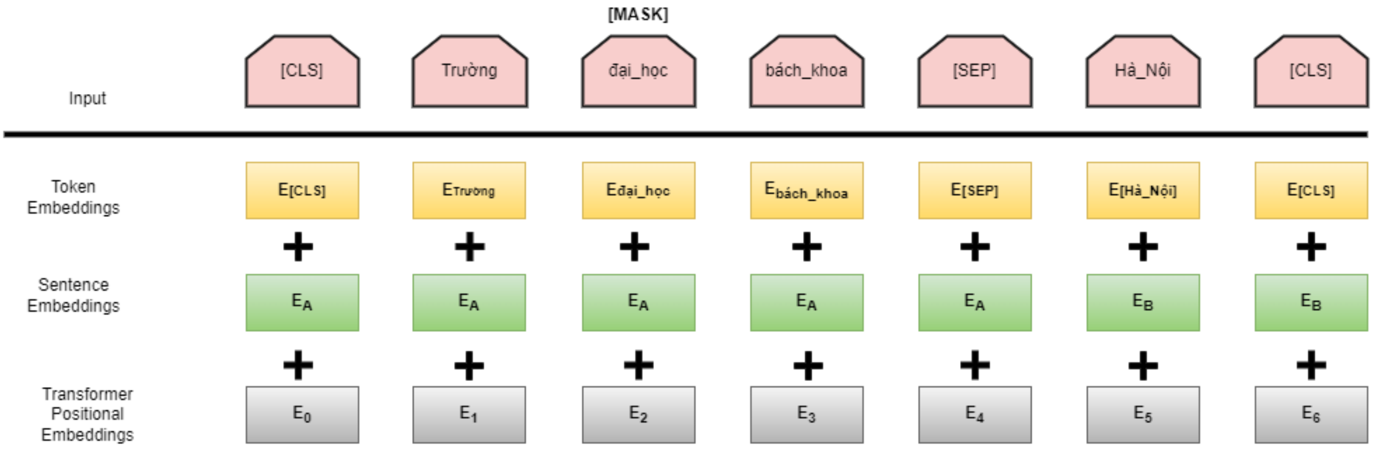


Fig. 4. Predicting the next sentence.

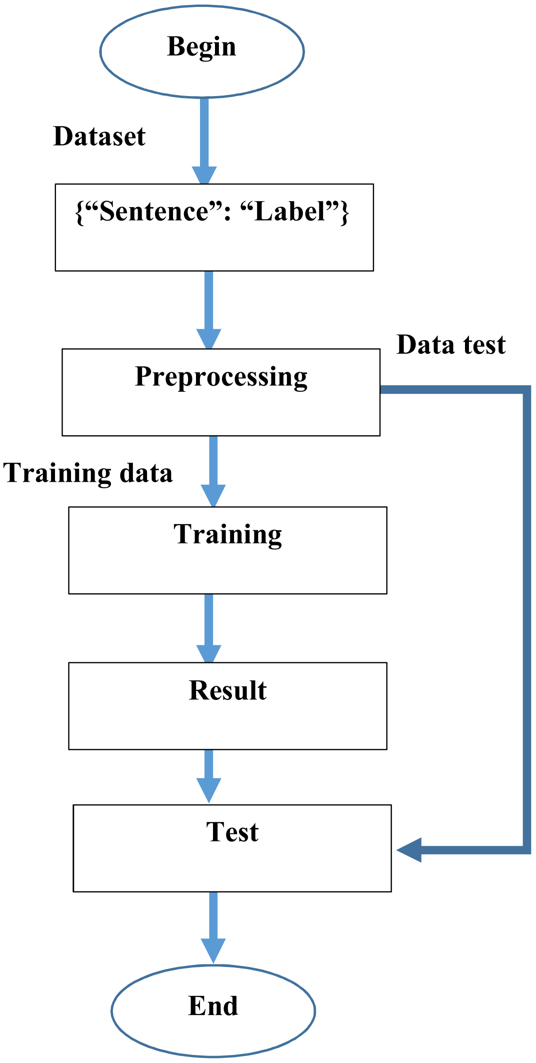
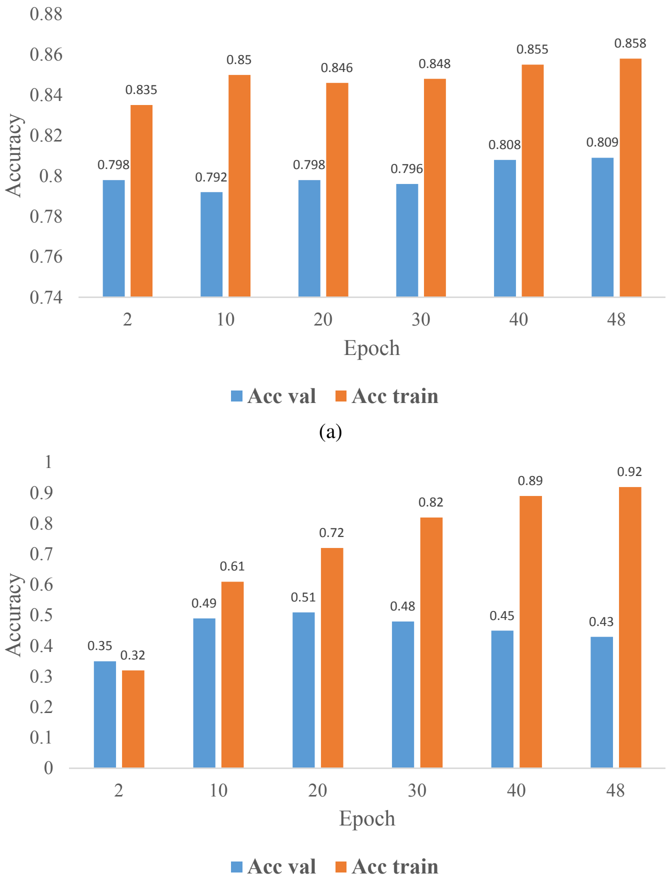


Fig. 5. Flowchart of the training algorithm.



(a)

(b)

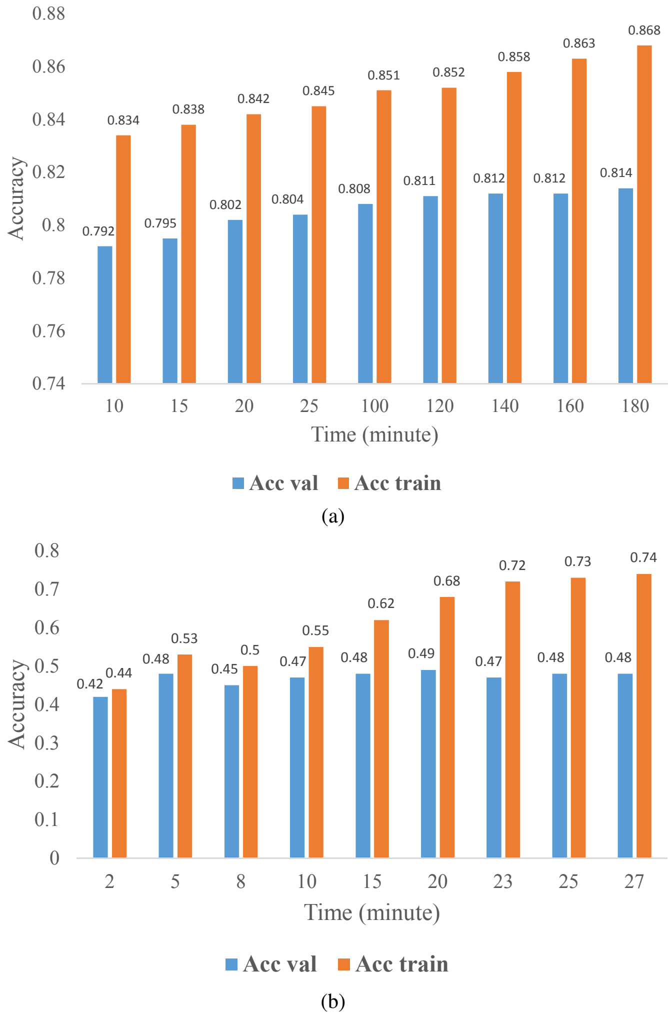
Fig. 6. Evaluating the accuracy of (a) proposal (using PhoBert) and (b) LSTM.

Acknowledgment

We would like to thank Hanoi University of Science and Technology (HUST) and Ho Chi Minh City University of Technology (HCMUT), VNU-HCM for the support of time and facilities for this study.

References

[1] D. T. Tuan, D. Van Thin, V.-H. Pham, and N. L.-T. Nguyen, "Vietnamese facebook posts classification using fine-tuning bert," in *2020 7th NAFOSTED Conference on Information and Computer Science (NICS)*, 2020, pp. 314–319.



(a)

(b)

Fig. 7. Evaluating the processing time of (a) proposal (using PhoBert) and (b) LSTM.

[2] N. Q. Nhan, V. T. Son, H. T. T. Binh, and T. D. Khanh, "Crawl topical vietnamese web pages using genetic algorithm," in *2010 Second International Conference on Knowledge and Systems Engineering*, 2010, pp. 217–223.

[3] T. Xu, N. Polyakova, and S. Shipilova, "Social internet-networks in the life of vietnamese students," *SHS Web of Conferences*, vol. 28, p. 01101, Jan. 2016.

[4] T. N. Thi Thu, D. N. D. Chi, T. n. Chi, H. V. Quang, and T. N. Thi Van, "Influent factors to individual online consumer behavior: A vietnamese case study," in *2020 12th International Conference on Knowledge and Systems Engineering (KSE)*, 2020, pp. 230–235.

[5] B. T. Sinh and D. T. Hoa, "Actual status of development and application of internet of things in vietnam," *Journal SCIENCE AND TECHNOLOGY POLICIES AND MANAGEMENT*, vol. 7, no. 3, p. 5668, Feb. 2019.

[6] N. Sriratanaviriyakul, A. L. Felipe, A. Shillabeer, O. Pui, M. Nkhoma, Q. Ha Tran, and T. Cao, "Awareness and impact of vietnamese security concerns in using online social networks," July 2014.

[7] V. Ho, D. Nguyen, D. Nguyen, L. Pham, D.-V. Nguyen, K. Nguyen, and N. Nguyen, *Emotion Recognition for Vietnamese Social Media Text*, July 2020, pp. 319–333.

[8] N. Nguyen, P. Phan, L. Thanh Nguyen, K. Nguyen, and N. Nguyen, "Vietnamese complaint detection on e-commerce websites," April 2021.

[9] H. N. Nguyen, V. Le, H. Le, and T. V. Pham, "Domain specific sentiment dictionary for opinion mining of vietnamese text," Dec. 2014.

[10] S. Chaffar and D. Inkpen, "Using a heterogeneous dataset for emotion analysis in text," May 2011, pp. 62–67.

[11] P. Nandwani and R. Verma, "A review on sentiment analysis and emotion detection from text," *Social Network Analysis and Mining*, vol. 11, 2021.

[12] A. Seyeditabari, N. Tabari, S. Gholizadeh, and W. Zadrozny, "Emotion detection in text: Focusing on latent representation," in *arXiv*, 2019.

[13] K. Cherry, "The 6 types of basic emotions and their effect on human behavior," in *Version 1*, 2021, retrieved from official website: https://www.verywellmind.com.

[14] A. Mohanta and V. K. Mittal, "Human emotional states classification based upon changes in speech production features in vowel regions," in *2017 2nd International Conference on Telecommunication and Networks (TEL-NET)*, 2017, pp. 1–6.

[15] S. S. Shanta, M. Sham-E-Ansari, A. I. Chowdhury, M. M. Shahriar, and M. K. Hasan, "A comparative analysis of different approach for basic emotions recognition from speech," in *2021 International Conference on Electronics, Communications and Information Technology (ICECIT)*, 2021, pp. 1–4.

[16] W. S. S. Khine, P. Siritanawan, and K. Kotani, "Generation of compound emotions expressions with emotion generative adversarial networks (emogans)," in *2020 59th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE)*, 2020, pp. 748–755.

[17] V. T. Tran, "Python vietnamese toolkit," in *Accessed 25, June 2021*, 2022, retrieved from official website: https://github.com/trungtv/pyvi.

[18] V.-D. Le, "Python vietnamese toolkit," in *Accessed 25, June 2022*, 2022, retrieved from official website: https://github.com/stopwords/vietnamese-stopwords.

[19] D. Q. Nguyen and A. Nguyen, "Phobert: Pre-trained language models for vietnamese," Jan. 2020, pp. 1037–1042.

[20] S. Liu, H. Tao, and S. Feng, "Text classification research based on bert model and bayesian network," in *2019 Chinese Automation Congress (CAC)*, 2019, pp. 5842–5846.

[21] A. Gillioz, J. Casas, E. Mugellini, and O. A. Khaled, "Overview of the transformer-based models for nlp tasks," in *2020 15th Conference on Computer Science and Information Systems (FedCSIS)*, 2020, pp. 179–183.

[22] J. Dong, F. He, Y. Guo, and H. Zhang, "A commodity review sentiment analysis based on bert-cnn model," in *2020 5th International Conference on Computer and Communication Systems (ICCCS)*, 2020, pp. 143–147.

[23] M. Ramina, N. Darnay, C. Ludbe, and A. Dhruv, "Topic level summary generation using bert induced abstractive summarization model," in *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*, 2020, pp. 747–752.