Multi-task Solution for Aspect Category Sentiment Analysis on Vietnamese Datasets

Hoang-Quan Dang[12], Duc-Duy-Anh Nguyen[12], Trong-Hop Do[12]
[1]Faculty of Information Science and Engineering, University of Information Technology, Ho Chi Minh City, Vietnam.
[2]Vietnam National University, Ho Chi Minh City, Vietnam.
{18520339, 18520455}@gm.uit.edu.vn, hopdt@uit.edu.vn

***Abstract*—In this article, we solved two tasks in the Vietnamese Aspect-based Sentiment Analysis problem: Aspect Category Detection (ACD) and Sentiment Polarity Classification (SPC). Besides, we proposed end-to-end models to handle the above tasks simultaneously for two domains (Restaurant and Hotel) in the VLSP 2018 Aspect-based Sentiment Analysis dataset using PhoBERT as Pre-trained language models for Vietnamese in two ways: Multi-task and Multi-task with Multi-branch approach. Both models give very good results when applied preprocessing. Specifically, the Multi-task model achieves state-of-the-art (SOTA) results in the Hotel domain of the VLSP 2018 ABSA dataset, with the F1-score being 82.55% for ACD and 77.32% for ACD with SPC. For the Restaurant domain, our Multi-task model also achieved SOTA in the ACD with SPC task by an F1-score of 71.55% and 83.29% for the ACD.**

***Index Terms*—Aspect-based Sentiment Analysis, PhoBERT, Aspect Category Detection, Sentiment Polarity Classification.**

I. Introduction

Nowadays, the popularity of the Internet has led to an explosion of enormous data sources from users, especially in fields such as e-commerce, social networks, and search engines. This has brought about continuous Artificial Intelligence (AI) development, especially Natural Language Processing (NLP). Among them, the problem of Sentiment Analysis is increasingly popular and successful in both research and commerce with the purpose is to understand the level of satisfaction through reviews of customers. However, it has not fully exploited the valuable data from the Internet because a user review can contain helpful information for companies or research institutions. Aspect-based Sentiment Analysis (ABSA) is an improvement of Sentiment Analysis that will solve that problem.

Aspect-based Sentiment Analysis is a technique of analyzing text, classifying data by aspect, and identifying sentiment polarities for each of those aspects. For example, in the Hotel domain, aspects might be the experience of customers about service, the response time of hotel complaints, or quality of room amenities. With the popularity of the Internet, smart mobile, everyone can now easily evaluate aspects of a product or a service quickly and conveniently.

The huge review response data source is a valuable resource for companies. Companies collect and extract valuable information to understand what customers want and need to drive growth. The problem of Aspect-based Sentiment Analysis is the key point to extract that information. In the field of Natural Language Processing, the ABSA problem was first proposed at SemEval-2014 Task 4 [8] by Pontiki et al. After that, there were many big competitions on analyzing the sentiment aspect were organized and achieved outstanding results such as SemEval 2015 task 12 [7], SemEval 2016 task 5 [6]. In addition, with the appearance of large, powerful language models such as BERT [1], Natural Language Processing problems have made great strides.

This article experiments with powerful, modern deep learning models for the ABSA problem on the VLSP 2018 ABSA dataset with two domains: Restaurants and Hotels. The ABSA problem that we performed determines two problems: Aspect Category Detection (ACD) and Sentiment Polarity Classification (SPC). The metric we use for the evaluation is the average micro F1-score.

In this article, we focus on introducing information related to the problem of Aspect-based Sentiment Analysis in Vietnamese language. In section II, we will present subproblems. In section III, we will present related work. Section IV is an introduction to the VLSP 2018 ABSA dataset. Section V will describe the methods used to process the data and construct the model. The experimental process will be in section VI and section VII is the discussion for the results. Finally is the conclusion and our future work VIII.

II. Problem Description

Specifically, the problem we perform is Aspect Category Sentiment Analysis (ACSA), and we divide this problem into two subproblems: Aspect Category Detection and Sentiment Polarity Classification based on detected Aspect:

A. Aspect Category Detection

Identify the entity E and attribute A of a review expressed in a given sentence. E and A should be selected from a predefined set of entity types (e.g. "ROOMS","HOTEL") and attribute labels (e.g. "PRICE","QUALITY").

B. Sentiment Polarity Classification

Each identified E#A pair must be assigned one of the following sentiment polarization labels: "Positive", "Negative", "Neutral". For example:

Vietnamese: *"Phòng rộng rãi, thoáng mát, nhân viên phục vụ tận tình."*

English: *"The room is spacious, airy, staff is enthusiastic."*
⇒ {cate: "ROOMS#DESIGN", pol: "positive"}.
⇒ {cate: "SERVICE#GENERAL", pol: "positive"}.
In which, cate is the Category or Aspect type, pol is the Polarity label.

III. Related Work

In the SemEval-2016 Task 5: Aspect Based Sentiment Analysis [6] article, Maria Pontiki et al. described the ABSA task. During the third year, the task provided 19 training datasets and 20 test datasets for 8 languages and 7 domains and a general evaluation process. Out of these datasets, 25 for the sentence level and 14 for the text level ABSA; the last one was first introduced as a subtask in SemEval. The mission attracted 245 entries from 29 teams. The three sub-tasks in the paper are Sentence analysis based on sentence-level aspect, Sentiment analysis based on text level aspect, and Sentiment analysis based on aspects in external domains. SemEval ABSA 16 is the first to include a text allowance quest.

The article Deep Learning for Aspect-based Sentiment Analysis: A Comparative Review [2] by Hai Ha Do et al. in 2018 pointed out that the increasing volume of user-generated content on the web has made Sentiment Analysis an essential tool for extracting information about people's sentiment states. This article provides a comparative review of deep learning for Aspect-based Sentiment Analysis to put different approaches in context.

In the paper Aspect-based Sentiment Analysis of movie reviews on discussion boards [10], the authors implemented an automated Sentiment Analysis method of proposed, performed and rated film reviews. This method uses detailed analysis to determine both the sentiment orientation and sentiment strength of the reviewer towards different aspects of a film. It applies a linguistic approach to calculate the sentiment of a term from a previous sentiment score.

The paper A Hierarchical Model of Reviews for Aspect-based Sentiment Analysis is proposed by Sebastian Ruder et al. in 2016 [9] represents methods of extracting opinions from customer reviews that have become popular in recent years. Sentences are usually classified independently, even though they are part of the review's argument structure. The paper shows that hierarchical two-way LSTM outperforms two non-hierarchical baselines and competes with the most advanced technology on five multilingual, multi-domain datasets without any throttling features, any manual or external resource.

The paper Aspect Based Sentiment Analysis with Gated Convolutional Networks [14] published in 2018 by Wei Xue, Tao Li showed that Gated Tanh-ReLU Units could selectively output affective features according to aspect or a specific entity. The architecture is much simpler than the attention layers used in current models. They propose a model based on an accumulative neural network and a localization mechanism.

SemEval-2015 Task 12: Aspect Based Sentiment Analysis [7] by author Maria Pontiki et al. described the SE-ABSA15 task that attracted 93 submissions from 16 teams that were evaluated in three positions: aspect classification, expression express objective point of view, and polar classification. Future works include applying the new framework to other languages (e.g., Spanish, Greek) and annotating topics or events. There are two sub-tasks in this task: Aspect-based Sentiment Analysis in the inner domains and outer domains.

The paper A BERT-based Hierarchical Model for Vietnamese Aspect Based Sentiment Analysis [11] by author Oanh Thi Tran et al. described (ABSA) represents the task of identifying sentiment poles for specific entities and their aspects mentioned in customer reviews. This paper presents a new and efficient hierarchical model using a pre-trained language model representing a bidirectional encoding from a transformer (BERT).

The paper VLSP shared task: Sentiment Analysis [5] by Nguyen Thi Minh Huyen et al. presented a series of VLSP seminars on Sentiment Analysis (SA) for Vietnamese that was organized to measure customer evaluation performance (quality) of Sentiment Analysis tools. This paper describes the built data sets and the evaluation results of the systems participating in these campaigns.

The paper Multi-task Learning for Aspect and Polarity Recognition on Vietnamese Datasets [12] is a study conducted by Thin Van Dang et al. that developed a deep neural network model to solve two tasks in Sentiment Analysis based on the document-level aspect of the Vietnamese dataset. They found that their model performed better than the steady baseline approaches on both tasks for the two domains and achieved an F1-score of 64.78% for the restaurant domain and 70.90% for the hotel domain.level aspect, Sentiment analysis based on text level aspect, and Sentiment analysis based on aspects in external domains. SemEval ABSA 16 is the first to include a text allowance quest.

The article Deep Learning for Aspect-based Sentiment Analysis: A Comparative Review [2] by Hai Ha Do et al. in 2018 pointed out that the increasing volume of user-generated content on the web has made Sentiment Analysis an essential tool for extracting information about people's sentiment states. This article provides a comparative review of deep learning for Aspect-based Sentiment Analysis to put different approaches in context.

In the paper Aspect-based Sentiment Analysis of movie reviews on discussion boards [10], the authors implemented an automated Sentiment Analysis method of proposed, performed and rated film reviews. This method uses detailed analysis to determine both the sentiment orientation and sentiment strength of the reviewer towards different aspects of a film. It applies a linguistic approach to calculate the sentiment of a term from a previous sentiment score.

The paper A Hierarchical Model of Reviews for Aspect-based Sentiment Analysis is proposed by Sebastian Ruder et al. in 2016 [9] represents methods of extracting opinions from customer reviews that have become popular in recent years. Sentences are usually classified independently, even though they are part of the review's argument structure. The paper shows that hierarchical two-way LSTM outperforms two non-hierarchical baselines and competes with the most advanced

technology on five multilingual, multi-domain datasets without any throttling features, any manual or external resource.

The paper Aspect Based Sentiment Analysis with Gated Convolutional Networks [14] published in 2018 by Wei Xue, Tao Li showed that Gated Tanh-ReLU Units could selectively output affective features according to aspect or a specific entity. The architecture is much simpler than the attention layers used in current models. They propose a model based on an accumulative neural network and a localization mechanism.

SemEval-2015 Task 12: Aspect Based Sentiment Analysis [7] by author Maria Pontiki et al. described the SE-ABSA15 task that attracted 93 submissions from 16 teams that were evaluated in three positions: aspect classification, expression express objective point of view, and polar classification. Future works include applying the new framework to other languages (e.g., Spanish, Greek) and annotating topics or events. There are two sub-tasks in this task: Aspect-based Sentiment Analysis in the inner domains and outer domains.

The paper A BERT-based Hierarchical Model for Vietnamese Aspect Based Sentiment Analysis [11] by author Oanh Thi Tran et al. described (ABSA) represents the task of identifying sentiment poles for specific entities and their aspects mentioned in customer reviews. This paper presents a new and efficient hierarchical model using a pre-trained language model representing a bidirectional encoding from a transformer (BERT).

The paper VLSP shared task: Sentiment Analysis [5] by Nguyen Thi Minh Huyen et al. presented a series of VLSP seminars on Sentiment Analysis (SA) for Vietnamese that was organized to measure customer evaluation performance (quality) of Sentiment Analysis tools. This paper describes the built data sets and the evaluation results of the systems participating in these campaigns.

The paper Multi-task Learning for Aspect and Polarity Recognition on Vietnamese Datasets [12] is a study conducted by Thin Van Dang et al. that developed a deep neural network model to solve two tasks in Sentiment Analysis based on the document-level aspect of the Vietnamese dataset. They found that their model performed better than the steady baseline approaches on both tasks for the two domains and achieved an F1-score of 64.78% for the restaurant domain and 70.90% for the hotel domain.

IV. Dataset

In this article, we have used the VLSP 2018 Aspect based sentiment analysis[1] dataset for 2 data domains: Restaurants and Hotels.

Table I: VLSP 2018 Aspect-based Sentiment Analysis dataset

Domain	Dataset	#Reviews	#Aspects	AvgLength	VocabSize	#DiffVocab
Restaurant	Training	2,961	9,034	54	5,168	-
	Dev	1,290	3,408	50	3,398	1,702
	Test	500	2,419	163	3,375	1,729
Hotel	Training	3,000	13,948	47	3,908	-
	Dev	2,000	7,111	23	2,745	1,059
	Test	600	2,584	30	1,631	346

[1]https://vlsp.org.vn/vlsp2018/eval/sa

Table I presents the VLSP 2018 Aspect-based Sentiment Analysis dataset on two domains: Hotels and Restaurants. The columns are respectively domain of reviews (Domain), number of reviews (#Reviews), number of aspects (#Aspects), average review length (AvgLength), vocabulary size (VocabSize), number of words in Test or Dev set that do not appear in the Training set (#DiffVocab).

Table II: VLSP 2018 ABSA Hotel reviews.

Review	Label
Vietnamese: Rộng rãi KS mới nhưng rất vắng. Các dịch vụ chất lượng chưa cao và thiếu. *English: Wide new hotel but very empty. The qualified services are not high and lack*	{HOTEL#DESIGN&FEATURES,positive}, {HOTEL#DESIGN&GENERAL,negative}
Vietnamese: Phục vụ, view đẹp, vị trí *English: Service, nice view, location*	{SERVICE#GENERAL,positive}, {HOTEL#GENERAL,positive}, {LOCATION#GENERAL,positive}

Table III: VLSP 2018 ABSA Restaurant reviews.

Review	Label
Vietnamese: Ngon tuyệt vời luôn! *English: That's fantastic!*	{FOOD#QUALITY,positive}
Vietnamese: Gà rán giòn ướp vị chua cay rất ngon, bánh ăn kèm giòn và thơm mùi bơ *English: Crispy fried chicken marinated with sour and spicy taste is very delicious, side dish is crispy and have a smell of butter*	{FOOD#STYLE&OPTIONS,positive}, {FOOD#QUALITY, positive}

Tables II and III are examples of the hotel and the restaurant domain of the VLSP 2018 dataset, respectively.

V. Method

A. Data preprocessing

Natural language processing in general and Vietnamese processing in particular, the first step is to preprocess the data. If this step is done well, it can increase performance. Therefore, data preprocessing is the process of normalizing data and removing non-significant components. So we designed a pipeline to process the data in order to get better information from the original data. These steps are described as below 1:

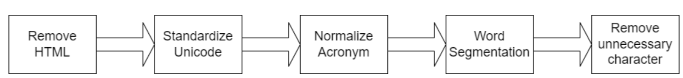

Figure 1 Preprocessing Pipeline

First, we deleted the HTML codes in the original dataset. We then performed the charset standardization from Windows-1252 to UTF-8 and standardized common Vietnamese abbreviations for each domain in the dataset. In Vietnamese, a word can be made up of two or more other words (known as a compound word), such as "nhà hàng" (each word has its meaning when standing alone and another meaning when combined). Therefore, it is necessary to do the segmentation for Vietnamese words before starting further processing, and we used the VnCoreNLP toolkit [13] to implement that step.

Finally, we removed unnecessary characters, this helps to reduce the number of feature dimensions for sentence, increase the speed, and avoids bad influence on the model results.

B. PhoBERT

PhoBERT [4] is a large-scale monolingual language model pre-trained for Vietnamese. PhoBERT outperforms previous monolingual and multilingual approaches, obtaining new state-of-the-art performances on four downstream Vietnamese NLP tasks: Part-of-speech tagging, Dependency parsing, Named entity recognition, and Natural language inference.

PhoBERT comes in two versions, PhoBERT$_{base}$ and PhoBERT$_{large}$. PhoBERT pre-training approach is based on RoBERTa [3], which optimizes the BERT pre-training procedure for more robust performance.

C. Model Architecture

We implemented end-to-end models with the same architecture in two approaches: Multi-task and Multi-task with Multi-branch to handle both ACD and SPC task simultaneously using PhoBERT$_{base}$ as a pre-trained language model for Vietnamese. Based on the original BERT paper [1], the model achieved the best results when concatenating the last four layers of BERT together. So we applied that method to our model architecture with the input transformed into a low-dimensional vector $x_I \in R^d$, where d is the length of the vector.

Table IV: Results with different methods of BERT

Feature-based approach (BERT$_{\text{BASE}}$)	
Embeddings	91.0
Second-to-Last Hidden	95.6
Last Hidden	94.9
Weighted Sum Last Four Hidden	95.9
Concat Last Four Hidden	96.1
Weighted Sum All 12 Layers	95.5

D. Multi-task Approach

The model is constructed with the output as a list of C one-hot vectors, where C is the number of aspects present in the dataset used. For the VLSP dataset, the number of aspects C will be 12 for the Restaurant dataset and 34 for the Hotel dataset, respectively. Each vector has four components corresponding to three labels of the polarities, including Positive, Negative, Neutral, and one None label to indicate whether or not the input has this aspect so that it can have a polarity. Since this is a one-hot vector, any label that exists will be assigned 1, and the rest will be 0.

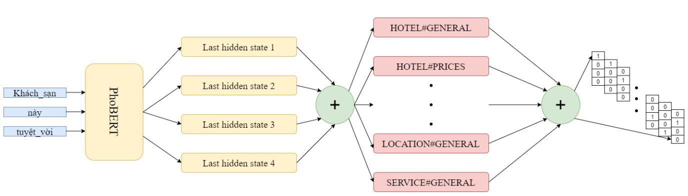

Figure 2 Multi-task model for Hotel domain

The learned feature (g) is passed through a fully connected layer created by concatenating C dense layers corresponding to C one-hot vectors. Thus, we will get a dense layer consisting of 48 neurons for the Restaurant domain (12 aspects $\times$ 4 polarities), 136 neurons for the Hotel domain (34 aspects $\times$ 4 polarities) 2, and using use the softmax function to calculate the score for the predicted value y_{pred} for each aspect A:

$$y_{\text{pred}} = \text{softmax}(W^{(a)} \cdot g + b^{(a)}). \tag{1}$$

So we can predict an aspect a and its polarity in one step by:

$$\text{output}^{(a)} = \arg\max_i \hat{y}_i^{(a)} \quad \text{where } i = 0, 1, 2, 3. \tag{2}$$

For the binary classification problem, the loss function will take the form:

$$L(y, \hat{y}) = -\sum_{i=1}^{N} y_i, \log(\hat{y}_i) + (1 - y_i) \cdot \log(1 - \hat{y}_i). \tag{3}$$

In the case of classification problems with C labels (C > 2) and at the same time we use the softmax function to calculate the output probability distribution, the loss function will be a cross-entropy function as follows:

$$L(y, \hat{y}) = -\sum_{i=1}^{N} \sum_{j=1}^{C} y_{ij} \cdot \log(\hat{y}_{ij}) \tag{4}$$

In the Multi-task learning approach, for each classification task, there will be a loss function with a value of equation (3). So when there are C different classification tasks, the loss function of multi-task learning will be the sum of the loss functions (in binary cross-entropy form) of each binary classification problem for each task:

$$L(y, \hat{y}) = -\sum_{i=1}^{N} \sum_{j=1}^{C} y_{ij} \cdot \log(\hat{y}_{ij}) + (1 - y_{ij}) \cdot \log(1 - \hat{y}_{ij}). \tag{5}$$

E. Multi-task with Multi-branch Approach

The only difference of this approach from the above one is that it will branch into many submodels by using C fully connected layers but not concatenating them into a single one 3. Each model predicts each task independently.

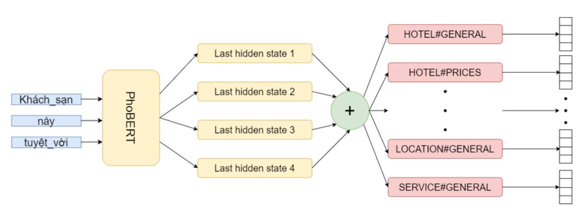

Figure 3 Multi-task with Multi-branch approach for Hotel domain

The model uses the softmax function to calculate the scores and can simultaneously predict its aspect and polarity in a similar way to the above model. The final loss function will be equal to the sum of loss functions in all C branches:

$$L(W; X) = -\sum_{i=1}^{N} \sum_{j=1}^{C} y_{ij} \cdot \log(\hat{y}_{ij}). \tag{6}$$

VI. Experimental Setup

We did the training in Google Colab environment with a TPU distribution strategy that we used to distribute model training onto the 8 separate TPU cores available on this one TPU worker. We use a batch size of 21 for the Restaurant dataset and a batch size of 20 for the Hotel dataset with the input padded according to the maximum number of tokens recommended by PhoBERT, namely 256 tokens. We also choose 20 epochs and uses Early Stopping to stop training if the validation set loss has not improved after 5 epochs.

For fine-tuning, we use the same optimization algorithm for which BERT pre-training was set up: Adaptive Moments (Adam). This optimizer minimizes the loss in prediction and performs regularization by decomposing or decreasing weight (without using moments), also known as AdamW. For the learning rate, we use the same learning rate schedule as BERT pre-training: linear decay of an initial learning rate, preceded by a linear warm-up period in 10% the first of the training steps. Here, We initialize a learning rate of 2e-4.

VII. Result and Discussion

In addition to the Multi-task approach and Multi-task with Multi-branch approach, we have presented some different results to compare our outcome. Specifically, we referenced the best submission model of VLSP from VLSP SHARED TASK: SENTIMENT ANALYSIS [5], the model Bi-LSTM+CNN referenced from Multi-task Learning for Multi-task Learning for Aspect and Polarity Recognition on Vietnamese Datasets [12], and the BERT-based Hierarchical model from A BERT-based Hierarchical Model for Vietnamese Aspect Based Sentiment Analysis [11].

Table V: Model results on the VLSP 2018 ABSA Hotel dataset, with and without preprocessing.

VLSP 2018 ABSA - Hotel + Preprocessing				
Task	**Method**	**Precision**	**Recall**	**F1-score**
Aspect Detection	VLSP best submission	76.00	66.00	70.00
	Bi-LSTM+CNN	84.03	72.52	77.85
	BERT-based Hierarchical	-	-	82.06
	Multi-task	**87.45**	**78.17**	**82.55**
	Multi-task Multi-branch	63.21	57.86	60.42
Aspect + Polarity	VLSP best submission	66.00	57.00	61.00
	Bi-LSTM+CNN	76.53	66.04	70.90
	BERT-based Hierarchical	-	-	74.69
	Multi-task	**81.90**	**73.22**	**77.32**
	Multi-task Multi-branch	57.55	52.67	55.00
VLSP 2018 ABSA - Hotel				
Task	**Method**	**Precision**	**Recall**	**F1-score**
Aspect Detection	VLSP best submission	76.00	66.00	70.00
	Bi-LSTM+CNN	84.03	72.52	77.85
	BERT-based Hierarchical	-	-	**82.06**
	Multi-task	87.32	76.51	81.56
	Multi-task Multi-branch	62.75	57.70	60.12
Aspect + Polarity	VLSP best submission	66.00	57.00	61.00
	Bi-LSTM+CNN	76.53	66.04	70.90
	BERT-based Hierarchical	-	-	74.69
	Multi-task	**80.79**	**70.78**	**75.45**
	Multi-task Multi-branch	57.20	52.59	54.80

Table V shows the results of the VLSP 2018 ABSA dataset in the hotel domain when there isn't preprocessing and without preprocessing. Applying preprocessing data, the Multi-task model gives the best results on Aspect Category Detection, Aspect Category Detection with Sentiment Polarity Classification higher than the BERT-based Hierarchical's current SOTA. Specifically, the micro F1-score of ACD and ACD + SPC is 82.55% and 77.32%. For results without preprocessing data, Multi-task gives the best results on ACD + SPC task with F1-score micro of 75.45% and good results on ACD task with F1-score micro of 81.56% only lower to BERT-based Hierarchical (F1-score micro-82.06%). The results between Multi-task and Multi-task with Multi-branch differ by approximately 30% in preprocessing and without preprocessing for ACD, ACD + SPC tasks.

Table VI: Model results on the VLSP 2018 ABSA Restaurant dataset, with and without preprocessing.

VLSP 2018 ABSA - Restaurant + Preprocessing				
Task	**Method**	**Precision**	**Recall**	**F1-score**
Aspect Detection	VLSP best submission	79.00	76.00	77.00
	Bi-LSTM+CNN	82.02	77.51	79.70
	BERT-based Hierarchical	-	-	**84.23**
	Multi-task	81.09	85.61	83.29
	Multi-task Multi-branch	80.81	87.39	83.97
Aspect + Polarity	VLSP best submission	62.00	60.00	61.00
	Bi-LSTM+CNN	66.66	63.00	64.78
	BERT-based Hierarchical	-	-	71.30
	Multi-task	**69.66**	**73.54**	**71.55**
	Multi-task Multi-branch	68.69	74.29	71.38
VLSP 2018 ABSA - Restaurant				
Task	**Method**	**Precision**	**Recall**	**F1-score**
Aspect Detection	VLSP best submission	79.00	76.00	77.00
	Bi-LSTM+CNN	82.02	77.51	79.70
	BERT-based Hierarchical	-	-	**84.23**
	Multi-task	80.07	81.07	80.57
	Multi-task Multi-branch	81.56	79.33	80.43
Aspect + Polarity	VLSP best submission	62.00	60.00	61.00
	Bi-LSTM+CNN	66.66	63.00	64.78
	BERT-based Hierarchical	-	-	**71.30**
	Multi-task	68.68	69.53	69.10
	Multi-task Multi-branch	69.02	67.14	68.06

Table VI shows the results of the VLSP 2018 ABSA dataset for the restaurant domain when there isn't preprocessing and without preprocessing. Applying the preprocessing data, Multi-task and Multi-task with multi-branch models gives the best results on Aspect Category Detection with Sentiment Polarity Classification higher than the BERT-based Hierarchical model's current SOTA. Specifically, the F1-score micro of ACD + SPC is 71.55% and 71.38%, while both models' Aspect Category Detection task on the F1-score micro is only 83.29% and 83.97%, the difference with BERT-based Hierarchical is about approx 1% - 0.26%. For results without preprocessing, the BERT-based Hierarchical model gives the best results in both ACD and ACD + SPC tasks, obtained from Multi-task and Multi-task with Multi-branch are also very good. Specifically, Multi-task is 80.57%, 69.10%, and Multi-task with Multi-branch is 80.43%, 68.06% in both ACD and ACD + SPC tasks. Multi-task has slightly better results than Multi-task with Multi-branch, about 1% in preprocessing data and without preprocessing data in tasks except for ACD tasks with preprocessing.

In general, preprocessing improves the accuracy of models in both the restaurant and hotel sectors.

VIII. Conclusion and Future work

In this article, we have solved the problem of Aspect-based Sentiment analysis on the VSLP 2018 ASBA dataset. We have implemented models as PhoBERT + Multi-task and PhoBERT + Multi-task with Multi-branch approach. In which, the Multi-task model gives the best results on two tasks for both Hotel and Restaurant domains when applying the preprocessing. The Multi-task model with Multi-branch approach gives good results for the Restaurant domain. Specifically, the Multi-task model achieves state-of-the-art (SOTA) results for the Hotel domain, while with the Restaurant domain, the model also achieves SOTA but not significantly compared to the old one. In addition, through the experimental process, we found that preprocessing helps to increase model performance.

In the future, we plan to explore other approaches for this problem such as Graph models, BERT variants. Changed Embedding classes for multi-task model: PhoW2V, fastText, multi Embedding, vELECTRA. Using data augmentation methods for the training set such as Translate languages, Paraphrase, Word similarity, etc... to increase the model's efficiency.

References

[1] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[2] Hai Ha Do, PWC Prasad, Angelika Maag, and Abeer Alsadoon. Deep learning for aspect-based sentiment analysis: a comparative review. *Expert Systems with Applications*, 118:272–299, 2019.

[3] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*, 2019.

[4] Dat Quoc Nguyen and Anh Tuan Nguyen. Phobert: Pre-trained language models for vietnamese. *arXiv preprint arXiv:2003.00744*, 2020.

[5] Huyen TM Nguyen, Hung V Nguyen, Quyen T Ngo, Luong X Vu, Vu Mai Tran, Bach X Ngo, and Cuong A Le. Vlsp shared task: sentiment analysis. *Journal of Computer Science and Cybernetics*, 34(4):295–310, 2018.

[6] Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Mohammad AL-Smadi, Mahmoud Al-Ayyoub, Yanyan Zhao, Bing Qin, Orphée De Clercq, Véronique Hoste, Marianna Apidianaki, Xavier Tannier, Natalia Loukachevitch, Evgeniy Kotelnikov, Nuria Bel, Salud María Jiménez-Zafra, and Gülşen Eryiğit. SemEval-2016 task 5: Aspect based sentiment analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 19–30, San Diego, California, June 2016. Association for Computational Linguistics.

[7] Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Suresh Manandhar, and Ion Androutsopoulos. SemEval-2015 task 12: Aspect based sentiment analysis. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 486–495, Denver, Colorado, June 2015. Association for Computational Linguistics.

[8] Maria Pontiki, Dimitris Galanis, John Pavlopoulos, Harris Papageorgiou, Ion Androutsopoulos, and Suresh Manandhar. SemEval-2014 task 4: Aspect based sentiment analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 27–35, Dublin, Ireland, August 2014. Association for Computational Linguistics.

[9] Sebastian Ruder, Parsa Ghaffari, and John G Breslin. A hierarchical model of reviews for aspect-based sentiment analysis. *arXiv preprint arXiv:1609.02745*, 2016.

[10] Tun Thura Thet, Jin-Cheon Na, and Christopher SG Khoo. Aspect-based sentiment analysis of movie reviews on discussion boards. *Journal of information science*, 36(6):823–848, 2010.

[11] Oanh Thi Tran and Viet The Bui. A bert-based hierarchical model for vietnamese aspect based sentiment analysis. In *2020 12th International Conference on Knowledge and Systems Engineering (KSE)*, pages 269–274, 2020.

[12] Dang Van Thin, Duc-Vu Nguyen, Kiet Van Nguyen, Ngan Luu-Thuy Nguyen, and Anh Hoang-Tu Nguyen. Multi-task learning for aspect and polarity recognition on vietnamese datasets. In *International Conference of the Pacific Association for Computational Linguistics*, pages 169–180. Springer, 2019.

[13] Thanh Vu, Dat Quoc Nguyen, Dai Quoc Nguyen, Mark Dras, and Mark Johnson. VnCoreNLP: A Vietnamese natural language processing toolkit. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations*, pages 56–60, New Orleans, Louisiana, June 2018. Association for Computational Linguistics.

[14] Wei Xue and Tao Li. Aspect based sentiment analysis with gated convolutional networks. *arXiv preprint arXiv:1805.07043*, 2018.