A Joint Multi-task Architecture for Document-level Aspect-based Sentiment Analysis in Vietnamese

Dang Van Thin, Lac Si Le, Hao Minh Nguyen, and Ngan Luu-Thuy Nguyen

***Abstract*—The increasing demands of e-commerce websites have caused a vast amount of opinions about Internet users' products and services. Therefore, the aspect-based sentiment analysis (ABSA) has attracted much attention from academia and industry due to its application in many real-life problems in recent years. This problem is a complex task that aims to extract both the aspects and sentiments from the text input. Aspect Category Detection and Sentiment Polarity Classification are two of the challenging subtasks of ABSA, which detect a set of pre-defined categories and corresponding sentiment polarity for a given review. This paper presents an effective joint multi-task architecture based on neural network models to solve two tasks in the document-level ABSA datasets. Our model is designed to predict the whole mentioned aspect categories and corresponding sentiment polarities on the document-level datasets. We trained our model jointly on two tasks simultaneously and utilized the additional information of aspect category detection task for predicting the aspect categories and its sentiments for the specific domain. Our architecture can explore the hidden correlated information between categories and polarities in the review. Experiments on two Vietnamese language datasets in the restaurant domain and hotel domain demonstrated that our model outperforms the previous state of the art methods on two benchmark document-level dataset.**

***Index Terms*—Aspect-based sentiment analysis, deep neural network, multi-task learning, Vietnamese corpora, document-level dataset.**

I. Introduction

With the development of information technology, the demand for online shopping has become a new social trend. The users can easily order an item via e-commerce websites as Amazon, eBay, Alibaba. Besides, we also book a hotel and restaurant through websites like TripAdvisor or Agoda. However, we always wonder about the quality and effectiveness of the products we want to use based on the website's descriptive information. Therefore, these websites often have sections for users to express their opinions about the products, services. There are many benefits when analyzing these comments. First, other users can refer to these reviews of the product before making a decision [1]. Second, the businessman can rely on those reviews to analyze the users' experiences about their product. In terms of business, the users' comments are precious free resources to discover the advantages and disadvantages of the product. Based on these analyses, they help the manufacturers to improve product quality [2]. With the vast amount of customer-generated data online, it is challenging to analyze manually by the human. For that reason, Sentiment Analysis has attracted the attention of a wide variety of researchers in academia and industry. Sentiment Analysis (SA) is an essential task of extracting opinions, subjectivity, sentiment information from online text documents. The primary purpose of this task is to identify the overall sentiment polarity from user-generated documents [3].

However, the users' reviews often contain also many different aspects, for example, "The staff serves good, but the food is bad". We can see that there are two aspects that are mentioned in the above example and the sentiment polarities of are different, i.e. the polarity of "Service" is positive. In contrast, the polarity of "Food#Quality" category is negative. A branch of SA task which identifies sentiments associated with specific aspects in the text input is called Aspect-based Sentiment Analysis (ABSA). This task helps us understand the users' opinions because it directly focuses on each aspect's sentiment rather than overall reviews. The first official international workshop about ABSA task was introduced in SemEval-2014 Task 4 [4] and continued over the next two years, SemEval-2015 Task 12 [5] and SemEval-2016 Task 5 [6]. In these workshops, three main tasks were introduced for the ABSA task included: Aspect Category Detection (ACD), Opinion Target Expressions (OTE), and Sentiment Polarity Classification (SPC) for two levels of text, including the sentence-level and document-level. The ACD task is to assign the aspect categories based on a pre-defined set. The OTE task is to detect the expression on the review mentioned the aspect while the SPC is to assign the sentiment polarity for ACD task and OTE task. For example, given a review for the restaurant domain as follow: "The staff serves good but the food is bad", the output of ACD task is "Service#General" and "Food#Quality", the output of OTE task is "staff" and "food", while the output of SPC is the "positive" for "Service#General" and "staff" term, "negative" for "Food#Quality" category and "food" term. We can see that it is a challenging task, especially if a review contains many aspects with different sentiment polarity polarities [7]. Besides, there are high quality datasets at the sentence-level with manual annotation on various language, i.e. English, Arabic, Chinese, are published for the research community from SemEval workshops.

For the Vietnamese language, Vietnamese Language and Speech Processing (VLSP) organized an ABSA workshop

Manuscript received February 1, 2021; revised December 2, 2021. This work is funded by Vietnam National University HoChiMinh City (VNU-HCM) under grant number B2019-26-01.

Dang V. Thin, Lac S. Le, Hao M. Nguyen and Ngan L.T Nguyen are with the University of Information Technology, Vietnam National University Ho Chi Minh City, Vietnam (e-mail: {thindv, ngannlt}@uit.edu.vn, 17520669@gm.uit.edu.vn, haohaoit@gmail.com).

doi: 10.18178/ijmlc.2022.12.4.1091

for the Vietnamese language in 2018. They provided two benchmark datasets for the restaurant and hotel domain [8]. Each dataset is annotated for the aspect category detection and sentiment polarity classification task. The participating teams were required to solve two tasks for the final output. It means the output of participation systems must detect the aspect categories and corresponding sentiment polarities mentioned in the reviews - called as Aspect Category with Polarity (ACP) task. The number of aspect categories follow by SemEval 2016 workshop Task 6 [6] and the number of sentiment polarities consist of three states: “positive”, “negative”, and “neutral”. These datasets are demanding for the ABSA task in terms of processing memory and training time, where each review can be assigned many aspect categories with different polarities. The length of each review is not uniform in three sets. Another challenge is the imbalance of the aspect categories and polarities in these datasets.

More recently, the pre-trained Bidirectional Encoder Representations from Transformer (BERT) language model [9] has established new state of the art results in aspect-based sentiment analysis datasets (Hoang *et al*. [10], Sun *et al.* [11], Li *et al.* [12], Li *et al*. [13]). However, the limit point of the BERT model [9] is that the input sequence length is limited to 512 tokens; therefore, the text pre-processing technique is applied for the reviews with a length larger than 512. To solve this problem, Sun *et al.* [14] proposed the truncation and hierarchical methods to fine-tune BERT for text classification tasks. The authors experimented and demonstrated the effectiveness of truncation methods for the long text. This method includes the head-only (keep the first 510 tokens of the input), tail-only (keep the last 510 tokens of the input), and head+tail (combination first 128 tokens and the last 382 tokens). However, we cannot apply the BERT architecture for the document-level datasets based on the truncation methods for two following reasons: (1) the maximum input sequence length of pre-trained BERT is 512, and the length of reviews at document-level often bigger than 512 in our case, (2) we cannot apply the truncation technique to reduce the length of the review because the aspect categories may appear anywhere in the long reviews, the truncations lead to missing information of aspect mentioned in the review. Therefore, it is difficult to apply the BERT model for the document-level datasets in ABSA problem.

In this paper, we present a joint multi-task architecture based on neural networks which are efficient for the prediction of the aspect categories and corresponding sentiment analysis for document-level data. Our architecture is able to predict whole aspect category with its sentiment for the specific domain. Our architecture train two tasks at the same time and utilize the information features of ACD task with the features of ACP task as the final representation of the long review input. This representation is put into the fully connected layer with the number of softmax layers to predict the aspect category with corresponding sentiment. Experimental results demonstrated the effectiveness of our framework on two datasets at the document-level for the Vietnamese language than previous methods include BERT architecture [15].

The rest of this paper is organized as follows: The related work is introduced in Section II. Section III presents our architecture, while Section IV gives detailed experiments on two benchmark datasets. Results and discussion are provided in Section V. Finally, Section VI summaries the paper and provides our future research directions.

II. Related Works

In recent years, there have been tremendous studies on aspect-based sentiment analysis problems. There are many survey papers about aspect-based sentiment analysis studies (Schouten and Frasincar [16], Zhang *et al.* [1], Do *et al.* [17], Nazir *et al.* [3]). In these papers, various approaches are summarized, including directions based on knowledge-based, machine learning, and hybrid for each different task in ABSA. In the latest survey, Nazir *et al.* [3] discussed the issues and challenges. They also presented the progress of recent studies and given future research directions. Most recent studies try to solve through multi-task learning approach based on deep learning architectures. Xue *et al.* [18] proposed a MTNA framework based on BiLSTM and CNN for the Aspect Category Detection and Opinion Target Expressions tasks because they noted that there is a close relationship between the two tasks. Their framework can share the mutual information of two tasks and achieved positive results on various SemEval datasets. Then, Wang *et al.* [19] also proposed a novel multi-task neural learning for opinion target expressions and sentiment polarity classification task by leveraging attention mechanisms. Similarly, Schmitt *et al.* [20] presented a jointly model based on a deep neural network for the Aspect Category Detection and Sentiment Polarity Classification tasks. They experimented with the CNN and LSTM architecture combined with the fasttext embedding on the GermEval 2017 dataset. Their model is a type of end-to-end trainable model which predicts the number of aspects and corresponding sentiment polarity at the same time. They formatted the model output as a set of 4 class vectors (none, positive, negative and neutral) corresponding to each aspect category in the specific domain. Experimental results have shown that the end-to-end CNN outperformed to the best system on the GermEval datasets.

Then, Dai *et al*. [21] also presented a Multi-task Multi-head Attention Memory Network (MMAM), which use the shared document and category features for aspect category detection and sentiment polarity classification for the Chinese datasets. They compared their framework with other multi-task architectures and achieved the comparable results on two datasets. Recently, He *et al.* [22] introduced a interactive multi-task learning network (IMN) and shown the superior performance on three benchmark datasets. The IMN can train multiple related tasks simultaneously by using a message-passing mechanism to interact between tasks. The useful information will be sent back to a shared latent representation between tasks. This information will be repeated to update and propagate across multiple links for all tasks. The performance of the IMN framework is optimized through iterative training. The architectures used in the IMN framework are CNN model combined with word embedding. Their results have shown that the IMN model is better than multiple baselines for Opinion Target Expressions and Sentiment Polarity Classification task. Based on proposed of He *et al*., Liang *et al.* [23] presented an Iterative Knowledge

Transfer Network (IKTN) model for the end-to-end ABSA task. There is an iterative knowledge transfer network, which can exploit the semantic relationships between tasks. Their approach significantly achieved new state-of-the-art results on three benchmark datasets for the OTE and SPC tasks. In very recently, with the development of pre-trained language model BERT [9], there are many studies have combined the deep contextualized word embedding layer with neural models and achieved the new state-of-the-art results on various ABSA tasks (Hoang *et al.* [10], Sun *et al.* [11], Li *et al.* [12], Li *et al.* [14]). Most of the above studies experimented on the sentence-level benchmark dataset from SemEval workshops. However, in this paper, we focus on the problem of the document-level dataset, where each sample is usually with a length larger than the input of BERT architecture.

For the Vietnamese language, there are a few research studies on the ABSA in recent years included public benchmark datasets (Nguyen *et al.* [8], Thuy *et al.* [24], Nguyen *et al.* [25], Thuy *et al.* [26]) and presented the methodology (Thin *et al*. [27], Thin *et al*. [28], Tran and Phan [29], Le *et al*. [15]). To the best of our knowledge, the research of Nguyen *et al.* [8] was the first study to publish benchmark datasets for the research community on the ABSA problem, which have same format of shared-task SemEval 2016 [6]. Their datasets are annotated at document-level users' reviews and split into the training, validation, and testing set for the hotel and restaurant domain. These datasets are very challenging because of the difference between the training set and testing set related to the number of samples and the reviews' length. Following that, Thuy *et al.* [24] also presented a manually annotated dataset at the sentence-level for the ACD task with 6 472 sentences (3 796 in Vietnamese and 2 676 in English) for the restaurant domain. Next, Thuy *et al.* [26] continually annotated the SPC task for this dataset and combined with translated English dataset [6] for the final dataset. Similarly, Nguyen *et al.* [25] also presented a dataset at the document-level for ABSA with two tasks: ACD and SPC for the restaurant reviews. Compared with other datasets, their dataset was annotated with 7 aspect categories and 5 sentiment polarities. Most of the dataset studies presented different baseline methods, they mostly used the Support Vector Machine model combined with various handcraft features [24]-[26]. The works of Thin *et al.* [27], Thuy *et al.* [24] and Thuy *et al.* [26] also separate two tasks (ACD and SPC) as two components where each component consist of N binary classifiers corresponding to N aspect categories. In contrast, Nguyen *et al.* [25] combined the two outputs of two tasks in a classifier for each aspect category. In our opinions, these approaches resolve aspect categories individually without utilizing relevant information between the aspect categories. In recent, Thin *et al.* [30] proposed a joint architecture based on deep learning methods to address two tasks on two VLSP datasets [8]. Next, Le *et al.* [15] presented the experimental results using Multilingual BERT-based for the VLSP datasets [8]. Because the VLSP datasets are the document-level dataset where each review is made up of many sentences and the input length of BERT architecture is limited by 512 tokens; therefore, they break each review into sentences and then put them into BERT architecture. They also resolve two tasks in the VLSP dataset separately; they treated two tasks as the multi-label classification problem. However, their experimental results are not effective than the joint architecture of Thin *et al.* [30].

Accordingly, in this paper, we present a joint architecture based on neural networks for the document-level dataset, which can train all aspect categories and predict two tasks into an architecture. In our architecture, we take advantages of the ACD task features by combining with the ACP task for each aspect category. Our model is trained whole categories for each domain to explore the correlation information between them. The experimental results on two document-level benchmark datasets show the effectiveness of our architecture than others.

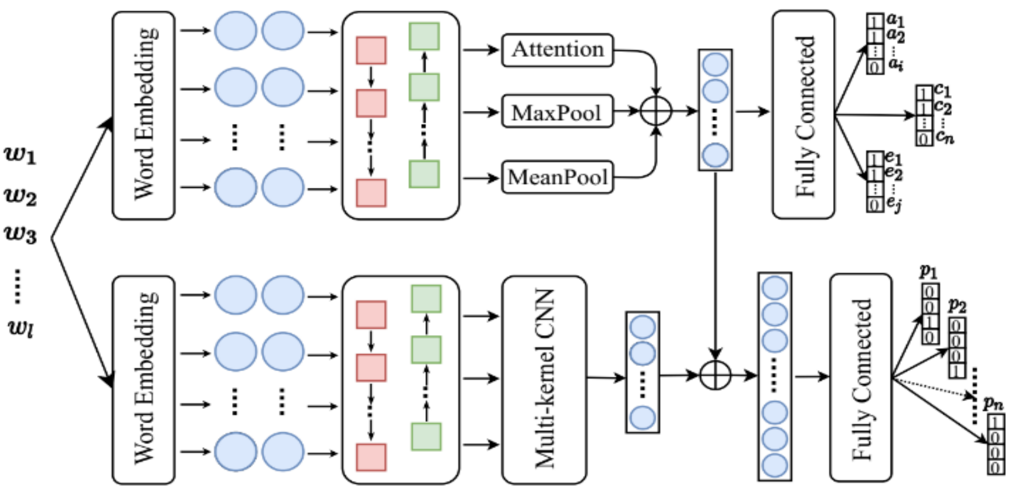

Fig. 1. The overview of our designed architecture.

III. Proposed Architecture

This section describes the details of our proposed architecture for two tasks of the ABSA problem in the Vietnamese language. We define two tasks in this paper as follows: Given an input text s with the length L, the objective of these tasks is to identify the n combination of j "Entity" and i "Attribute" or "Entity" for each domain. Each pair of n combination E#A mentioned in the commentary will be analyzed for sentiment polarity (simply called as sentiment) according to 3 levels: positive, neutral, and negative for the VLSP 2018 datasets or 5 levels: very positive, positive, neutral, negative, very negative for the UIT_ABSA 2019 dataset. Fig. 1 shows an overview of our neural model. At first, the model consists of two components corresponding to two tasks to be solved in this paper: Aspect Category Detection (ACD) and Aspect Category with Polarity (ACP). We now describe each component in our architecture in detail.

A. Component 1: Aspect Category Detection

This component aims to learn the features to predict the entities, the attributes, and aspect category (a combination of entities and attributes) of the input. The text input consists of the sequence of L words $w_1, w_2, \ldots w_L$ with L is the length of the input after padding. Every word is represented as a d-dimensional word vector retrieved from the pre-trained embedding. The embedding layer E with L × D dimension is generated by concatenating all word vectors. Then, we utilize a bidirectional Gated Recurrent Unit (GRU) [31] to encode both forward and backward sequences from the embedding layer. The Bi-GRU layer produces the concatenated output vectors H = $\{h_0, h_1, .., h_t\}$ where h_i is the combination of the i^{th} backward and forward hidden states. The attention is then applied to the output of Bi-GRU layer to capture the important words in the text. The word attention is designed to extract important words because there are many words in the document-level review. However, not all words can represent the aspect category meaning, therefore, we need an attention layer to extract the informative words. Our attention layer is implemented follow by Yang *et al.* [32]. In addition, we concatenate both the max pooling and mean pooling representation with attention vector to produce the input representation: $Reprensentation_{comp1} = Attentio(H) \oplus MaxPooling(H) \oplus MeanPooling(H)$. This representation is then fed into a fully connected layer before predicting the Entity, the Attribute and the Entity#Attribute pairs as in Fig. 1. A fully connected layer is composed of a hidden layer and three sigmoid output layers. The output layer's size is the number of the entity, attribute, and aspect category classes for each domain. For the UIT_ABSA dataset [25], we only use the Entity as the output of this component because their dataset treated the Entity as the aspect category.

$$fc_1 = D(Reprensentation_{comp1}) \tag{1}$$

$$attribute = sig(fc_1) \tag{2}$$

$$entity = sig(fc_1) \tag{3}$$

$$aspect\_category = sig(fc_1) \tag{4}$$

The representation $Reprensentation_{comp1}$ of input review is feed into a fully connected layer for classification. The output of this component includes three outputs corresponding to the entities, attributes and aspect categories (Entity#Attribute). The main output is the aspect categories, however, we also consider the entity and attribute as the output for the regularization effects. Therefore, we have three output layers where each output is a fully connected layer with sigmoid activation. The size of the output is the number of entities, attributes and aspect categories.

B. Component 2: Aspect Category with Polarity

Similar to the Component 1, this component aims to predict the sentiment polarity of aspects. The matrix-vector of the input with L words $w_1, w_2, \ldots w_L$ extracted from the embedding layer is fed into the Bi-GRU layer to extract the context semantic information in the text. The output of i^{th} word is shown as the following equation:

$$h_i = [\overrightarrow{h}_i \oplus \overleftarrow{h}_i] \tag{5}$$

Here, the concatenation is used to combine the forward and backward pass hidden states to obtain the comprehensive information of input. This layer produces the context input representation $H = [h_1, \ldots, h_L]$, H ϵ $\mathbb{R}^{L \times D}$. Next, the convolution operation is applied to the input representation H generated from the Bi-GRU layer. Following the Kim [33], we use the CNN architecture with three convolution layers on the Bi-GRU output to extract the higher-level features. This paper uses three parallel convolution layers with 3,4,5 kernel size to capture the n-gram feature from the input representation H. We can obtain a filter F with the w word window to produce a feature map vector $fm \epsilon R^{k \times (|L|-w+1)}$ where each element is calculated as follows:

$$fm_i = f\ (H_{i:i+n-1} \cdot F + b_0) \tag{6}$$

where f is a non-linear activation function, $b0$ is a bias value. Each convolution layer uses the same k filters generate a matrix of feature map M ϵ $R^{k \times (|L|-w+1)}$. Three convolution operations produce three feature map matrices M_1, M_2, M_3 corresponding to 3, 4, 5 kernel size, respectively. We then apply a max-pooling operation and global pooling operation over each matrix to capture important features of particular filter. Finally, we obtain three vector v_1, v_2, v_3, where v_i ϵ R^k corresponding to M_1, M_2, M_3 feature map matrix, respectively. Then, we concatenate three vectors together as a final vector.

Because the output of Component 2 is designed as a 4-dimension vector (not-mentioned, positive, neutral, negative) or 6-dimension vector (not-mentioned, very positive, positive, neutral, negative, very negative) for each aspect category, we concatenated this vector with Component 1 for the aspect category detection task to provide the final representation of this component. It helps this component have more useful information from the aspect category detection task to predict each sentiment polarity of aspect.

$$v = v_1 \oplus v_2 \oplus v_3 \tag{7}$$

$$Representation_{comp2} = v \oplus Representation_{comp1} \tag{8}$$

The vector $Representation_{comp2}$ is put into a fully connected layer and n softmax layers, where n is the number of aspect categories for each texture domain.

$$fc_2 = D(Representation_{comp2}) \tag{9}$$

$$o_i = sof(fc_2) \tag{10}$$

where fc_2 is the output of a fully connected layer and o_i ϵ R^4 or o_i ϵ R^6 is the output of the i^{th} category for the VLSP 2018 and UIT_ABSA 2019 dataset, respectively.

C. Training Objective

As shown in Fig. 1, our architecture has three output layers for the aspect category detection task (the above half of the Fig. 1) and n output layers for detection of aspects and the classification of their polarity (the below half of the Fig. 1), where n is the number of aspect categories for each domain. For Component 1, the prediction outputs of these two datasets are multi-labels, therefore we use the Binary Cross-Entropy loss function for each output layer. This function and the loss function of Component 1 can be formalized as follows:

$$(\Theta) = -\sum_{i=1}^{N}(y_i.\log(\hat{y}_i)) + (1-y_i)\, log(1-\hat{y}_i)) \quad (11)$$

$$J_{component1}(\Theta) = L_{aspect}(\Theta) + L_{entity}(\Theta) + L_{attribute}(\Theta) \quad (12)$$

where N is the output size, $\hat{y}_i$ is the i^{th} value in the model prediction, y_i is the corresponding target value. The sigmoid activation is used in the output layer. The L_{aspect} is the Binary Cross Entropy loss on the aspect category labels. The L_{entity} and $L_{attribute}$ are the loss function of the entity and attribute labels.

For the Component 2, the loss function is the sum of categorical cross-entropy loss in each aspect category. The loss function of Component 2 is given as follows:

$$J_{component2}(\Theta) = -\sum_{a \in T}\sum_{i=1}^{p} y_i^a.\log(\hat{y}_i^a) \quad (13)$$

where T is the number of aspect category for each domain, $y_i^a \epsilon\ R^p$ is the one-hot vector corresponding to sentiment classes, p is the number of polarities + 1. Finally, we define the loss function of our architecture as follows:

$$(\Theta) = \alpha \cdot J_{component1}(\Theta) + \beta \cdot J_{component2}(\Theta) \quad (14)$$

where α and β are the coefficients of weighting schemes to optimize the objective functions. The Component 2 is the primary output while the Component 1 is the auxiliary output; therefore, we set α = 0.1 and β = 1 after using the grid search technique.

IV. Experimental Results

A. Datasets

In this paper, we use two standard document-level ABSA datasets in Vietnamese: VLSP 2018 [8] and UIT_ABSA 2019 [25]. The VLSP 2018 [8] datasets are annotated for the restaurant and hotel domain. For the restaurant domain, there are 12 different aspect categories (Entity#Attribute) where each category is assigned one of three sentiments: "positive", "negative", and "neutral". There are 34 different aspect categories for the hotel domain. In contrast, the UIT_ABSA 2019 dataset [25] is only annotated for the restaurant domain with 7 aspect categories and 5 sentiment polarities. Here is a brief summarization of the two datasets used in this paper.

- **VLSP 2018**: This dataset was developed by the Nguyen *et al*. [8] research team to organize the ABSA shared-task for the Vietnamese language in 2018. This dataset is built at the document-level data and consists of two datasets where each dataset includes 4 100 reviews divided into the training, validation, and testing set. As the Table I and Table II show the number of reviews, the number of vocabularies, etc. The average lengths of review vary across in three sets. For example, this value is 54 tokens per review for the training set in the restaurant domain, while this value is 163 tokens per a review for the testing set. We can also see the difference between the number of aspect categories per a review, which is usually higher in the testing set. Besides, this dataset is annotated at the document-level; therefore, the number of aspect categories with different sentiment polarities often appears in the samples. For those reasons, this is a challenging dataset for the Vietnamese language.
- **UIT_ABSA 2019**: Similarly, this dataset is directly crawled from Foody[1] website for restaurant domain. It is the first official dataset with 5 sentiment polarities for the ABSA problem in Vietnamese. There are seven aspect categories where each category is assigned one of five polarity levels: "very positive", "positive", "neutral", "negative", and "very negative". Unlike VLSP 2018, this dataset includes only 7 different aspect categories, which are not a combination of attributes and entities. In total, this dataset includes 7 828 reviews divided into three sets with the ratio of 7/1/2. The authors used the iterative stratification[2] technique to the split overall dataset to balance the number of aspect categories and corresponding sentiment polarities in three sets. However, there is still an imbalance between the aspect categories and the different sentiment polarities in the training, validation, and testing set. For example, the "Quality" and "Service" are the most annotated aspect categories compared with others in the entire dataset. For the sentiment labels, the "very positive" sentiment is most annotated for all aspects with a percentage of 50%.

TABLE I: The General Information of the VLSP 2018 [8] Dataset for the Restaurant Domain

Information	Train	Validation	Test
N.o Reviews	2 961	1 290	500
N.o Tokens	147 706	59 102	76 755
N.o Vocabularies	5 168	3 398	3 375
N.o Aspect	9 034	3 408	2 419
Ave. Aspects per Review	3.05	2.64	4.38
Average Length	54	50	163

TABLE II: The General Information of the VLSP 2018 [8] Dataset for the Hotel Domain

Information	Train	Validation	Test
N.o Reviews	3 000	600	500
N.o Tokens	105 919	11 907	30 046
N.o Vocabularies	3 908	2 745	1 631
N.o Aspect	13 948	7 111	2 584
Ave. Aspects per Review	4.65	11.85	5.17
Average Length	47	23	30

TABLE III: The General Information of the UIT_ABSA 2019 [25] Dataset

Information	Train	Validation	Test
N.o Reviews	5 479	776	1 573
N.o Tokens	287 660	81 198	161 804
N.o Vocabularies	8 982	3 464	4 852
N.o Aspect	17 227	2 458	4 968
Ave. Aspects per Review	3.14	3.18	3.16
Average Length	52.5	52.3	51.4

A common point of the two datasets is comprising users' comments crawled directly on the famous websites about the restaurant and hotel. Hence, there are still many grammatical errors, sentence structure, teen code, and icons in these

[1] https://www.foody.vn/
[2] https://github.com/trent-b/iterative-stratification

datasets. The two datasets also reflect the ABSA dataset challenges of a very limited amount of annotated data and high-class imbalances.

B. Baselines

In this section, we present the various architectures as baseline methods which was compared with our framework.

- SVM + handcraft features [25] [27]: A traditional SVM model with various handcraft features (n-grams, POS, words, etc.).
- DCNN [28]: A very deep Convolutional Neural Network with various kernel sizes for multi-label classification.
- BiLSTM-CNN [30]: A combination framework with the BiLSTM and CNN model. We use the BiLSTM to encode the document representation, then put them to CNN model. The output layer consists of the number of softmax layers corresponding to each aspect categories.
- HAN [32]: A hierarchical attention networks with the number of softmax layer on the final states for aspect categories. Because the reviews contain many errors in the text structure, therefore it is difficult to apply the sentence segmentation technique. Hence, we split the reviews as a combination of phrases as the input of Sentence Encoder component.

C. Experimental Setup

Model Configuration: In the following, we describe the model hyper-parameters during our experiments. To initialize the word embedding, we have trained a Skip-gram algorithm with 100-dimensions on the domain corpus. Specifically, our domain corpus includes $\tilde{2}27K$ and $\tilde{3}00K$ sentences corresponding to the hotel and restaurant domain, respectively. We used Gensim Library [34] to train our domain embeddings.

Because each Vietnamese word can be made up of one or more syllables, it is necessary to segment the text input into words. However, Vietnamese word segmentation tools are currently trained on the news corpora; it is very challenging to segment words on users' comments, which contained many abbreviations or wrong grammatical. It will lead to many words that cannot be segmented in the right way and do not exist in pre-trained word embedding. To solve this problem, we used a technique by taking advantage of the largest prefix and suffix of segmented words. The average value is presented as the word vector. This technique reduces the number of Out of Vocabulary (OOV) words in the training set. In our architecture, each GRU layer's hidden unit was set to 256 and we employed 3 different kernel sizes (3,4,5), and the number of filters were set to 300 for the CNN architecture. The first fully connected layer was also set to 300 neurons. We trained our model for 100 epochs with a batch size of 50. For optimization algorithm, we used RAdam [35] optimizer with default settings. We also experimented with other optimization algorithms such as Adam and Stochastic Gradient Descent (SGD); however, these optimization algorithms were not effective for our architecture. We used dropout operation for regularization with the rate of 0.5 for the word embeddings and the fully connected layer as the default parameter. To processing the text inputs, we conducted a series of pre-processing steps before putting them to our model as previous works [30] to show the effectiveness of our architecture. All models will be run 5 times with different random seeds and reported as the average scores.

TABLE IV: RESULTS OF OUR PROPOSED FRAMEWORK IN COMPARISON WITH PREVIOUS METHODS ON THE VLSP 2018 DATASETS [8] FOR THE RESTAURANT DOMAIN

Task	Methods	Precision	Recall	F1-score
Aspect Category Detection	SVM [27]	79.00	76.00	77.00
	DCNN [28]	84.75	76.48	80.40
	HAN [32]	73.21	80.94	76.88
	BiLSTM-CNN [30]	82.02	77.51	79.70
	BERT [15]	79.16	83.67	81.35
	Ours	83.43	81.23	**82.29**
Aspect Category with Polarity	SVM [27]	62.00	60.00	61.00
	DCNN [28]	-	-	-
	HAN [32]	58.25	64.40	61.17
	BiLSTM-CNN [30]	66.66	63.00	64.78
	BERT [15]	63.27	65.03	64.14
	Ours	68.91	67.08	**67.96**

V. RESULT AND ANALYSIS

A. Result

In this section, we compare our experimental results to previous approaches on two different datasets: VLSP 2018 [8] and UIT _ABSA 2019 [25]. It is noted that the Aspect Category Detection (ACD) task is to identify the aspect category (e.g., Service#General, Food#Quality) mentioned in user comments, while the Aspect Category with Polarity (ACP) task is computed based on the pairs of Entity#Attribute and corresponding polarity. Therefore, the effectiveness of ACP task depends on precisely identifying the aspect categories mentioned in a given review. In this paper, we consider the AS task as main task based on the results of previous studies. As follows, we show the experimental results on two tasks on both datasets. To evaluate our architecture's performance compared with other methods, we use Precision, Recall, and F1-score for each task.

For the VLSP 2018 datasets, our architecture achieves results superior to the best previous methods on both tasks for two domains. For the restaurant domain, our result is higher than others, specifically, +5.29%, +1,89%, +5.41% ,+3.59%, and +1.14% of F1-score for the SVM [27], DCNN [28], HAN [32], BiLSTM-CNN [30] and BERT [15] on the ACD task, respectively. For the main task - AS, our model achieves the rise in F1-score compared to other methods, in detail, it gives an improved performance as compared to the SVM [27] with +6.96%, BiLSTM-CNN [30] with +3.18%, HAN [32] with +6.75% and BERT [15] +3.82%, respectively. For the hotel domain of VLSP 2018, our architecture is also higher than other approach except BERT approach [15] on the ACD task. We outperform SVM [27], HAN [32] and DCNN [28] with +9.66%, +9.41% and +10.16% of F1-score. Compared with BiLSTM-CNN [30] and BERT [15] approach, our model is +0.81% higher and -1.0% lower of F1-score. On the contrary, for the ACP task, our proposed model outperforms previous methods (SVM, HAN, BiLSTM-CNN, BERT) with +12.16%, +14.04%, +2.26%, +5.42% of F1-score, respectively. For the UIT_ABSA 2019 dataset, our architecture also outperforms the previous methods on both two tasks. Our experimental results are higher than the baseline of [25] with +5.95%, DCNN [28] with +0.21%, HAN [32] with +8.26% and BiLSTM-CNN [30] with 1.14% of F1-score on the ACD task. For the main task, our model improves the performance by +9.93% of F1-score than the SVM [25], +14.04% of F1-score than HAN [32] model and +1.84% of F1-score than the BiLSTM-CNN [30]. The experiments indicate that our method generates state-of-the-art results for two tasks on both the VLSP 2018 and UIT_ABSA 2019 datasets.

Next, we show the results of our architecture for the whole aspect categories in both datasets. Fig. 2 and Fig. 3 present the F1-score of each aspect category for two domains of VLSP 2018 and UIT_ABSA 2019 dataset, respectively. For the restaurant domain, we can see that there are many aspect categories are achieved with the high score, such as "Food#Style&Options", "Food#Quality", and "Food#Prices". However, there are still a few aspect categories with a low score, such as "Restaurant#Miscellaneous" and "Restaurant#Prices" because these aspect categories are less annotated in both the training and testing set. For the hotel domain, we show the 12/34 aspect categories results on the testing set. However, there are many aspect categories which can not predict, for example, "Room_Amenities # Prices" (0 samples), "Room_Amenities # Miscellanous" (1 sample), "Rooms# Miscellanous" (3 samples), "Food&Drink_Miscellanous" (13 samples) and "Facilities#Miscellanous" (33 samples).

TABLE V: RESULTS OF OUR PROPOSED FRAMEWORK IN COMPARISON WITH PREVIOUS METHODS ON THE VLSP 2018 DATASETS [8] FOR THE HOTEL DOMAIN

Task	Methods	Precision	Recall	F1-score
Aspect Category Detection	SVM [27]	75.00	64.00	69.00
	DCNN [28]	82.35	59.75	69.25
	HAN [32]	76.69	63.54	69.50
	BiLSTM-CNN [30]	84.03	72.52	77.85
	BERT [15]	79.60	79.72	**79.66**
	Ours	87.38	71.52	78.66
Aspect Category with Polarity	SVM [27]	66.00	57.00	61.00
	DCNN [28]	-	-	-
	HAN [32]	68.24	56.54	61.84
	BiLSTM-CNN [30]	76.53	66.04	70.90
	BERT [15]	79.83	58.82	67.74
	Ours	81.28	66.52	**73.16**

Fig. 2. The chart shows our results of whole aspect categories on two VLSP 2018 datasets. The left chart is the F1-score of 12 aspect categories for the restaurant domain and the right chart is the F1-score of 12/34 highest aspect categories for the hotel domain.

TABLE VI: RESULTS OF OUR PROPOSED MODEL IN COMPARISON WITH PREVIOUS METHODS ON THE UIT_ABSA 2019 DATASETS [25]

Task	Methods	Precision	Recall	F1-score
Aspect Category Detection	SVM [25]	92.02	82.73	87.13
	DCNN [28]	92.04	93.71	92.87
	HAN [32]	86.31	83.37	84.82
	BiLSTM-CNN [30]	94.19	89.90	91.94
	Ours	93.10	93.05	**93.08**
Aspect Category with Polarity	SVM [25]	62.52	56.21	59.20
	DCNN [28]	-	-	-
	HAN [32]	56.06	54.15	55.09
	BiLSTM-CNN [30]	68.93	65.72	67.29
	Ours	69.15	69.11	**69.13**

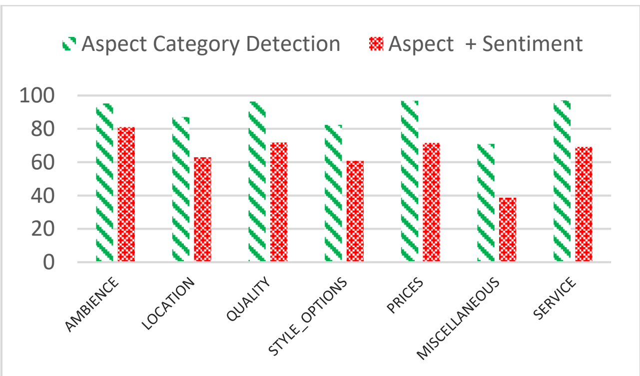

Fig. 3. The chart shows our results of whole aspect categories of the UIT_ABSA 2019 dataset.

As shown in Fig. 3, we observe that our model can achieve satisfying results for categories in terms of F1 in the ACD task except for "Miscellaneous". The results of the ACP task are relatively low for all aspect categories. One of the reasons for poor performance of our architecture is the imbalance between the polarities of each aspect. Another reason might because the polarity sentiments in this dataset are assigned in 5 levels, which is more complicated than 3 level sentiments, may yield worse results for the ACP task. In conclusion, these results demonstrate that our architecture can utilize an aspect category information of ACD task to generate the final representation for aspect category with sentiment prediction.

B. *Analysis*

B.1 Sensitivity Analysis. This section shows the effects of two important parameters in our architecture, the dropout rate and the word embedding size. The other parameters are held as the default when we analyze the effect of one parameter.

+ **Dropout rate**: Our architecture is a type of ensembles of neural networks with different model configurations; therefore, it is easy to overfit in trying to achieve good accuracy with small size dataset. Dropout is one of the regularization methods to reduce overfitting in deep neural networks effectively. Fig. 4 and Fig. 5 show the F1-score of different dropout values on the validation set corresponding to the hotel and restaurant, respectively. It can be found that the dropout rate has significant impacts on the performance of the model. As shown in the three charts (Fig. 4 and Fig. 5), the dropout value in the range of 0.5 and 0.7 will help our model achieved the best performance on the validation set. It proves that the value of dropout significantly affects the performance.

+ **Embedding dimension**: In this paper, we trained two-word embeddings on domain corpus, respectively. We conduct a comparative experiment to discover the effects of different dimensions of pre-trained word embedding. We choose the vector size as follow: dim ϵ {50, 100, 150, 200, 300, 400, 500}. As the results are shown in Fig. 6, the embedding dimension affects both data domains differently - high dimension helps our model achieves the better results for the hotel domain, but the difference between these dimension values is not significant. As shown in Fig. 7, the best result is achieved with the embedding dimensions with value of 100 on the UIT_ABSA 2019 dataset.

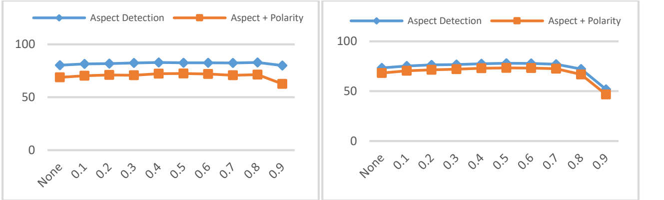

Fig. 4. The chart shows the effect of dropout probability with F1-score on the VLSP 2018. The left and right chart corresponding to the restaurant and hotel domain, respectively.

TABLE VII: TWO ACTUAL EXAMPLES AND OUR PREDICTIONS IN THE VALIDATION SET OF THE VLSP 2018 DATASET

Review	Gold Labels	Predicted Labels
Khu này là khu chợ nên có rất nhìu đồ ăn khá ok, đặc biệt trong đó là quán chè ngay phía đầu đường sau khi vừa vượt qua nhã 3. Cô bán nhìu loại chè, và đồng giá 10k nhé. Vừa vô gặp cái này là nhìn ghiền liền, tấp vô ngồi ăn ngay và luôn nhé. Chè bánh lọt 10k nè. Bánh bên trong nè, nhìn hấp dẫn chưa. Bánh lọt k nè. Ở đây có cả đá me nữa nha. Đậu nành 4k/bịch, chè 10k bịch đem về nè. Có cả bánh cam 4k/cái nữa nè. Bánh bò cũng 4k nhé. Há cảo 18k hộp 5 cái. Giá thành chợ thì khá ok rồi. Ăn cũng ngon, chè rẽ mà ăn thích lắm luôn ấy. Mỏi lần đợi vk đi chợ là phải làm ở đây vài ly mới chịu dc. (This area is a marketplace so there is a lot of OK food, especially the tea shop, which is right at the beginning of the road after crossing "three-way crossroads". She sells many types of tea with the same price - 10k. When I come to, I immediately feel attracted and rushed in to eat. 10k for cake sweet soup. Cake inside, look attractive yet. The Vietnamese cendol is here. There is also tamarind ice. Soybeans with 4k a bag and tea bags with 10k for one can buy and take away. There is also a 4k / orange cake. The sponge cake is 4k. Dumplings is 18k for a box of 5. The price at this market is quite ok. The food was delicious, the tea was small but the taste was yummy. "Every time" I wait for my wife to go to the market, have to eat a few things at this store.)	+Location#General, ***neutral*** +Food#Style&Option, positive +Food#Prices, ***positive*** +Food#Quality, positive +Drinks# Prices, positive +Restaurant#General, positive	+Location#General, ***positive*** +Food#Style&Option, positive +Food#Prices, ***neutral*** +Food#Quality, positive +***Service#General, negative*** +Restaurant#General, positive
Ksan mới, dễ tìm, gần biển, đi bộ chỉ mất vài phút. Nhân viên nhiệt tình. Bữa sáng bình thường, dao cắt bánh mì ko sạch, cắt miếng bánh mì đen thui. Dra giường ko sạch, phòng có vài chú muỗi, vòi sen trong phòng tắm bị chảy nước làm gia đình tôi phải yêu cầu đổi phòng. Dù sao nhân viên ksan cũng nhiệt tình và nhanh chóng đổi phòng khác cho tôi. (New hotel, easy to find, close to the sea, it only takes a few minutes to walk. Enthusiastic staff. Normal breakfast, unclean bread knife, slicing black bread. The bed is not clean and the room has a few mosquitoes. The shower in the bathroom is also watery, so my family have to request to change rooms. Anyway, the hotel staff are enthusiastic and quickly change the room for me.)	+Hotel#Design&Feature, positive +Location#General, positive +Service#General, positive +Food&Drink#Quality, neutral +Food&Drink#Style&Option, negative +Room_Amenities#Cleanliness, nega-tive +Rooms#Cleanliness, negative +Room_Ameninites#Design& Features, negative	+Hotel#Design&Feature, positive +Location#General, positive +Service#General, positive +Food&Drinks #Quality, neutral +Room_Amenities# Cleanlines, negative +Rooms#Cleanlines, negative +***Room_Amenities# Quality, negative***

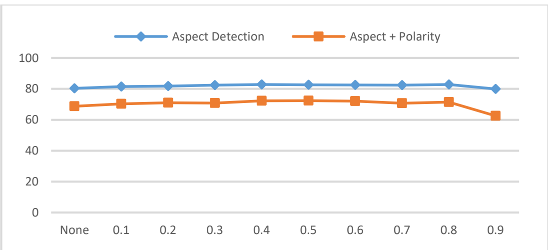

Fig. 5. The chart shows the effect of dropout probability with F1-score on the UIT_ABSA 2019.

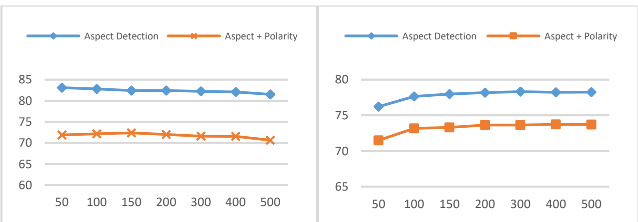

Fig. 6. The chart shows the F1-score of embedding dimensions trained on domain corpus for VLSP 2018 dataset. The left and right chart corresponding to the restaurant and hotel domain, respectively.

B.2 Case study. In this part, we make a case study on the results of the testing set of two datasets. Table VII shows two examples with ground-truth and our prediction labels for the VLSP 2018 datasets. As for the first example in Table VII, our model predicts the number of aspect categories nearly correct for the review, except the "Service#General" category. Because the user mentions a sentence "Mỏi lần đợi vk đi chợ là phải làm ở đây vài ly mới chịu dc" ("Every time" I wait for my wife to go to the market, I have to eat a few things at this store.) leads to wrong this category in our prediction. However, this sentence, describes the writer's supplementary information rather than referring to the "Service#General" category. We verify this error by removing this sentence in the review, and our model does not predict this category for the review. Compared to the sentiment of aspect category, our predictions are different from the ground-truth label in the "Location#General" and "Food#Prices" categories. The ground-truth of sentiment of "Location#General" is "neutral" while our prediction is "positive" because our model notes that the phrase "quán chè ngay phía đầu đường sau khi vừa vượt qua nhã 3"(the tea shop, which is right at the beginning of the road after crossing "three-way crossroads") means the "positive" for this category.

Therefore, we examined this example in the guideline of this dataset, and in the training and validation sets, we discovered ambiguous cases of sentiment annotation for this category. This leads to an incorrect prediction of our model.

In contrast, our model assigned the “neutral” sentiment to “Food#Prices”. In terms of the guideline, “if the user only mentions specific prices and does not express the opinion for the item will be assigned “neutral” for the category. Based on this example, we can see that the user mentioned many specific prices of foods such as “Đậu nành 4k/bịch, chè 10k bịch đem về nè” (Soybeans with 4k a bag and tea bags with 10k for one can buy and take away), “bánh cam 4k/cái”(There is also a 4k / orangecake), etc. Therefore, our model predicts the “neutral” sentiment for “Food#Prices” category. After listing the food names with its prices, however, the user concludes that the price of food is pretty okay in the phrase “Giá thành chợ thì khá ok rồi” (The price at this market is quite ok). Besides, there is also a bunch of spelling mistakes. For the words we put in quotation marks, it means we did translate it into the correct one. In the first example, the word "nhã3" should be "ngã3" instead.

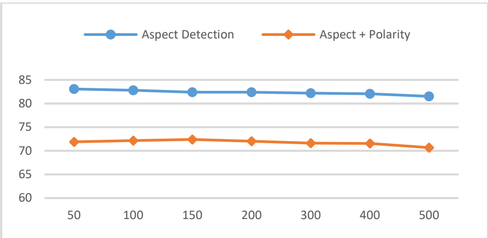

Fig. 7. The chart shows the F1-score of embedding dimensions trained on domain corpus for UIT_ABSA 2019 dataset.

As for the second example, our model is unable to find the “Food&Drink#Style&Option” because the phrase “cắt miếng bánh mì đen thui”(Cutting pieces of bread and make it black) - this is a type of implicit aspect category; and “Room_Ameninites#Design&Features” is wrong predict as “Room_Ameninites#Quality”. For the sentiment prediction of each aspect category, our model can capture the correct sentiments corresponding to each detected category from the review. In this paper, we focus on predict the pairs of aspect category and corresponding sentiment, therefore, the overall scores of our architecture depend on the ACD task.

VI. Conclusion and Future Work

This paper presented an effective joint multi-task architecture based on neural networks for the Aspect Category Detection and Aspect Category with Polarity Classification tasks for Vietnamese document-level ABSA datasets. Our model jointly trains the aspect category detection and corresponding polarity tasks simultaneously and combines the information feature of the ACD task with the ACP task for the final representation. Our model can predict the whole aspect categories with corresponding sentiment polarities for each domain. We conducted various experiments and compared it against several previous methods to show the effectiveness of our model. Experimental results on two benchmark document-level datasets demonstrated that our model has good performance for the document-level input. Up to now, our model established the new state-of-the-art results for the VLSP 2018 and UIT_ABSA 2019 datasets.

For future works, we intend to explore the new other neural networks to solve this problem in the Vietnamese language. Our framework can be applied to other languages in the document-level datasets for ABSA problem. Besides, we also want to combine domain knowledge and sentiment features to provide additional features for the sentiment polarity task. Moreover, the idea of using rich-resource languages is additional data using translation approaches, and multilingual embeddings can be the potential direction for these tasks. Recently, the Iz Beltagy *et al.*, [36] presented a new model for the long document, however, this model is not available for Vietnamese up to now. For the future works, the combination this architecture and our approach for the document-level for ABSA problem is also a new potential research direction.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

D. V. Thin proposed the main ideas of this research, conducted the experiments and wrote the paper. Lac S. Le and Hao M. Nguyen discuss the ideas and evaluated the experimental results. Ngan L.T Nguyen discussed the ideas and contributed to writing the paper. All authors approved the final version.

References

[1] L. Zhang, S. Wang, and B. Liu, “Deep learning for sentiment analysis: A survey,” *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 8, no. 4, 2018.

[2] R. Ireland and A. Liu, “Application of data analytics for product design: Sentiment analysis of online product reviews,” *CIRP Journal of Manufacturing Science and Technology*, vol. 23, pp. 128–144, 2018.

[3] A. Nazir, Y. Rao, L. W. Wu, and L. Sun, “Issues and challenges of aspect-based sentiment analysis: A comprehensive survey,” *IEEE Transactions on Affective Computing*, 2020.

[4] M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, and S. Manandhar, “SemEval-2014 Task 4: Aspect Based sentiment analysis,” in *Proc. the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, Association for Computational Linguistics, Dublin, Ireland, 2014, pp. 27–35.

[5] M. Pontiki, D. Galanis, H. Papageorgiou, S. Manandhar, and I. Androutsopoulos, “Semeval-2015 task 12: Aspect based sentiment analysis,” in *Proc. the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, 2015, pp. 486–495.

[6] M. Pontiki, D. Galanis, H. Papageorgiou *et al.*, “Semeval-2016 task 5: Aspect based sentiment analysis,” in *Proc. the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, 2016, pp. 19–30.

[7] Q. N. Jiang, L. Chen, R. F. Xu, X. Ao, and M. Yang, “A challenge dataset and effective models for aspect-based sentiment analysis,” in *Proc. the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 2019, pp. 6281–6286.

[8] H. Nguyen, H. Nguyen, Q. Ngo, L. Vu, V. Tran, B. Ngo, and C. Le, “VLSP shared task: Sentiment analysis,” *Journal of Computer Science and Cybernetics*, vol. 34, no. 4, pp. 295–310, 2019.

[9] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805, 2018.

[10] M. Hoang, O. A. Bihorac, and J. Rouces, “Aspect-based sentiment analysis using BERT,” in *Proc. the 22nd Nordic Conference on Computational Linguistics*, Link ̈oping University Electronic Press, Turku, Finland, 2019, pp. 187–196.

[11] C. Sun, L. Y. Huang, and X. P. Qiu, “Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence,” in *Proc. NAACL-HLT*, 2019, pp. 380–385.

[12] X. Li, L. D. Bing, W. X. Zhang, and W. Lam, “Exploiting BERT for end-to-end aspect-based sentiment analysis,” in *Proc. the 5th Workshop on Noisy User-generated Text (W-NUT 2019)*, 2019, pp. 34–41.

[13] X. L. Li, X. Y. Fu, G. L. Xu, Y. Yang, J. N. Wang, L. Jin, Q. Liu, and T. Y. Xiang, "Enhancing BERT representation with context-aware embedding for aspect-based sentiment analysis," *IEEE Access*, vol. 8, pp. 46868–46876, 2020.

[14] C. Sun, X. P. Qiu, Y. G. Xu, and X. J. Huang, "How to fine-tune bert for text classification?" in *Proc. China National Conference on Chinese Computational Linguistics*, Springer, 2019, pp. 194–206.

[15] N. C. Le, N. T. Lam, S. H. Nguyen, and D. T. Nguyen, "On Vietnamese sentiment analysis: A transfer learning method," in *Proc. 2020 RIVF International Conference on Computing and Communication Technologies (RIVF)*, 2020, pp. 1–5.

[16] K. Schouten and F. Frasincar, "Survey on aspect-level sentiment analysis," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 3, pp. 813–830, 2015.

[17] H. H. Do, P. W. C. Prasad, A. Maag, and A. Alsadoon, "Deep learning for aspect-based sentiment analysis: A comparative review," *Expert Systems with Applications*, vol. 118, pp. 272–299, 2019.

[18] W. Xue, W. B. Zhou, T. Li, and Q. Wang, "MTNA: A neural multi-task model for aspect category classification and aspect term extraction on restaurant reviews," in *Proc. the Eighth International Joint Conference on Natural Language Processing*, 2017, pp. 151–156.

[19] F. X. Wang, M. Lan, and W. T. Wang, "Towards a one-stop solution to both aspect extraction and sentiment analysis tasks with neural multi-task learning," in *Proc. 2018 International Joint Conference on Neural Networks*, IEEE, 2018, pp. 1–8.

[20] M. Schmitt, S. Steinheber, K. Schreiber, and B. Roth, "Joint aspect and polarity classification for aspect-based sentiment analysis with end-to-end neural networks," in *Proc. the 2018 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Lin guistics, Brussels, Belgium, pp. 1109–1114.

[21] Z. H. Dai, W. Dai, Z. H. Liu *et al*., "Multi-task multi-head attention memory network for fine-grained sentiment analysis," in *Proc. CCF International Conference on Natural Language Processing and Chinese Computing*, Springer, 2019, pp. 609–620.

[22] R. D. He, W. S. Lee, H. T. Ng, and D. Dahlmeier, "An Interactive Multi-Task Learning Network for End-to-End aspect-based sentiment analysis," in *Proc. the 57th Annual Meeting of the Association for Computational Linguistics*, 2019, pp. 504–515.

[23] Y. L. Liang, F. D. Meng, J. C. Zhang, J. N. Xu, Y. F. Chen, and J. Zhou, "An iterative knowledge transfer network with routing for aspect-based sentiment analysis," arXiv preprint arXiv:2004.01935, 2020.

[24] N. T. T. Thuy, N. X. Bach, and T. M. Phuong, "Cross-language aspect extraction for opinion mining," in *Proc. 2018 10th International Conference on Knowledge and Systems Engineering (KSE)*, 2018, pp. 67–72.

[25] M.-H. Nguyen, T. M. Nguyen, D. V. Thin, and N. L.-T. Nguyen, "A corpus for aspect-based sentiment analysis in Vietnamese," in *Proc. 2019 11th International Conference on Knowledge and Systems Engineering (KSE)*, IEEE, 2019, pp. 1–5.

[26] N. T. T. Thuy, N. X. Bach, and T. M. Phuong, "Leveraging foreign language labeled data for aspect-based opinion mining," in *Proc. 2020 RIVF International Conference on Computing and Communication Technologies (RIVF)*, 2020, pp. 1–6.

[27] D. V. Thin, V. Nguyen, N. Kiet, and N. Ngan, "A transformation method for aspect-based sentiment analysis," *Journal of Computer Science and Cybernetics*, vol. 34, no. 4, pp. 323–333, 2019.

[28] D. V. Thin, V. D. Nguyen, K. V. Nguyen, and N. L. Nguyen, "Deep learning for aspect detection on Vietnamese reviews," in *Proc. 2018 5th NAFOSTED Conference on Information and Computer Science (NICS)*, 2018, pp. 104–109.

[29] T. K. Tran and T. T. Phan, "Towards a sentiment analysis model based on semantic relation analysis," *International Journal of Synthetic Emotions (IJSE)*, vol. 9, no. 2, pp. 54–75, 2018.

[30] D. V. Thin, K. V. Nguyen, D.-V. Nguyen, and N. L.-T. Nguyen, "Multi-task learning for aspect and polarity recognition on vietnamese datasets," in *Proc. The 16th International Conference of the Pacific Association for Computational Linguistics*, 2019.

[31] K. Cho, B. van Merri ̈enboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder–decoder for statistical machine translation," in *Proc. the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 1724–1734.

[32] Z. C. Yang, D. Y. Yang, C. Dyer, X. D. He, A. Smola, and E. Hovy, "Hierarchical attention networks for document classification," in *Proc. the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2016, pp. 1480–1489.

[33] Y. Kim, "Convolutional Neural networks for sentence classification," in *Proc. the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Association for Computational Linguistics, Doha, Qatar, 2014, pp. 1746–1751.

[34] R. Reh˚u˘rek and P. Sojka, "Software framework for topic modelling with large corpora," in *Proc. the LREC 2010 Workshop on New Challenges for NLP Frameworks*, ELRA, Valletta, Malta, 2010, pp. 45–50.

[35] L. Y. Liu, H. M. Jiang, P. C. He *et al*., "On the variance of the adaptive learning rate and beyond," arXiv preprint arXiv: 1908.03265, 2019.

[36] B. Iz, M. E. Peters, and A. Cohan, "Longformer: The long-document transformer," arXiv preprint arXiv: 2004.05150, 2020.

Dang Van Thin was born in Nghe An Province, Viet Nam in 1995. He graduated with the B.S. in computer science at the University of Information Technology – Vietnam National University Ho Chi Minh city, Vietnam in 2017. He has currently got the master degree at the University of Information Technology – VNU HCM. His research interests are about natural language processing, machine learning.

Lac Si Le was born in An Giang province, Viet Nam in 1999. Currently, he is a junior student of the Faculty of Software Engineering at the University of Information Technology – Vietnam National University Ho Chi Minh city, Vietnam in 2020. His research interests include natural language processing, neural network and data mining.

Hao Minh Nguyen was born in Kien Giang province, Viet Nam in 1985. He graduated with the B.S. from Posts and Telecommunications Institute of Technology, Vietnam in 2015. He received a master degree at the University of Information Technology – Vietnam National University Ho Chi Minh city, Vietnam in 2020. He has currently been a director of a Software Limited Company 102 and visiting lecturers. His research interests are natural language and website technology.

Ngan Luu-Thuy Nguyen is a scientist at the University of Information Technology, Vietnam National University, Ho Chi Minh City, Vietnam. She received her PhD degree in information science and technology from the University of Tokyo, Japan. She was a postdoctoral researcher at the National Institute of Informatics, Japan from 2012 to 2013. Her research interests include natural language processing and data analysis.