# Aspect-based Sentiment Analysis for Vietnamese Reviews about Beauty Product on E-commerce Websites

**Quang-Linh Tran, Phan Thanh Dat Le, Trong-Hop Do**
University of Information Technology, Ho Chi Minh City, Vietnam
Vietnam National University, Ho Chi Minh City, Vietnam
{18520997, 18520570}@gm.uit.edu.vn, hopdt@uit.edu.vn

## Abstract

Millions of reviews are generated on e-commerce platforms and analyzing them deeply brings a lot of useful information for sellers and buyers. This paper deals with the aspect-based sentiment analysis problem to analyze the aspect and polarity of Vietnamese reviews about beauty products on e-commerce websites. The contribution of this paper is three-fold. Firstly, we introduce a dataset containing 16,227 reviews about lipsticks. There are 32,775 pairs of aspects and sentiments in this dataset. Besides, we conduct some baseline experiments using Deep Learning-based models to build detection and classification models for extracting the aspects of reviews and classifying the sentiment of each aspect. In addition, a comprehensive comparison is also performed to see whether single-task learning or multi-task learning is the better approach to predict the aspect and sentiment of reviews. The experimental results show that the BiGRU+Conv1D model in the single-task learning approach outperforms others with the F1-score in the aspect detection task at 98.09% and 91.01% for the sentiment classification task.

## 1 Introduction

Aspect-based sentiment analysis (ABSA) is a task in Natural Language Processing (NLP). Instead of only extracting the polarity of the text in sentiment analysis, ABSA analyzes the sentiment more deeply by showing the polarity of each aspect of the text. In reviews on e-commerce platforms, a review contains more than one aspect so ABSA is more robust to analyze the polarity than sentiment analysis. For example, in lipstick reviews, the customers do not always give general reviews about the products, they tend to review based on the aspect of the lipstick such as price, colour, et cetera. The customers may like the colour, but the texture of the lipstick is so bad that they are disappointed with this. This is a reason why ABSA is really necessary to understand more deeply what customers think about the products.

With the explosion of e-commerce, millions of reviews are generated every day from customers. They leave a lot of useful information for not only the sellers to understand what their customers like and dislike about their products but also for other customers to read and consider previous buyers' experiences before making a purchase. There are many kinds of products in e-commerce, but beauty products are one of the most favorable products to buy online. In addition, the reviews of beauty products, especially about the lipsticks contain a lot of aspects, from price, staying power to colour, thanks to which, several analyses can be conducted to deeply understand the attitude of customers toward the lipstick. This is the reason why we choose reviews about lipsticks to build a dataset to deal with ABSA problem in e-commerce reviews.

There are three main contributions of this paper. Firstly, a novel Vietnamese dataset about reviews of lipsticks in e-commerce platforms for the aspect-based sentiment analysis task is created to handle the problem. To the best of our knowledge, there is no dataset for reviews about beauty products in Vietnamese and the size of this dataset is also bigger than some other Vietnamese datasets for ABSA. Secondly, we propose an effective deep learning model

architecture to detect the aspect and corresponding sentiment of reviews. Finally, a comprehensive comparison between single-task learning and multi-task learning is conducted to find the best approach for the aspect-based sentiment analysis problem.

## 2 Related Works

Aspect-based sentiment analysis has drawn a lot of attention in recent years. Several workshops such as SemEval2014 (Pontiki et al., 2014), SemEval2015 (Pontiki, Galanis, Papageorgiou, Manandhar, & Androutsopoulos, 2015), SemEval2016 (Pontiki et al., 2016), VLSP2018 (H. T. Nguyen et al., 2018) were organized to find the best solution for aspect-based sentiment analysis problem. There are many approaches to resolving this problem. Single task ABSA contains several sub-tasks such as aspect category detection, and aspect sentiment classification (Zhang, Li, Deng, Bing, & Lam, 2022). (Zhou, Wan, & Xiao, 2015) achieved the F1-score of 90.10% in aspect category detection in the SemEval2014 dataset (Pontiki et al., 2014) with representation learning. (Wang, Huang, Zhu, & Zhao, 2016) used attention-based LSTM for aspect sentiment classification. Besides single task ABSA, compound ABSA is also an effective approach. In this approach, aspect category sentiment analysis is the task to extract aspects and the corresponding sentiment simultaneously. (He, Lee, Ng, & Dahlmeier, 2019) proposed an interactive multi-task learning network for extracting aspects and sentiment of documents.

In Vietnamese, there are several datasets about sentiment analysis in many domains. (H. T. Nguyen et al., 2018) published the dataset SA-VLSP2018 for aspect-based sentiment analysis about hotel and restaurant domains in the VLSP workshop. In addition, (K. T.-T. Nguyen et al., 2021) applied span detection for aspect-based sentiment analysis and get the performance at 62.76% F1-score for the dataset UIT-ViSD4SA. (K. V. Nguyen, Nguyen, Nguyen, Truong, & Nguyen, 2018) published the dataset UIT-VSFC, which consists of over 16,000 sentences of feedback from students. This dataset is used for sentiment classification and topic classification.

## 3 Dataset

### 3.1 Task Definition

We built a Vietnamese dataset for the aspect-based sentiment analysis task. This dataset contains 16,227 Vietnamese reviews about 9 lipsticks from Shopee[1]. There are 2 sub-tasks in this dataset: aspect detection and sentiment classification. In the aspect detection sub-task, we focus on finding the aspects mentioned in the review. There are 7 aspects: SMELL, PRICE, SHIPPING, COLOUR, PACKING, TEXTURE, STAYING POWER and 1 aspect indicating spam: OTHERS. Table 1 shows the definition of all aspects. Another sub-task is classifying the polarity of these aspects into: Positive, Neutral or Negative. We split the dataset into three sets: 12,981 reviews for training, 1,623 reviews for validation, and 1,623 reviews for testing. The training and validation will help us to build detection and classification models, and the test set is used to evaluate the performance of the models.

### 3.2 Annotation process

We split our annotation process into 2 main phases, including training phase and labelling phase. There are 5 sub-phases in the training phase, each sub-phase has 200 reviews. We use Cohen's Kappa (Cohen, 1960) score to measure inter-annotator agreement as the metric for calculating the quality of annotation. When the score between annotators in each sub-phase is higher than 0.65, we stop training our annotators and move to the next sub-phase. Figure 1 illustrates the Cohen's Kappa score between 6 annotators of 5 training sub-phases.

After the training phase is done, annotators are able to label the rest of the dataset in the second phase, the labeling phase. There are 2 sub-phases, each sub-phase has about 7,500 reviews. Three annotators are responsible for annotating 7,500 separately. For disagreed reviews, we automatically choose the label which is chosen by 2/3 annotators.

### 3.3 Statistics

Our dataset contains 16,227 reviews, including 7 sentiment aspects and 1 aspect indicating spam reviews. Table 2 shows some examples of reviews in

[1] https://shopee.vn/

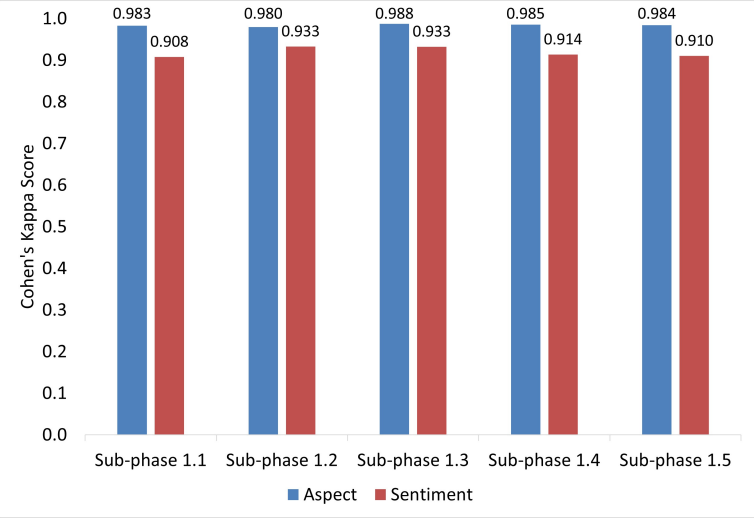



Figure 1: Inter Annotation Agreement Score of the training phase.

| Aspect | Definition |
|---|---|
| SMELL | Reviews mention the scent of the lipstick. |
| COLOUR | Reviews mention lipstick color such as dark or light |
| TEXTURE | Reviews mention the characteristic and quality of the lipstick such as information about the moisture and dryness of the lipstick. |
| PRICE | Reviews mention the price of lipsticks, whether it is affordable or not. |
| STAYINGPOWER | Reviews mention the adhesion of lipsticks on the lip. |
| SHIPPING | Reviews mention the delivery service such as the time, the shipper's attitude. |
| PACKING | Reviews mention the quality of packing, whether the lipsticks are well packed or not. |
| OTHERS | Spam reviews. |

Table 1: Aspect Definition.

our dataset and the corresponding aspects and polarities. Figure 2 illustrates the distribution for each aspect and sentiment in the dataset. Overall, the sentiment Positive accounts for the largest part in all aspects, except the aspect OTHERS. In addition, the number of reviews having aspect COLOUR is over 7000 reviews, accounting for nearly 50% of reviews, which shows the high concern of customers about this aspect when buying lipsticks. The high imbalance between sentiments can be seen in the aspect PRICE and PACKING. The reason is the price on e-commerce platforms is really competitive and cheaper than in physical stores, so this tends to receive good reviews. Meanwhile, the PACKING aspect receives good reviews because this is an aspect that sellers can control and manage to receive a good initial impression from buyers. Table 3 shows overview statistics of our dataset.

## 4 Aspect-based Sentiment Analysis model

This section gives information about the model that is used for the aspect-based sentiment analysis problem. The model has three main components: input layer & hidden layers, output layer for single-task learning, and output layer for multi-task learning approach. For the input layer and hidden layers in subsection 4.1, the model architecture is the same between single-task learning and multi-task learning. Based on the learning approach, the output layer can be different. The subsection 4.2 and 4.3 give more details about the output layers of two approaches.

| Review | Aspect | Polarity |
|---|---|---|
| về mẫu mã khá cute mới đánh thử thì thấy oke màu lên môi đẹp mùi cũng thơm *(the sample is so cute, I just used it, the colour is great on my lip and the smell is also fragant too.)* | SMELL, COLOUR | SMELL:Positive, COLOUR:Positive |
| Công dụng: màu lì Son đẹp lắm ạ đóng gói rất kĩ hàng không bị móp méo xài rất mướt *(Uses: the colour is very adhesive, the beautiful lipsticks. Very well packed, no dents)* | STAYINGPOWER | STAYINGPOWER: Positive |
| Giao hàng lâu. Chờ mòn mỏi luôn í. Chak do hàng quốc tế. Màu khá ok. Sẽ ủng hộ shop *(Long delivery. Tired of waiting. Maybe because of international delivery. The colour is ok. I will support the shop.)* | COLOUR, SHIPPING | COLOUR:Positive, SHIPPING:Negative |

Table 2: Several examples of the dataset.

| Set | Review | Avg aspect/ review | Positive | Neutral | Negative | Total sentiment |
|---|---|---|---|---|---|---|
| Train | 12,981 | 2.02 | 18,694 | 3,822 | 3,715 | 26,231 |
| Dev | 1,623 | 2.01 | 2,336 | 462 | 466 | 3,264 |
| Test | 1,623 | 2.02 | 2,298 | 511 | 471 | 3280 |

Table 3: Statistics about the experimental dataset.

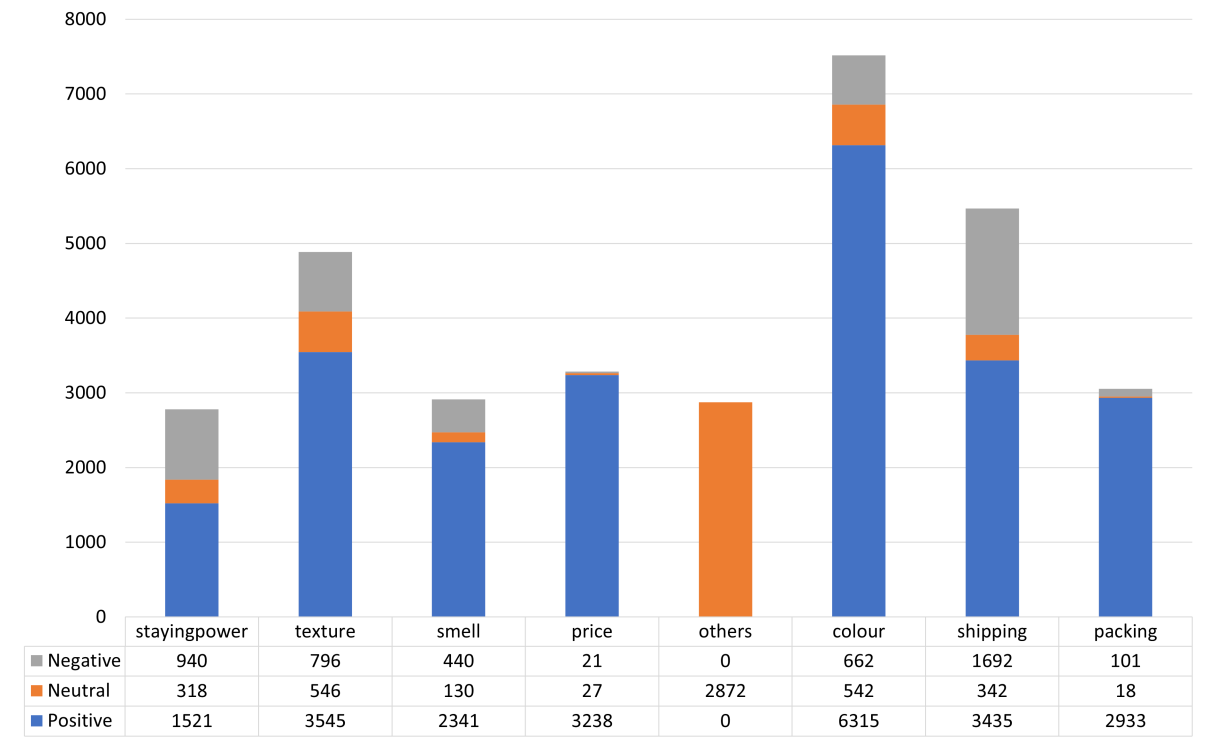
### 4.1 Input and Hidden layers

The input and hidden layers are illustrated in the figure 3. After pre-processing the reviews, a tokenizer layer will be used to convert from words to indexes based on the vocabulary index. The ELMO pre-trained word embedding (Vu, Vu, Tran, & Jiang, 2019) is used as the initialization for the embedding layer and this layer creates a representative vector for every word. The SpatialDropout1D layer helps to reduce the overfitting problem. A Bi-LSTM (Hochreiter & Schmidhuber, 1997) and Bi-GRU (Chung, Gulcehre, Cho, & Bengio, 2014) are used parallelly to obtain as much information as possible. The Bi-LSTM layer can save valuable information from the beginning of the reviews and utilize it to predict the label. After the Bi-LSTM or Bi-GRU layers, the Conv1D layers convert multi-dimensional matrices to 1D matrices and GlobalMaxPooling and GlobalAveragePooling will extract the maximum element of the matrices as well as the average element. The reason why parallel RNN-based neuron networks are used is that they can extract more features than a single neuron network. All pooling layers are concatenated and go through a dense layer before passing to output layers. It is worth noting that depending on the type of learning, which will be described at 4.2 and 4.3 below, the output layer can be different.

To find the best model architecture, we stack layers one by one from the Bi-LSTM layer or Bi-GRU layer to Bi-LSTM+Conv1D or Bi-GRU+Conv1D to the full layer version as in figure 3.

### 4.2 Output layer for Single-Task Learning approach

Single-Task Learning (STL) is a type of Deep Learning, in which a model is only specific to a task. In ABSA, there are many sub-tasks and they can be categorized as single-task learning. There are two main sub-tasks in ABSA, which are aspect category detection, and aspect sentiment classification (Zhang et al., 2022). After the aspect category detection model extracts the aspects in a review, the sentiment estimation model will predict the polarity of that aspect in the review. Because of this, for every aspect, a sentiment classification model needs to be built to estimate the polarity of that aspect.




| | stayingpower | texture | smell | price | others | colour | shipping | packing |
|---|---|---|---|---|---|---|---|---|
| Negative | 940 | 796 | 440 | 21 | 0 | 662 | 1692 | 101 |
| Neutral | 318 | 546 | 130 | 27 | 2872 | 542 | 342 | 18 |
| Positive | 1521 | 3545 | 2341 | 3238 | 0 | 6315 | 3435 | 2933 |

Figure 2: Aspect and Sentiment distribution in the dataset.

For our problem, there is one multi-label classification model to detect whether one or more aspects exist in the reviews. After that, seven sentiment classification models for seven aspects will be used to predict the sentiment of that aspect in the review. However, if the aspect detection model does not detect an aspect, the corresponding sentiment classification model will not estimate the polarity of that aspect in the reviews.

The output for the single-task learning approach has 8 units corresponding to 8 aspects for the aspect detection model or 3 units corresponding to 3 polarities for the sentiment classification models. The figure 4 illustrates the example of the sentiment classification for the aspect COLOUR. We have to build 6 other models like this for the sentiment classification and 1 model for the aspect detection task.

### 4.3 Output layer for Multi-Task Learning approach

In the multi-task learning (MTL) approach, the aspects and corresponding sentiment are predicted simultaneously. This approach has been used in a lot of previous research (He et al., 2019), (Luo, Li, Liu, & Zhang, 2019) and it proves its effectiveness in the aspect-based sentiment analysis task. Inspirited by these researches, we design a multi-task recurrent neural network for our own problem. The input layer and hidden layers in the multi-task model are similar to model architecture in 4.1, and they serve as the shared layers, however, for the task-specific layers, there are an aspect detection task and seven sentiment classification tasks. An example of a multi-task learning model is illustrated in the figure 5. The aspects and corresponding sentiment will be learned and predicted simultaneously.

The example of the Multi-task Learning model is illustrated in figure 5. For the aspect detection task, the output layer is a dense layer with 8 nodes, corresponding to 8 aspects. The sigmoid activation is used because this task is multi-label classification and a review can have more than one aspect. For the sentiment classification tasks, there are 7 tasks corresponding to 7 aspects that are needed to classify the sentiment (except the aspect others). In each task, the output has 4 nodes indicating positive, neutral, negative, or nan (the aspect is absent in this review so there is no sentiment for this aspect). The softmax

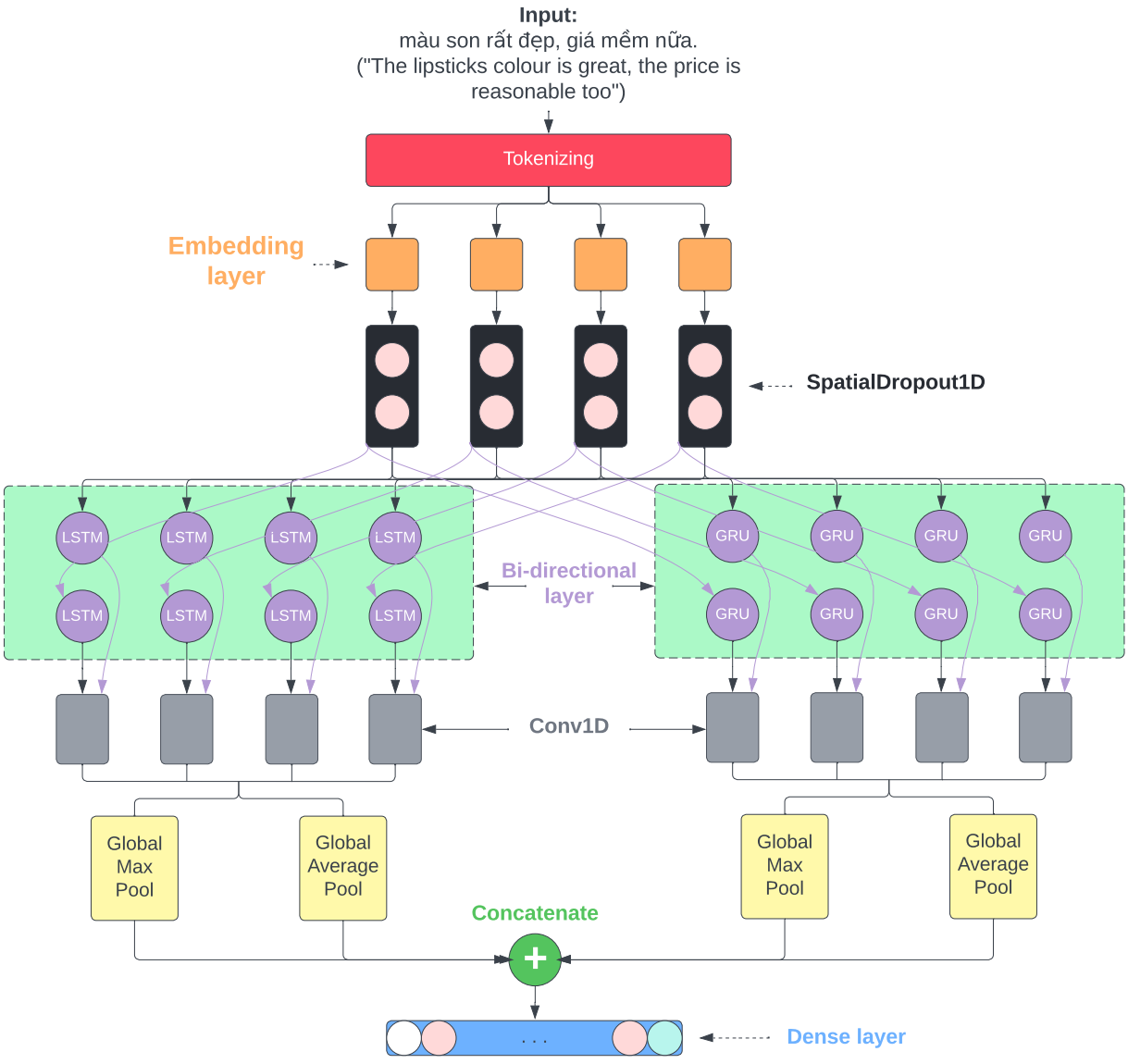



Figure 3: The proposed input and hidden layers

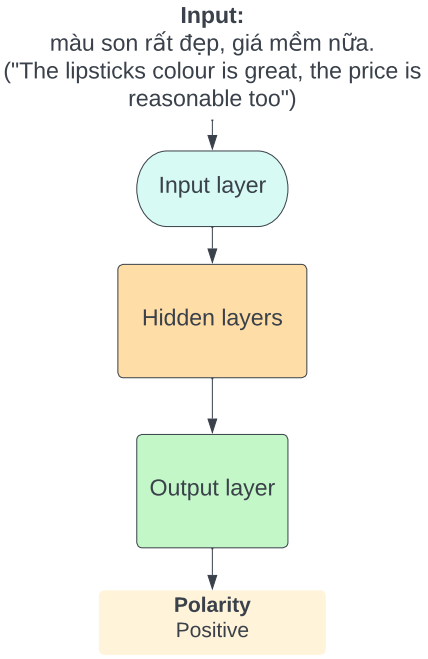



Figure 4: The example STL model for classifying the sentiment of aspect COLOUR

activation function is used for these tasks.

## 5 Experiment

### 5.1 Experimental settings

The embedding layer has the dimension of 1024, which is also the dimension of the ELMO pre-trained word embedding. The number of units in BiLSTM or BiGRU is 400 units and the activation function in these layers is tanh function. In the Conv1D layer, the kernel size is 2 and the filter is 128, which means reducing the input dimension from 400 to 128.

We use binary_crossentropy loss function for the aspect detection task and categorical_crossentropy for the sentiment classification tasks. The Adam (Kingma & Ba, 2014) is used as the optimizer with the learning rate at 0.0001. The batch size is 128 and the number of epochs for training is 50 epochs. Early Stopping is used to reduce the overfitting problem. For the implemented code, one can see it at this repository at Github: `https://github.com/linh222/absa_vn_lipsticks_review`.

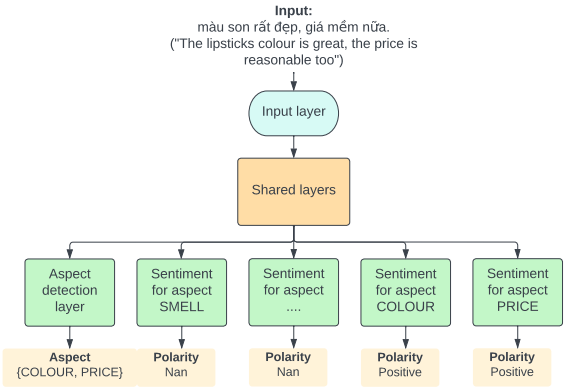



Figure 5: The example MTL model

### 5.2 Evaluation metrics

Because of the imbalance of aspect and sentiment in the experimental dataset, F1-score is used to evaluate the performance of models to take into account the imbalance and give a proper view of the effectiveness of the predictive models. The $F1_{ad}$ and $F1_{sc}$ denote for F1-score in aspect detection and sentiment classification tasks. The formula of F1-score (according to (Sokolova & Lapalme, 2009)) is as follow with n is the number of samples and TP, FP, and FN denote for TruePosituve, FalsePositive, and FalseNegative, respectively:

$$Precision = \frac{\sum_{i=1}^{n} TP_i}{\sum_{i=1}^{n} TP_i + FP_i}$$

$$Recall = \frac{\sum_{i=1}^{n} TP_i}{\sum_{i=1}^{n} TP_i + FN_i}$$

$$F1 - score = \frac{2(Precision * Recall)}{(Precision + Recall)}$$

### 5.3 Results and Discussion

Weighted averages of F1-Score are used to measure the performance of our models. The table 4 illustrates the performance of implemented model architectures on the dataset. Most models achieve a performance higher than 87% for single-task learning and over 80% for multi-task learning. The best model in both approaches is BiGRU + Conv1D with 98.09% F1_ad, 91.01% F1_se for the single-task learning approach, and 97.1% F1_ad, 86.92% F1_se for the multi-task learning approach.

| Model | STL | | MTL | |
|---|---|---|---|---|
| Metric | F1_ad | F1_sc | F1_ad | F1_sc |
| BiLSTM | 97.62 | 87.56 | 94.12 | 81.44 |
| BiGRU | 97.90 | 87.22 | 95.61 | 80.94 |
| BiLSTM +Conv1D | 98.00 | 89.90 | 96.89 | 86.26 |
| **BiGRU +Conv1D** | **98.09** | **91.01** | **97.51** | **86.92** |
| BiLSTM +BiGRU +Conv1D | 97.92 | 90.15 | 96.63 | 85.65 |

Table 4: Results of implemented models on the dataset (%)

| Aspect | STL | MTL |
|---|---|---|
| SMELL | 99.23 | 98.01 |
| COLOUR | 97.06 | 96.60 |
| STAYINGPOWER | 98.48 | 95.67 |
| PRICE | 98.33 | 96.37 |
| SHIPPING | 97.69 | 97.15 |
| PACKING | 98.18 | 95.89 |
| TEXTURE | 96.80 | 94.88 |
| OTHERS | 97.48 | 94.68 |

Table 5: The F1-score for aspect detection in each aspect(%)

| Aspect | Positive | Neutral | Negative |
|---|---|---|---|
| SMELL | 94.22 | 0.00 | 74.01 |
| COLOUR | 95.66 | 36.90 | 67.48 |
| STAYING-POWER | 84.97 | 47.63 | 86.49 |
| PRICE | 99.01 | 60.00 | 0.00 |
| SHIPPING | 93.92 | 33.23 | 91.97 |
| PACKING | 98.55 | 0.00 | 44.90 |
| TEXTURE | 92.22 | 53.87 | 77.57 |

Table 6: The F1-score of sentiment in each aspect in Single-Task Learning(%)

| Aspect | Positive | Neutral | Negative |
|---|---|---|---|
| SMELL | 93.41 | 0.00 | 76.32 |
| COLOUR | 94.35 | 28.04 | 63.95 |
| STAYING-POWER | 82.98 | 27.84 | 89.73 |
| PRICE | 96.32 | 0.00 | 0.00 |
| SHIPPING | 92.01 | 33.28 | 90.42 |
| PACKING | 94.28 | 0.00 | 17.83 |
| TEXTURE | 89.52 | 48.92 | 76.90 |

Table 7: The F1-score of sentiment in each aspect in Multi-Task Learning(%)

Table 5 proves that our models are robust for aspect detection task. The F1-score for the aspect detection task is always higher than 96% for the single-task learning approach and 94% for the multi-task learning approach.

For sentiment classification task, the results on table 6 and 7 show that the performance in classifying the sentiment Positive is better than other sentiments. The model can detect the sentiment of some aspects such as STAYINGPOWER, SHIPPING, TEXTURE very accurately. However, some other aspects are poorly in sentiment classification such as SMELL, PACKING, especially on the sentiment Neutral and Negative. The reason is there are a lot of positive reviews for beauty products on Shopee while very few neutral and negative reviews. The imbalance of the dataset is one of the big challenges which will be addressed in future work.

## 6 Conclusion and Future Work

This paper deals with the aspect-based sentiment analysis of beauty products reviews on e-commerce websites. In this paper, we presented a new dataset containing 16,277 reviews about lipstick in e-commerce platforms for the task aspect-based sentiment analysis. There are 32,775 pairs of aspect and sentiment in the dataset. For the task of predicting the aspect and sentiment of reviews, we compared single-task learning and multi-task learning and received the result that single-task learning is better than multi-task learning. However, the implementation and complexity of single-task learning are significantly higher than multi-task learning so this is a trade-off between accuracy and complexity. For model architecture, the

combination of BiGRU and Conv1D outperformed other model architecture in both single-task learning and multi-task learning. The best F1-score belonged to BiGRU+Conv1D in the single-task learning approach at 98.09% for aspect detection and 91.01% for sentiment classification.

For future work, we are considering building an automatic pipeline to collect reviews in e-commerce platforms, process the reviews, predict the aspect and sentiment of reviews and visualize the results on a dashboard. This pipeline will bring a broader look for sellers to understand their products and for customers to consider before making a purchase. In addition, we will apply some state-of-the-art models such as transformer models to improve the accuracy of sentiment classification.

## References


Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). *Empirical evaluation of gated recurrent neural networks on sequence modeling.* arXiv. doi: 10.48550/ARXIV.1412.3555

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*(1), 37-46. doi: 10.1177/001316446002000104

He, R., Lee, W., Ng, H., & Dahlmeier, D. (2019, 01). An interactive multi-task learning network for end-to-end aspect-based sentiment analysis. In (p. 504-515). doi: 10.18653/v1/P19-1048

Hochreiter, S., & Schmidhuber, J. (1997, 12). Long short-term memory. *Neural computation*, *9*, 1735-80. doi: 10.1162/neco.1997.9.8.1735

Kingma, D. P., & Ba, J. (2014). *Adam: A method for stochastic optimization.* arXiv. doi: 10.48550/ARXIV.1412.6980

Luo, H., Li, T., Liu, B., & Zhang, J. (2019). *Doer: Dual cross-shared rnn for aspect term-polarity co-extraction.* arXiv. doi: 10.48550/ARXIV.1906.01794

Nguyen, H. T., Nguyen, H. V., Ngo, Q. T., Vu, L. X., Tran, V. M., Ngo, B. X., & Le, C. A. (2018). Vlsp shared task: sentiment analysis. *Journal of Computer Science and Cybernetics*, *34*(4), 295–310.

Nguyen, K. T.-T., Huynh, S. K., Phan, L. L., Pham, P. H., Nguyen, D.-V., & Van Nguyen, K. (2021). *Span detection for aspect-based sentiment analysis in vietnamese.* arXiv. doi: 10.48550/ARXIV.2110.07833

Nguyen, K. V., Nguyen, V. D., Nguyen, P. X. V., Truong, T. T. H., & Nguyen, N. L.-T. (2018). Uit-vsfc: Vietnamese students' feedback corpus for sentiment analysis. In *2018 10th international conference on knowledge and systems engineering (kse)* (p. 19-24). doi: 10.1109/KSE.2018.8573337

Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., AL-Smadi, M., ... Eryiğit, G. (2016, January). SemEval-2016 Task 5: Aspect Based Sentiment Analysis. In *International Workshop on Semantic Evaluation* (p. 19 - 30). San Diego, United States. doi: 10.18653/v1/S16-1002

Pontiki, M., Galanis, D., Papageorgiou, H., Manandhar, S., & Androutsopoulos, I. (2015). Semeval-2015 task 12: Aspect based sentiment analysis. In *Proceedings of the 9th international workshop on semantic evaluation (semeval 2015)* (pp. 486–495).

Pontiki, M., Galanis, D., Pavlopoulos, J., Papageorgiou, H., Androutsopoulos, I., & Manandhar, S. (2014, August). SemEval-2014 task 4: Aspect based sentiment analysis. In *Proceedings of the 8th international workshop on semantic evaluation (SemEval 2014)* (pp. 27–35). Dublin, Ireland: Association for Computational Linguistics. doi: 10.3115/v1/S14-2004

Sokolova, M., & Lapalme, G. (2009, 07). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, *45*, 427-437. doi: 10.1016/j.ipm.2009.03.002

Vu, X.-S., Vu, T., Tran, S. N., & Jiang, L. (2019). Etnlp: A visual-aided systematic approach to select pre-trained embeddings for a downstream task. In *Proceedings of the international conference recent advances in natural language processing (ranlp).*

Wang, Y., Huang, M., Zhu, X., & Zhao, L. (2016, November). Attention-based LSTM for aspect-level sentiment classification. In *Proceedings of the 2016 conference on empirical methods in natural language processing* (pp. 606–615). Austin, Texas: Association for Computational

Linguistics. doi: 10.18653/v1/D16-1058

Zhang, W., Li, X., Deng, Y., Bing, L., & Lam, W. (2022). *A survey on aspect-based sentiment analysis: Tasks, methods, and challenges.* arXiv. doi: 10.48550/ARXIV.2203.01054

Zhou, X., Wan, X., & Xiao, J. (2015). Representation learning for aspect category detection in online reviews. In *Proceedings of the twenty-ninth aaai conference on artificial intelligence* (p. 417–423). AAAI Press.