

**VIETNAM NATIONAL UNIVERSITY  
UNIVERSITY OF ENGINEERING AND TECHNOLOGY**



**Magneto @ DS&KTLab**

**ASPECT-ORIENTED SENTIMENT ANALYSIS  
FOR VIETNAMESE E-COMMERCE REVIEWS**

**STUDENT SCIENTIFIC RESEARCH REPORT**

**Faculty of Information Technology**

**HANOI - 2021**

**VIETNAM NATIONAL UNIVERSITY  
UNIVERSITY OF ENGINEERING AND TECHNOLOGY**

**Le Thi Phuong  
Le Minh Binh  
Bui Khanh Huyen  
Tran Khanh Hung**

**ASPECT-ORIENTED SENTIMENT ANALYSIS  
FOR VIETNAMESE E-COMMERCE REVIEWS**

**STUDENT SCIENTIFIC RESEARCH REPORT**

**Faculty of Information Technology**

**Supervisors:**

**MSc. Le Hoang Quynh**

**MSc. Can Duy Cat**

**HANOI - 2021**

# Tóm tắt

Tóm tắt để sau

***Keywords:*** *Graph Attention Networks(GAT), transductive, Vietnamese multi-aspect dataset.*

# Acknowledgements

Lời đầu tiên, chúng em xin cảm ơn Cô Nguyễn Thị Cẩm Vân và các thầy cô trong lab Data Science and Knowledge Technology Laboratory at University of Engineering and Technology. Các thầy cô đã luôn giúp đỡ và hướng dẫn chúng em viết ra bài báo này.

Đồng thời em cũng xin cảm ơn khoa Công nghệ thông tin - Trường Đại học Công nghệ đã tạo điều kiện để cho chúng em có được trải nghiệm thực tế trong môi trường nghiên cứu khoa học và tiếp thu nhiều kiến thức bổ ích.

# Declaration

Chúng tôi tuyên bố rằng tác phẩm của chúng tôi là do chính chúng tôi sáng tác và tác phẩm đó không được nộp cho bất kỳ bằng cấp hoặc trình độ chuyên môn nào khác. Chúng tôi xác nhận rằng tác phẩm được gửi là của riêng bạn, trừ khi tác phẩm đã tạo thành một phần của các ấn phẩm đồng tác giả đã được đưa vào. Đóng góp của chúng tôi và của các tác giả khác cho tác phẩm này đã được chỉ ra rõ ràng dưới đây. Chúng tôi xác nhận rằng tín dụng thích hợp đã được đưa ra trong bài viết này khi tài liệu tham khảo đã được thực hiện cho công việc của người khác.

Chúng tôi xác nhận rằng, theo hiểu biết tốt nhất của chúng tôi, bài viết của chúng tôi không vi phạm bản quyền của bất kỳ ai cũng như không vi phạm bất kỳ quyền sở hữu nào và rằng mọi ý tưởng, kỹ thuật, trích dẫn hoặc bất kỳ tài liệu nào khác từ tác phẩm của người khác được đưa vào luận án của tôi, đã xuất bản hoặc mặt khác, được thừa nhận đầy đủ theo các thông lệ tham chiếu tiêu chuẩn. Hơn nữa, trong phạm vi mà chúng tôi đã bao gồm tài liệu có bản quyền, chúng tôi xác nhận rằng chúng tôi đã nhận được sự cho phép bằng văn bản từ (những) chủ sở hữu bản quyền để đưa (những) tài liệu đó vào tác phẩm của chúng tôi và có đầy đủ quyền tác giả để cải thiện những tài liệu này.

Nhóm sinh viên

**Đoàn Văn Nguyên**  
**Nguyễn Đức Trọng**  
**Nguyễn Trần Đạt**

# Table of Contents

<b>Tóm tắt</b>	<b>iii</b>
<b>Acknowledgements</b>	<b>iv</b>
<b>Declaration</b>	<b>v</b>
<b>Table of Contents</b>	<b>vi</b>
<b>1 Giới thiệu</b>	<b>1</b>
1.1 Nhận Diện Cảm Xúc Trong Hội Thoại (Emotion Recognition in Conversation - ERC)	1
1.2 Multimodal ERC	1
1.3 Graph-based Multimodal ERC	1
1.4 Feature Propagation	2
1.4.1 Graph Completion Network - GCNET	2
1.4.2 Missing Modality Imagination Network - MMIN	2
1.5 Đóng góp	2
<b>2 Nghiên Cứu Liên Quan</b>	<b>3</b>
2.1 Mạng Đồ Thị Hoàn Thiện (GC-NET)	3
2.2 Data Annotation	3
2.3 Dataset Analysis	5
<b>3 Phương Thức</b>	<b>8</b>
3.1 Chuẩn Bị Dữ Liệu	8
3.2 Mã Hoá Thông Tin Ngữ Cảnh	9
3.3 Graph Attentional Layer	10
3.4 Phân Lớp	10
<b>4 Experiments and Results</b>	<b>12</b>
4.1 Experiment Setup	12

4.1.1	Experimental Data . . . . .	12
4.1.2	Baseline Methods . . . . .	12
4.1.3	Evaluation Metrics . . . . .	12
4.2	Experiment Result and Analysis . . . . .	14
4.2.1	Classic Models with One-Hot Representation Method. . . . .	14
4.2.2	Logistics Regression with Different Representation Methods . . .	14
4.2.3	Chi-Squared with/without Word-Window Locating compared with CNN + PhoBERT . . . . .	15
4.2.4	WWL + CS detailed statistics in all domains . . . . .	16
4.2.5	Error Analysis . . . . .	16
4.3	Demonstrations . . . . .	17
4.3.1	Data Visualization . . . . .	17
4.3.2	Web Application Demonstration . . . . .	18
<b>5</b>	<b>Conclusions . . . . .</b>	<b>21</b>
	<b>Conclusions . . . . .</b>	<b>21</b>
	<b>References . . . . .</b>	<b>23</b>

# Chapter 1

## Giới thiệu

### 1.1 Nhận Diện Cảm Xúc Trong Hội Thoại (Emotion Recognition in Conversation - ERC)

Cảm xúc là một khía cạnh quan trọng trong giao tiếp hàng ngày của con người. Vì vậy, trong lĩnh vực xử lý ngôn ngữ tự nhiên, nhận diện cảm xúc là một mảng nghiên cứu càng ngày càng được quan tâm. Công nghệ Nhận diện cảm xúc trong hội thoại (ERC) có vai trò xác định trạng thái cảm xúc của người nói trong một cuộc hội thoại.

ERC có nhiều ứng dụng tiềm năng như hỗ trợ đàm thoại, phân tích cho các thử nghiệm pháp lý và dịch vụ y tế điện tử, v.v. ERC có rất nhiều tiềm năng khai thác dữ liệu từ các mạng xã hội nổi tiếng trên thế giới như Facebook, Twitter, Youtube, Reddit etc. ERC cũng là một thành phần quan trọng để xây dựng các tương tác máy tính tự nhiên như con người và có thể trả lời một cuộc đối thoại có tính cảm xúc.

### 1.2 Multimodal ERC

Nhận thấy các nghiên cứu về ERC tập trung vào loại dữ liệu văn bản, \*trích bài báo MMGCN\* đã đề xuất kết hợp nhiều kiểu dữ liệu khác trong hội thoại (ảnh và tiếng) và ERC, từ đó công bố MMGCN.

### 1.3 Graph-based Multimodal ERC

Theo đó, mô hình đồ thị tập trung (Graph Attention Network - GAT) được áp dụng để liên kết các nút trong đồ thị hiệu quả hơn các mô hình trước đây, bằng việc gán trọng số



tập trung khác nhau cho từng nút.

## 1.4 Feature Propagation

Tuy nhiên, trong nhiều dự án thực tế, các dữ liệu thành phần không phải lúc nào cũng đầy đủ. Ví dụ, một câu thoại có thể khuyết thông tin do lỗi dịch thuật hay lỗi của phần mềm chuyển âm thanh thành giọng nói; hay bản ghi âm bị nhiễu do môi trường hay do thiết bị ghi âm. Đây là một chướng ngại lớn trong việc nhận diện cảm xúc trong hội thoại.

\*trích báo FP\* đã đề xuất một phương pháp xử lý dữ liệu bị khuyết, gọi là Lan Truyền Đặc Trưng, sử dụng tối ưu hoá Dirichlet. Phương pháp này được hai nhóm tác giả khác kế thừa và phát triển:

### 1.4.1 Graph Completion Network - GCNET

GCNET gồm hai mô-đun đồ thị mạng nơ-ron Speaker và Temporal để tách hai loại dữ liệu tương ứng, sau đó được phân loại và tối ưu hoá.

### 1.4.2 Missing Modality Imagination Network - MMIN

MMIN sử dụng học máy để dự đoán các phần dữ liệu bị khuyết từ dữ liệu có sẵn, trong các điều kiện khác nhau, bằng cách tìm liên hệ giữa các modality với nhau.

## 1.5 Đóng góp

Chúng tôi kết hợp phương pháp lan truyền đặc trưng vào mạng chú ý đồ thị, để xử lý các trường hợp khuyết dữ liệu trong học máy. Mô hình này gồm...

## Chapter 2

# Nghiên Cứu Liên Quan

### 2.1 Mạng Đồ Thị Hoàn Thiện (GC-NET)

To serve our dedicated approach, we have to handle textual data that are reviews on E-commerce platforms. We choose Scrapy<sup>1</sup> to crawl the data from online shopping websites including <http://www.tiki.vn> and <http://www.shopee.vn>. They are two of Vietnam's most common and trusted e-commerce platforms. We focus on technology and mother & baby domains since they are the most interested domain in Vietnamese e-commerce. These raw data are then used as input for Data Annotation step that we will describe in Section 2.2.

### 2.2 Data Annotation

We used Docanno as a tool for data annotation. First of all, we overview the data set collected from crawling process to figure out aspects that were frequently mentioned on users' reviews.

After discussion, we extracted aspect terms in the sentences and labeled the sentiment polarities with respect to the aspect terms using Docanno. Table 1 illustrates how we handle each aspect. For Technology domain, we predefined eight coarse aspect categories: price, service, delivery, performance, hardware, authenticity, accessories, design. For Mother & Baby, these are: price, service, delivery, safety, quality and authenticity. In each aspect, it is considered negative when there is at least a complaint regarding that aspect, otherwise it is considered positive.

---

<sup>1</sup><https://scrapy.org/>

<i>Domain</i>	<i>Aspect</i>	<i>Description</i>
<b>Technology</b>	Price	Cut/ reduce/ slash/ low price considered positive while increase/ put up/ raise/ high price is considered negative.
	Service	Nice, fast, efficient, enthusiastic support is considered positive while no reply, irresponsibility or carelessly packing is considered negative.
	Delivery	The quality, speed and cost of shipping process. If it is fast, carefully shipped and the cost is low, it is considered positive, otherwise is negative.
	Performance	Fast processing speed is considered positive while lag or latency in the middle of an app or apps shut down unexpectedly is considered negative.
	Hardware	The quality of the hardware of the device including display, chip, battery, cameras, storage, RAM, etc.
	Authenticity	Genuinely produced product is considered positive while fake, imitation is consider negative.
	Accessories	Fully provided, good quality accessories are considered positive; low quality or missed are considered negative.
	Appearance	Product with nice design, nice color, luxury, etc. are considered positive, product with scratches, awful design is consider negative.
<b>Mother &amp; Baby</b>	Price	Cut/ reduce/ slash/ low price considered positive while increase/ put up/ raise/ high price is considered negative.
	Service	Nice, fast, efficient, enthusiastic support is considered positive while no reply, irresponsibility or carelessly packing is considered negative.
	Delivery	The quality, speed and cost of shipping process. If it is fast, carefully shipped and the cost is low, it is considered positive, otherwise negative.
	Safety	Product having obvious expiration date and has not expired, safe to use is considered positive, while product that is expired or causes allergy, rashes is considered negative.
	Quality	Customer experience on the product: the softness, absorbency of diapers, the taste and smell of milk, etc.
	Authenticity	Genuinely produced product is considered positive while fake, imitation is consider negative.

Table 2.1: Aspect description

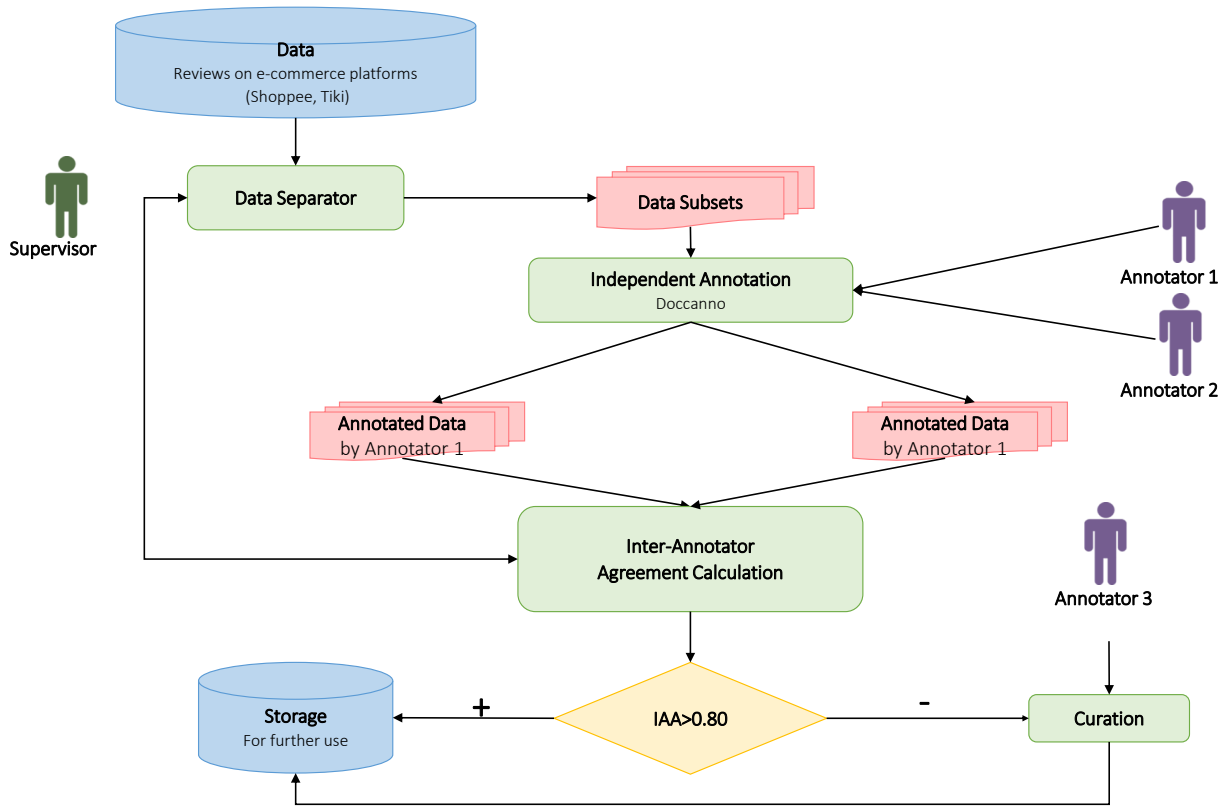


Figure 2.1: Data Annotation process

Data then separated in subsets, each of them was then evaluated and labeled by 2 annotators independently. Next, we measured the agreement between two annotators, if the result is too low (Inter-Annotator Agreement or  $IAA < 0.8$ ), manually reviewing – process will be applied. The labeled data then stored as input for Pre-processing step.

## 2.3 Dataset Analysis

The statistics of Vietnamese E-commerce dataset (VECD) including Shopee and Tiki in two different domains: Technology and Mother & Baby are reported in figure 1 to 4. VECD consists of 3016 instances of Shopee Mother & Baby, 2986 instances of Tiki , 3002 instances of Shopee Technology and 3236 instances of Tiki Technology, making up 12240 instances in total.

	Mother & Baby		Technology	
	<i>Shopee</i>	<i>Tiki</i>	<i>Shopee</i>	<i>Tiki</i>
<b>Total aspect</b>	6	6	8	8
<b>Aspect count per sentence</b>	1.753234	1.497347	2.087627	1.746551

Table 2.2: The average number of aspects mentioned

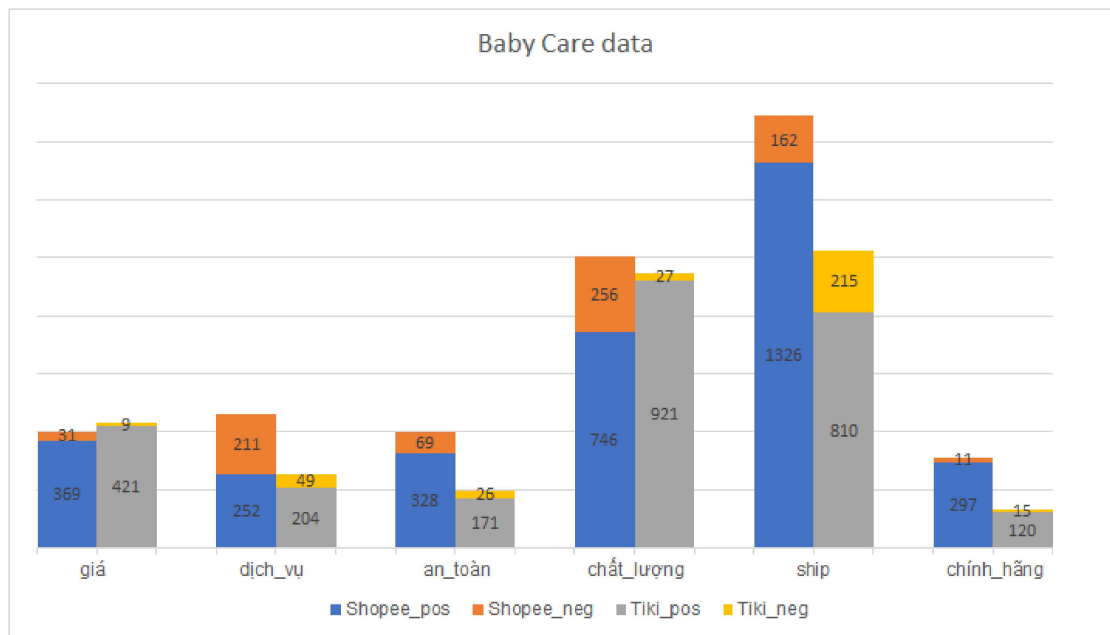


Figure 2.2: Mother & Baby data statistics

In Mother & Baby domain, Shipping and Quality appear to be the most concerned aspects. Shopee has 1541/3216 comments on Shipping and 773/3216 comments on Quality, while in Tiki they are 972/2986 and 1177/2986 respectively. Authenticity only occasionally mentioned in Tiki's comments, just in 131 cases while the others are mentioned in over 200 comments, in both platforms, including Authenticity mentioned in Shopee.

Similarly, in Technology domain, Appearance is frequently mentioned in Shopee while it is Shipping in Tiki. More specifically 1064 out of 3002 comments on Shopee have information relating to products' appearance, compared to 1021 out of 3236 comments on Tiki having information about Shipping. Price and Authenticity are the most imbalanced aspects, with only 4 negative comments on Price and 12 negative comments on Authenticity in Shopee while in Tiki the figures are 31 and 24, respectively. In Tiki, although the number of comments on Service, Hardware, Performance and Accessories is comparatively low, the data on these aspects are relatively balanced, with the difference below 16%.

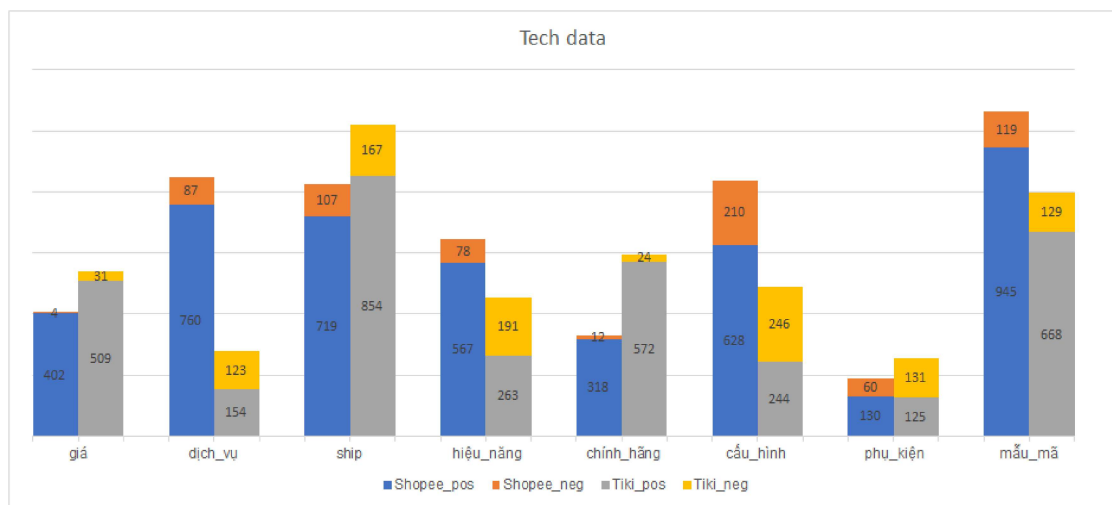


Figure 2.3: Technology data statistics

# Chapter 3

## Phương Thức

Trong chương này, chúng tôi sẽ trình bày phương thức chuẩn bị dữ liệu hội thoại,

### 3.1 Chuẩn Bị Dữ Liệu

Chúng ta xác định một cuộc trò chuyện  $C = \{(u_i, y_i)\}_{i=1}^L$ , trong đó  $L$  là số lượng các câu nói trong cuộc trò chuyện,  $u_i$  là câu nói thứ  $i^{th}$  trong  $C$  và  $y_i$  là nhãn đúng của  $u_i$ . Ở đây,  $y_i \in \{1, 2, \dots, c\}$  và  $c$  là tổng số nhãn. Mỗi câu nói  $u_i$  của người nói  $p_{s(u_i)}$ , trong đó hàm  $s(\cdot)$  ánh xạ chỉ số của câu nói vào người nói tương ứng.

Cho mỗi câu nói  $u_i$ , chúng ta trích xuất đặc trưng đa mô hình  $x_i = \{x_i^m\}_{m \in \{a, l, v\}}$ . Ở đây,  $x_i^a \in \mathbb{R}^{d_a}$ ,  $x_i^l \in \mathbb{R}^{d_l}$  and  $x_i^v \in \mathbb{R}^{d_v}$  là các đặc trưng mức câu nói của các mô hình âm thanh, từ vựng và hình ảnh tương ứng. Và  $\{d_m\}_{m \in \{a, l, v\}}$  là chiều của đặc trưng của mỗi mô hình.

Để tạo ra các trường hợp mất mô hình giống như trong thế giới thực, chúng ta sẽ loại bỏ ngẫu nhiên một số mô hình, nhưng đảm bảo ít nhất có một mô hình cho mỗi mẫu, theo các công trình trước đó.

Do đó, một tập dữ liệu  $M$  không hoàn thiện có  $(2^M - 1)$  mẫu thiếu sót khác nhau. Hình ?? minh họa một tập dữ liệu trimodal ( $M = 3$ ) với bảy mẫu thiếu sót. Giả sử  $\sigma_i$  là mẫu thiếu sót của  $u_i$  và  $\phi(\cdot)$  là một hàm mô tả mỗi mẫu thiếu sót với các chế độ có sẵn. Biểu diễn không hoàn thiện của  $u_i$  được đánh dấu là  $\tilde{x}_i = \{\lambda_i^m x_i^m\}_{m \in \{a, l, v\}}$ , trong đó  $\lambda_i^m$  được định nghĩa như sau:

$$\lambda_i^m = \begin{cases} 1, m \in \phi(\sigma_i) \\ 0, m \notin \phi(\sigma_i) \end{cases} \quad (3.1)$$

## 3.2 Mã Hoá Thông Tin Ngữ Cảnh

Như đã đề cập ở trên, thông tin ngữ cảnh hội thoại là quan trọng để dự đoán nhãn cảm xúc của mỗi lời nói. Vì vậy, việc mã hóa thông tin ngữ cảnh vào biểu diễn tính năng của lời nói là có lợi. Chúng ta tạo ra mã hóa tính năng của lời nói được nhận thông tin ngữ cảnh cho mỗi modality qua mã hóa modality tương ứng. Cụ thể, chúng ta sử dụng mạng LSTM kết hợp hai chiều để mã hóa thông tin ngữ cảnh liên tục cho modality văn bản. Đối với modality âm thanh và hình ảnh, chúng ta sử dụng mạng kết nối đầy đủ. Mã hóa tính năng của lời nói được nhận thông tin ngữ cảnh có thể biểu diễn như sau:

$$\begin{aligned} h_i^t &= [\overrightarrow{\text{LSTM}}(u_i^t, h_{i-1}^t), \overleftarrow{\text{LSTM}}(u_i^t, h_{i+1}^t)] \\ h_i^a &= W_e^a u_i^a + b_i^a \\ h_i^v &= W_e^v u_i^v + b_i^v \end{aligned} \quad (3.2)$$

trong đó  $u_i^a, u_i^v, u_i^t$  là biểu diễn tính năng thô cảnh của câu thoại  $i$  từ modality âm thanh, hình ảnh và văn bản, tương ứng. Mã hóa modality xuất ra mã hóa tính năng thô cảnh cho các modality  $h_i^a, h_i^v$ , and  $h_i^t$  tương ứng.

## Missing Mask

### Feature Propagation

**Feature Propagation Algorithm.** We can notice that the update in Equation ?? is equivalent to first multiplying the feature vector  $\mathbf{x}$  by the original diffusion matrix  $\tilde{\mathbf{A}}$ , and then resetting the known features to their true value. This gives us Algorithm ??, an extremely simple and scalable iterative algorithm to reconstruct the missing features on a graph, which we refer to as *Feature Propagation* (FP). While  $\mathbf{x}_u$  can be initialized to any value, in practice we initialize  $\mathbf{x}_u$  to zero and find 40 iterations to be enough to provide convergence for all datasets we experimented on. At each iteration, the diffusion occurs from the nodes with known features to the nodes with unknown features as well as among the nodes with unknown features.



**Extension to Vector-Valued Features.** Algorithm ?? extends seamlessly to vector-valued features by simply replacing the feature vector  $\mathbf{x}$  with a  $n \times d$  feature matrix  $\mathbf{X}$ , where  $d$  is the number of features. Multiplying the diffusion matrix  $\mathbf{A}$  by the feature matrix  $\mathbf{X}$  diffuses each feature channel independently. Interestingly, it would not be trivial to extend Equation ?? to vector-valued features without noticing its equivalence with Algorithm ??, as each node could have different missing features, leading to different sub-matrices  $\tilde{\mathbf{A}}_{uk}$  and  $\tilde{\mathbf{A}}_{uu}$  for each feature channel.

**Learning.** One significant advantage of FP is that it can be easily combined with any graph learning model to generate predictions for the downstream task. Moreover, FP is not aimed at merely reconstructing the node features. Instead, by only reconstructing the lower frequency components of the signal, it is by design very well suited to be combined with GNNs, which are known to mainly leverage these lower frequency components [? ]. Our approach is generic and can be used for any graph-related task for missing features, such as node classification, link prediction and graph classification. In this paper, we focus on node classification.

**Oversmoothing.** Figure ?? shows that the more features are missing, the smoother the reconstruction produced by FP is. Despite this, FP does not suffer from oversmoothing [? ], a term used when node representations converge to similar values. Oversmoothing is caused by repeated diffusion and occurs widely when stacking more than a few layers of the most popular GNNs such as GCN [? ], GAT [? ] or SGC [? ]. However, the boundary conditions in the Feature Propagation diffusion equation prevent the reconstructed features from becoming overly smooth, even when using an extremely high number of diffusion steps. This has also been studied by CGNN [? ] and GRAND++ [? ], which require soft boundary conditions in the form of a source term to prevent oversmoothing, although not in the context of missing features.

### 3.3 Graph Attentional Layer

Lorem Ipsum

### 3.4 Phân Lớp

Dựa theo phần \*\*\*, chúng tôi khởi tạo các nút bằng kết hợp của đặc trưng của câu nói và đặc trưng người nói, được biểu diễn là  $h'_i$ .

$$h'_i = [h'^a_i, h'^v_i, h'^t_i]. \quad (3.3)$$

Gọi  $g^a_i$ ,  $g^v_i$  và  $g^t_i$  là các đặc trưng của các mô hình khác nhau được mã hóa bởi GCN. Các đặc trưng tương ứng với cùng một câu thoại được nối lại:

$$g_i = [g^a_i, g^v_i, g^t_i]. \quad (3.4)$$

Chúng ta sau đó có thể nối lại  $g_i$  và  $h_i$  để tạo ra biểu diễn đặc trưng cuối cùng cho mỗi câu nói:

$$e_i = [h'_i, g_i], \quad (3.5)$$

Sau đó,  $e_i$  được cho vào một MLP với các lớp đầy đủ được kết nối để dự đoán nhãn cảm xúc  $\hat{y}_i$  cho câu thoại:

$$\begin{aligned} l_i &= \text{RELU}(W_l e_i + b_l) \\ P_i &= \text{Softmax}(W_{\text{softmax}} l_i + b_{\text{softmax}}) \\ \hat{y}_i &= \arg \min_k (P_i[k]) \end{aligned} \quad (3.6)$$

# Chapter 4

## Experiments and Results

### 4.1 Experiment Setup

#### 4.1.1 Experimental Data

In preparation to evaluate the effectiveness of our model, we proposed experiments on the Baby Care dataset collected from Shopee and Tiki, which has been described previously.

#### 4.1.2 Baseline Methods

We compare various combinations of classifiers and different representation methods to figure out which one is most effective for our situation. The evaluation process is:

(1) Classic models: Logistics Regression, Random Forest, Decision Tree, Naive Bayes, K-nearest Neighbors and Support Vector Machine combine with different representation methods: One-Hot, Chi-Squared, PhoBERT.

(2) Logistics Regression with different representation methods.

(3) General comparison with/without WWL to find out the best model.

#### 4.1.3 Evaluation Metrics

**Precision** Precision is the ratio of correctly predicted positive observations to the total predicted positive observations. High precision relates to the low false positive rate.

$$Precision = \frac{TP}{TP + FP} \quad (4.1)$$

where

$TP$  = True Positive - correctly predicted positive values which means that the value of actual class is 1 and value of predicted class is also 1.

$FP$  = False Positive - actual class is 0 and predicted class is 1.

**Recall** Recall is the ratio of correctly predicted positive observations to the all observations in actual class. High recall is synonymous with the low false negative rate.

$$Recall = \frac{TP}{TP + FN} \quad (4.2)$$

where

$FN$  = False Negative - actual class is 1 and predicted class is 0.

**F1-score** The F1 is a way of combining the precision and recall of the model. The higher the F1 score the better, with 0 being the worst possible, which means the precision or recall is zero; and 1 being the best, indicating perfect precision and recall. F1-score is defined as the harmonic mean of the model's precision and recall.

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \quad (4.3)$$

**Micro-average and Macro-average Performance** Micro Average and Macro Average Performance are used to evaluate multi-label classification model. A macro-average will compute the metric independently for each class and then take the average, whereas a micro-average will aggregate the contributions of all classes to compute the average metric. Micro- and macro-average for Precision, Recall and F1-score is calculated by the formulas as shown in the Equations from 4.4 to 4.7.

$$P_{macro} = \frac{\sum_i P_i}{N} \quad R_{macro} = \frac{\sum_i R_i}{N} \quad F1_{macro} = \frac{\sum_i F1_i}{N} \quad (4.4)$$

$$P_{micro} = \frac{\sum_j TP_j}{\sum_j TP_j + \sum_j FP_j} \quad (4.5)$$

$$R_{micro} = \frac{\sum_j TP_j}{\sum_j TP_j + \sum_j FN_j} \quad (4.6)$$

$$F1_{micro} = \frac{2 * P_{micro} * R_{micro}}{P_{micro} + R_{micro}} \quad (4.7)$$

## 4.2 Experiment Result and Analysis

In context of facing imbalance data (positive outnumbered negative) as analysed in Section 2.3, we appreciate the results of class negative in particular. In this section, we will use Micro/Macro-average metrics calculated for negative class as the measure of evaluation.

### 4.2.1 Classic Models with One-Hot Representation Method.

With the OH data representation method, we use the Mother & Baby data collected from Tiki as the input data for traditional machine learning models. Applying this simple method surprisingly brings fairly good results for such unbalanced data (compared to the methods below) but the negative value of some aspects is not yet assigned.

Based on the results of applying models, we find that the model using the LR method has the most outstanding results. We will use a combination of this method and the data representation methods below to compare results.

Evaluation results are illustrated in table 4.1.

	<i>SVM</i>	<i>DT</i>	<i>KN</i>	<i>LR</i>	<i>NB</i>	<i>RF</i>
<b>micro-p</b>	0,587	0,622	0,541	0,810	0,205	0,818
<b>micro-r</b>	0,228	0,599	0,370	0,525	0,746	0,500
<b>micro-f1</b>	0,329	0,610	0,440	<b>0,637</b>	0,322	0,621
<b>macro-p</b>	0,330	0,422	0,389	0,541	0,183	0,548
<b>macro-r</b>	0,159	0,397	0,258	0,351	0,641	0,329
<b>macro-f1</b>	0,176	0,406	0,299	0,405	0,277	0,395

Table 4.1: Micro/Macro-average for class negative of OH + traditional models

### 4.2.2 Logistics Regression with Different Representation Methods

We use the model using the LR method for comparison since it is the model with the best results as calculated above. Mother & Baby data collected from Tiki are used to evaluated results among three different representation methods: PhoBERT, CS and OH. OH and PhoBERT showed relatively good results, but in general CS has the best performance. Table 4.2 illustrates micro- and macro-average performance for class negative of LR with multiple representation methods.

	<i>PhoBERT</i>	<i>CS</i>	<i>OH</i>
<b>micro-p</b>	0,764	0,704	0,810
<b>micro-r</b>	0,540	0,660	0,525
<b>micro-f1</b>	0,633	<b>0,682</b>	0,637
<b>macro-p</b>	0,506	0,560	0,541
<b>macro-r</b>	0,368	0,532	0,351
<b>macro-f1</b>	0,423	<b>0,545</b>	0,405

Table 4.2: Micro/Macro-average for class negative of LR with representation methods

### 4.2.3 Chi-Squared with/without Word-Window Locating compared with CNN + PhoBERT

	<i>CS + WWL</i>	<i>CS</i>	<i>CNN + PhoBERT</i>
<b>micro-p</b>	0,738	0,704	0,801
<b>micro-r</b>	0,728	0,660	0,747
<b>micro-f1</b>	0,733	0,682	<b>0,773</b>
<b>macro-p</b>	0,781	0,560	0,512
<b>macro-r</b>	0,632	0,532	0,518
<b>macro-f1</b>	<b>0,674</b>	0,545	0,515

Table 4.3: Micro/Macro-average of CS with/without WWL and CNN + PhoBERT

We continue to use the model using LR method to compare between CS, CS + WWL and CNN + PhoBERT.

The results indicate that CS + WWL outperformed CS alone in every figures. In comparison of CS + WWL and CNN + PhoBERT, it is indicated that micro-F1 of CNN + PhoBERT is slightly better with a percentage of 4% but CS + WWL surpassed CNN + PhoBERT with a significant percentage of 16%.

Evaluation proves effective use of the WWL method.

Table 4.3 shows micro- and macro-average performance of CS with/without WWL and CNN + PhoBERT, while table 4.4 shows CS + WWL vs. CNN + PhoBERT detailed statistical results in all aspects.

Method	Class	Price	Service	Safety	Quality	Delivery	Authenticity
<b>CS + WWL</b>	Positive	0.983	0.757	0.892	0.907	0.928	0.960
	Negative	0.571	0.804	0.774	0.718	0.676	0.500
<b>CNN</b>	Positive	0.978	0.875	0.759	0.940	0.943	0.943
	Negative	<b>0.000</b>	0.884	0.632	0.792	0.763	<b>0.000</b>

Table 4.4: WWL + CS vs. CNN + PhoBERT detailed statistics in all aspects

#### 4.2.4 WWL + CS detailed statistics in all domains

Given the proportional difference between the positive and negative classes of different datasets in the domain and the data source in the item (item name), we find that the model uses WWL with the CS method of representing the level data showed quite stable results. The statistical results are described in table 4.5.

	<i>mb_tiki</i>	<i>mb_shopee</i>	<i>tech_tiki</i>	<i>tech_shopee</i>
<b>micro-p</b>	0,738	0,615	0,707	0,709
<b>micro-r</b>	0,728	0,635	0,621	0,629
<b>micro-f1</b>	0,733	0,625	0,661	0,667
<b>macro-p</b>	0,781	0,541	0,582	0,593
<b>macro-r</b>	0,632	0,462	0,510	0,572
<b>macro-f1</b>	0,674	0,482	0,540	0,545

Table 4.5: CS + WWL detailed statistics for class negative in all domains

#### 4.2.5 Error Analysis

- Unbalanced data results affect model training leading to low results for the indicated aspect.
- Lack of summarized data.
- Errors occur in the data annotation phase due to manual process.
- The investigation of the data is not in-depth, yield threshold selected for locating not matching the data.

## 4.3 Demonstrations

### 4.3.1 Data Visualization

We use Kibana and Elasticsearch to analyze and visualize data based on their labels and polarity. Figure 4.1 and 4.2 shows how our dataset are visualized based on labels and keywords.



Figure 4.1: Dataset visualization based on labels and polarity



Figure 4.2: Dataset visualization based on keywords



### 4.3.2 Web Application Demonstration

With a view to bring the model into practical situation, we built a web interface for users to manipulate easily. The users will put in data in two ways: insert each comment by the 'Add' button (figure 4.3) or import a CSV file containing comment data along with the respective aspect (figure 4.4).

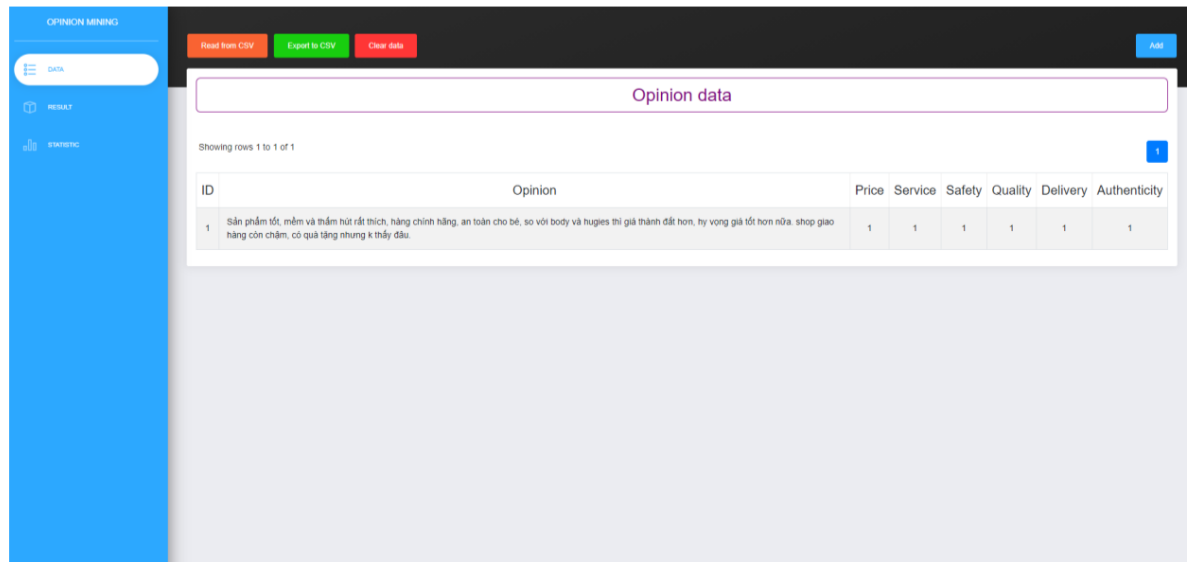


Figure 4.3: Insert comment by the 'Add' button

The model runs when the user selects 'Result' button. The results as illustrated in figure 4.5 and 4.6 show -1 and 1 on the aspects that appear in the comment corresponding to the negative and positive emotions related to that aspect.

In addition, we have built a few functions for the web application such as showing detailed statistical results of positive and negative sentiment on each aspects, exporting results in CSV, PNG, etc. formats for data statistics and visualization.

ID	Opinion	Price	Service	Safety	Quality	Delivery	Authenticity
0	lần thứ 2 nhận quà khuyến mãi thiếu món . lần đầu cho xe 3 bánh con bị thiếu cây sắt gần 2 bánh xe phía sau . lần này cho bộ leo xếp bệnh viện ghi 60 món mà thiếu hết mấy miếng to . quà khuyến mãi gì mà như đánh đổ . cho mà không có tâm gì hết ,...	0	1	0	0	0	0
3	gối bím rất cũ bạc màu và là loại mẫu mã từ rất lâu rồi . gối bím còn bị rách . , không hài lòng	0	0	1	1	0	0
5	sai cung được , thường	0	0	0	0	0	0
6	giao hàng trước dự kiến . hàng tốt . chính hãng . cảm ơn tác , ki hài lòng	0	0	0	1	1	1
7	size m nhưng hơi nhỏ phứt mua lên 1 kích thước . , lòng	0	0	0	0	0	0
8	sữa cho con uống rất tốt , tăng cả chiều cao và cân nặng . tiêu hoá cho bé cũng khá tốt , ki hài lòng	0	0	0	1	0	0
9	giao hàng nhanh , đúng loại như hình ( đc tặng 8m cùng loại ) kích thước s phù hợp với số cân nặng từ 4 - 8 kg , bé nhà mình hơn 5.5 kg mặc vẫn còn rộng , phần đai dán mềm mại , dính tốt . không như loại bobby soft màu vàng , kích thước thì bé , đ	0	0	0	1	1	0
10	mua lần thứ 2 r , bé sử dụng rất êm , cáo bóng dễ thương , ki hài lòng	0	1	0	1	0	0
12	sản phẩm rất tốt . , ki hài lòng	0	0	0	0	0	0
13	bé nhà mình mới 8 ki mà mặc vớ bị lằn đỏ hết hai bên hông , bím khá nhỏ bé mặc không thoải mái , hài lòng	0	0	0	1	0	0

Figure 4.4: Insert comments by importing CSV file

ID	Opinion	Price	Service	Safety	Quality	Delivery	Authenticity
1	Sản phẩm tốt , mềm và thấm hút rất thích , hàng chính hãng , an toàn cho bé , so với body và hugies thì giá thành đắt hơn , hy vọng giá tốt hơn nữa . shop giao hàng còn chậm , có quà tặng nhưng k thấy đâu .	-1	-1	1	1	-1	1

Figure 4.5: Model output when using 'Add' function

OPINION MINING

Export to CSV

### Opinion result

Showing rows 1 to 10 of 269

ID	Opinion	Price	Service	Safety	Quality	Delivery	Authenticity
0	lần thứ 2 nhận quà khuyến mãi thiếu món . lần đầu cho xe 3 bánh con bo thiếu cây sắt gắn 2 bánh xe phía sau . lần này cho bộ lều xếp bên vệ ghè 60 món mà thiếu hết mấy miếng to . quà khuyến mãi gì mà như đánh đổ . cho mà không có tâm gì hết . quà mà thiếu	0	-1	0	0	0	0
3	gói bím rất cũ bạc màu và là loại mẫu mã từ rất lâu rồi . gói bím còn bị rách . không hài lòng	0	0	-1	-1	0	0
6	giao hàng trước dự kiến . hàng tốt . chính hãng . cảm ơn tiki , ki hài lòng	0	0	0	1	1	1
8	sữa cho con uống rất tốt . tăng cả chiều cao và cân nặng . tiêu hoá cho bé cũng khá tốt . ki hài lòng	0	0	0	1	0	0
9	giao hàng nhanh , đúng loại như hình ( đc tăng âm công loa ) kích thước s phù hợp với số cân nặng từ 4 - 8 kg . bé nhà mình hơn 5,5 kg mặc vẫn còn rộng . phần đai dán mềm mại , dính tốt . không như loại babysoft màu vàng , kích thước thì bé . đai dán thì vừa cứng vừa chặt làm bé mình bị hằn hết lên . chẳng...	0	0	0	1	1	0
10	mua lần thứ 2 r . bé sử dụng rất êm . cáo bóng dễ thương . ki hài lòng	0	1	0	1	0	0
13	bé nhà mình mới 8 ki mà mặc vô bị lún đồ hết hai bên hông . bím khá nhỏ bé mặc không thoải mái . hài lòng	0	0	0	-1	0	0
19	đóng gói cẩn thận . hàng tốt r . ki hài lòng	0	0	0	1	1	0
21	tã mỏng hơn tã mình mua bên ngoài . có mùi thơm nhưng hơi nặng trong khi tã mua bên ngoài không có mùi . cảm giác sử dụng tã mua bên ngoài an toàn và dễ chịu hơn khi dùng tã trên tiki . . . . .	0	0	0	-1	0	0
22	bé nhà mình sử dụng tã bobby từ nhỏ . thun co giãn tốt . chống hăm . tiki bán hay giảm giá tại có quà . có thêm gói tiki now giao hàng cực kỳ nhanh . rất thích r . phẩm tốt . quà đẹp . giá tốt	1	1	1	1	1	0

Figure 4.6: Model output when using 'Import' function

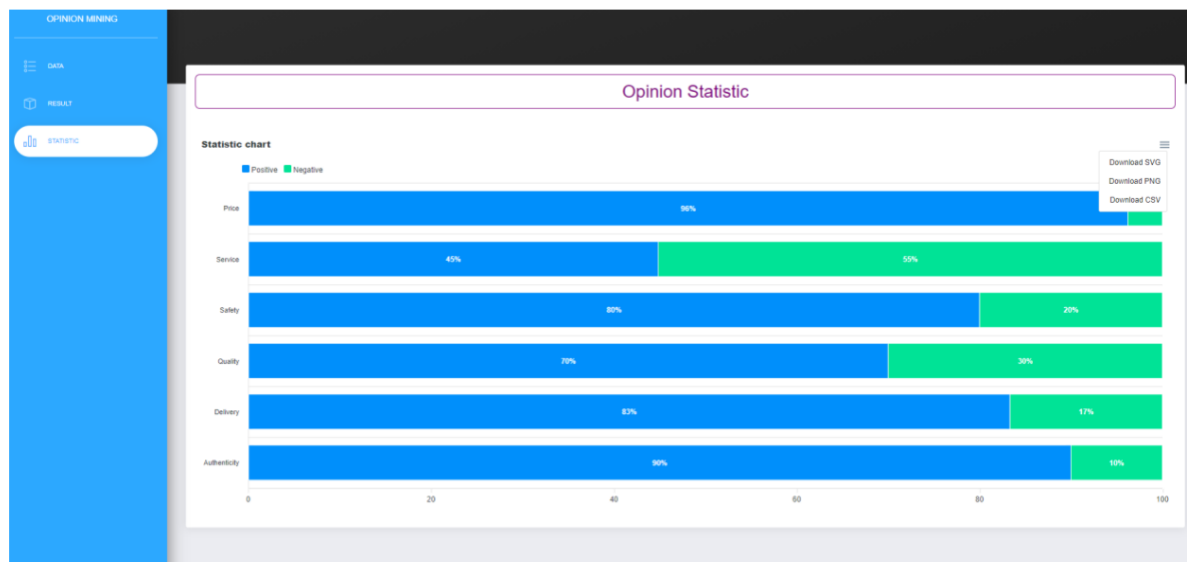


Figure 4.7: Detailed statistical results of positive and negative sentiment on each aspects

# Chapter 5

## Conclusions

### Contributions

In this article, we proposed multiple approaches for the sentiment polarity detection - a sub-task of ABSA problem. Experiments indicated significant result as WWL combine with classic models has reach the maximum percentage of 95% in Micro- and Macro-Average Performance of F1-score.

Through experiments using the locating method and apply multiple data representation techniques, we can determine that using the WWL method with the Chi-Squared data representation brings very prominent results when combined with traditional machine learning methods. Because when classify the sentiment for multi-label data (i.e., multiple aspects in a data sample) using multiple models, we need to find the words associated with the label and the word representing its sentiment. WWL technique step solves this problem. In addition, in order for the data to focus on important words, we use Chi-Squared score to weight word level in the vocab and use it to represent data. This combination brings the most outstanding advantage to address the problem.

### Limitations and Future Work

Selecting threshold for locating words inappropriately may be harmful to the result of the model. Window size (i.e., the number of words around from the center, which equals 3 in the proposed work) affects final result too. This requires manual data revision to be done carefully to choose which one is the best. This is the limitation of our method.

The weak point of proposed work indicates our next path to have in-depth research

in the future:

- (i) More data should be collected to enlarge the dataset.
- (ii) Handling the data imbalance problem.
- (iii) Developing WWL in order to overcome current disadvantage.

# References

- [1] N. Q. Dat and N. T. Anh, “PhoBERT: Pre-trained language models for Vietnamese,” in *Findings of the Association for Computational Linguistics: EMNLP 2020*, 2020, pp. 1037–1042.
- [2] M. Hu and B. Liu, “Mining and summarizing customer reviews,” in *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, 2004, pp. 168–177.
- [3] Q. Jiang, L. Chen, R. Xu, X. Ao, and M. Yang, “A challenge dataset and effective models for aspect-based sentiment analysis,” in *Proceedings of 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.
- [4] B. T. Kieu and S. B. Pham, “Sentiment analysis for vietnamese,” in *2010 Second International Conference on Knowledge and Systems Engineering*. IEEE, 2010, pp. 152–157.
- [5] Y. Ma, H. Peng, T. Khan, E. Cambria, and A. Hussain, “Sentic lstm: a hybrid network for targeted aspect-based sentiment analysis,” *Cognitive Computation*, vol. 10, no. 4, pp. 639–650, 2018.
- [6] L. Mai and B. Le, “Aspect-based sentiment analysis of vietnamese texts with deep learning,” in *Asian Conference on Intelligent Information and Database Systems*. Springer, 2018, pp. 149–158.
- [7] T. H. Nguyen and K. Shirai, “Phrasernn: Phrase recursive neural network for aspect-based sentiment analysis,” in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, 2015, pp. 2509–2514.

- [8] D.-K. Nguyen-Nhat and H.-T. Duong, “One-document training for vietnamese sentiment analysis,” in *International Conference on Computational Data and Social Networks*. Springer, 2019, pp. 189–200.
- [9] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [10] 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.
- [11] P. D. Turney, “Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews,” *arXiv preprint cs/0212032*, 2002.
- [12] T. Van Dang, V. D. Nguyen, N. Van Kiet, and N. L. T. Ngan, “A transformation method for aspect-based sentiment analysis,” *Journal of Computer Science and Cybernetics*, vol. 34, no. 4, pp. 323–333, 2018.
- [13] T. Vu, D. Q. Nguyen, D. Q. Nguyen, M. Dras, and M. 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*. New Orleans, Louisiana: Association for Computational Linguistics, Jun. 2018, pp. 56–60. [Online]. Available: <https://www.aclweb.org/anthology/N18-5012>
- [14] W. Xue and T. Li, “Aspect based sentiment analysis with gated convolutional networks,” *arXiv preprint arXiv:1805.07043*, 2018.