UTILIZING T5 MODEL FOR ASPECT-BASED SENTIMENT ANALYSIS IN VIETNAMESE

MÔ HÌNH T5 CHO BÀI TOÁN TRÍCH XUẤT CẢM XÚC DỰA TRÊN KHÍA CẠNH TRÊN DỮ LIỆU TIẾNG VIỆT

Đặng Bá Quí
Trường Đại học Nông lâm thành phố Hồ Chí Minh

ABSTRACT: *In recent years, with the explosion of user reviews on the Internet, analyzing the sentiment of comments based on specific aspects has become a crucial task in the field of natural language processing (NLP). This allows for a deeper understanding of emotions expressed toward specific aspects in textual sentences. In this paper, we present an approach to Aspect-Based Sentiment Analysis (ABSA) using the T5 (Text-to-Text Transfer Transformer) model, a state-of-the-art framework capable of converting various NLP tasks into text generation tasks. Our approach simplifies the complex processes of traditional ABSA models by directly generating aspect-sentiment pairs in a single step, reducing the need for separate modules for aspect extraction and sentiment classification. We tested this method on the standard dataset of the Vietnamese Language and Speech Processing (VLSP) project. The ViT5 model achieved F1-scores of 84.04% for Aspect Detection (AD) and 72.29% for Aspect Polarity (AP) in the restaurant domain, and 62.68% and 66.34% in the hotel domain, respectively. This research demonstrates the potential of the ViT5 model in context-aware sentiment analysis for the Vietnamese language.*

Keywords: *Aspect-based sentiment analysis, T5, BERT, SVM.*

TÓM TẮT: *Trong những năm gần đây với việc bùng nổ những đánh giá của người dùng trên Internet, việc phân tích cảm xúc của những bình luận dựa trên những khía cạnh đang trở thành một nhiệm vụ quan trọng trong lĩnh vực xử lý ngôn ngữ tự nhiên, cho phép hiểu rõ hơn về những cảm xúc được thể hiện đối với các khía cạnh cụ thể trong câu văn bản. Trong bài này chúng tôi sẽ trình bày một phương pháp tiếp cận bài toán phân tích cảm xúc dựa trên khía cạnh (Aspect-Based Sentiment Analysis) bằng cách sử dụng mô hình T5 (Text-to-Text Transfer Transformer), một SOTA (state-of-the-art) framework có khả năng chuyển đổi các nhiệm vụ NLP khác nhau thành nhiệm vụ tạo văn bản. Cách tiếp cận của chúng tôi đơn giản hóa quy trình phức tạp của các mô hình ABSA truyền thống bằng cách tạo trực tiếp các cặp khía cạnh và cảm xúc trong một bước duy nhất, giảm thiểu nhu cầu cho các mô-đun truy xuất khía cạnh và phân loại cảm xúc riêng biệt. Chúng tôi thử nghiệm trên bộ dữ tiêu chuẩn của Vietnamese Language and Speech Processing (VLSP) dựa trên mô hình ViT5 đạt được F1-scores 84.04% trong nhiệm vụ Aspect Dection (AD) và 72.29% trong nhiệm vụ Aspect Polarity (AP) trên miền nhà hàng, 62.68% và 66.34% trên miền khách sạn*[1].

Từ khóa: *Aspect-based sentiment analysis, T5, BERT, SVM.*

1. GIỚI THIỆU

Phân tích cảm xúc dựa trên khía cạnh (ABSA) là một nhiệm vụ quan trọng trong xử lý ngôn ngữ tự nhiên (NLP), đặc biệt trong bối cảnh thương mại điện tử, nơi khách hàng thường chia sẻ ý kiến về các

[1] Thử nghiệm và kết quả có sẵn tại: https://github.com/dbqui1706/ABSA-S2S

khía cạnh khác nhau của sản phẩm và dịch vụ. Sự gia tăng nhanh chóng của các bình luận trực tuyến đã biến việc khai thác ý kiến thành một lĩnh vực nghiên cứu sôi động. Các đánh giá này không chỉ là nguồn thông tin quý giá về chất lượng sản phẩm và dịch vụ mà còn là thước đo quan trọng cho cả doanh nghiệp và người tiêu dùng. Việc phân tích tự động các đánh giá này để trích xuất thông tin hữu ích là rất cần thiết. Thông tin này có thể được sử dụng để cải thiện sản phẩm, nâng cao trải nghiệm khách hàng, và hỗ trợ các hoạt động nghiên cứu thị trường và khảo sát. Do đó, phát triển các kỹ thuật ABSA hiệu quả là một yếu tố quan trọng trong việc tận dụng tối đa giá trị từ dữ liệu đánh giá trực tuyến.

Nhiệm vụ ABSA nhằm mục đích rút ra sự phân cực về tình cảm (ví dụ: positive, negative và neutral) đối với một mục tiêu, đó là khía cạnh của một thực thể cụ thể. Trong ABSA có thể được chia làm ba nhiệm con: Phát hiện khía cạnh (Aspect Detection), biểu hiện mục tiêu ý kiến (Opinion Target Expression), phân cực khía cạnh (Aspect Polarity). Tuy nhiên, bài báo này tập trung vào hai nhiệm vụ AD và AP trên bộ dữ liệu VLSP 2018. Nhiệm vụ AD liên quan đến việc xác định các khía cạnh từ một tập đã định nghĩa trước, trong khi nhiệm vụ AP tập trung vào việc xác định phân cực cảm xúc của các khía cạnh đó. Ví dụ, cho một câu đánh giá "**Về thức ăn và giá cả: Mình ăn phần ăn 279k nên được ăn khá nhiều các món, khá đa dạng, món nướng ướp ngon, đậm đà vừa phải.**" với nhiệm vụ AD sẽ có output là **"Food#Prices"** (Thức ăn#Giá cả), **"Food#Quality"** (Thức ăn#Chất lượng), và **"Food#Style&Options"** (Thức ăn#Cách&Tùy chọn); trong khi đó nhiệm vụ AP, phân loại cảm xúc cho **"Food#Frice"** là **neutral**, **"Food#Quality"** là **positive** và **"Food#Style&Options"** là **positive**. Từ ví dụ trên dùng chúng tôi thấy rằng những việc khai thác thông tin từ những bình luận của người dùng đóng vai trò quan trọng trong việc phát triển của các tổ chức kinh doanh.

Bài báo này trình bày việc sử dụng mô hình ViT5 [1] để giải quyết hai nhiệm vụ chính trong phân tích cảm xúc dựa trên khía cạnh (ABSA) trên dữ liệu tiếng Việt. Nghiên cứu đánh giá mô hình trên bộ dữ liệu VLSP 2018 và so sánh với các phương pháp khác. Kết quả cho thấy phương pháp đề xuất vượt trội hơn về hiệu suất so với các phương pháp khác được thử nghiệm.

2. CÁC CÔNG TRÌNH LIÊN QUAN

Aspect-based sentiment analysis (ABSA) là nhiệm vụ đã thu hút được rất nhiều sự quan tâm của công đồng nghiên cứu từ rất lâu. Một vài workshop như SemEval2014 [2], SemEval2015 [3], SemEval2016 [4] và VSLP2018 [5], đã tổ chức giải này ra nhằm tìm được những các giải quyết vấn đề cho bài toán của họ một cách tốt nhất, và đã có rất nhiều phương pháp đã giải quyết được bài toán. Do bài toán bao gồm nhiệm vụ con là AD (Aspect dection) và SP (Sentiment Polarity), [6] đã mô tả một hệ thống dùng máy học để trích xuất các khía cạnh từ các bài đánh giá khách hàng. Hệ thống này có thể trích xuất cả các khía cạnh rõ ràng (explicit) và ngầm (implicit), sử dụng Suppport Vector Machine (SVM) one-vs-rest, kết hợp với một pipeline tiền xử lý văn bản tiên tiến. Nó sử dụng các đặc trưng mới như vector trung bình (mean embedding) để cải thiện đáng kể độ chính xác của bộ phân loại SVM trên bộ dữ liệu của SemEval2016. sử dụng mô hình

Bảng 1. Một ví dụ về đầu vào và đầu ra trong tập dữ liệu

Domain	Input	Output
Restaurant	Mấy lần trước thì sốt sền sệt, lần này ko hiểu sao sốt lỏng le, ăn ko thích bằng nhưng được cái vị vẫn ngon.	{Food#Quality, positive} {Food#Style&Options, negative}
Hotel	Khách Sạn TTC Hotel Premium Ngọc Lan Ngọc Lan Đà Lạt ngày trước tôi có thường đặt phòng ở đây, theo tôi thấy thì nhân viên của khách sạn phục vụ tốt, phòng ở rộng và thoáng.	{Hotel#General, positive} {Service#General, positive} {Room#Comfort, positive} {Room#Design&Features, positive}

AT-LSTM [7] làm baseline, mô hình LSTM với cơ chế attention để có thể tập trung vào các từ quan trọng hơn liên quan đến từng khía cạnh (aspect) cần phán đoán, kết hợp thêm thông tin từ các từ điển sentiment (lexicon) vào mô hình, để tăng độ linh hoạt và sức mạnh của mô hình đồng thời họ cũng thêm regularizer cho attention vector để tránh tập trung quá vào một vài từ nhất định mà phân bổ sự chú ý rộng hơn sang các từ khác cũng có thông tin quan trọng.

Trong tiếng Việt có một số bộ dữ liệu về phân tích tình cảm trong nhiều lĩnh vực. Các tác giả [8] đã sử dụng PhoBERT [9] làm mô hình ngôn ngữ được đào tạo trước cho tiếng Việt theo hai cách: Đa tác vụ (Multitask) và Đa tác vụ với cách tiếp cận đa nhánh (Multitask with Multi-branch approach), cả hai phương pháp đều cho kết quả tốt cho tác vụ AD và SP. Đặc biệt với phương pháp Multitask thì mô hình đạt được kết quả tốt nhất khi ghép bốn lớp cuối cùng của BERT [10] lại với nhau. Với bộ dữ liệu VSLP 2018 ABSA trên Hotel domain họ đạt được kết quả cho F1-score là 82.55% trên nhiệm vụ AD và 77.32% cho nhiệm vụ SP, với miền Restaurant domain F1-score 83.29% và 71.55%. Ngoài ra còn [11] đề xuất phương pháp biến đổi (Transformation method) để giải quyết bài toán, họ tiếp cận bài toán như bài bài toán phân loại nhiều nhãn (Multi-label classification) và áp dụng phương pháp biến đổi để chuyển nó thành nhiều bài toán phân loại nhị phân (binary classification), để huấn luyện các bộ phân loại nhị phân, họ trích xuất các đặc trưng khác nhau từ các bài đánh giá và sử dụng tuyến tính SVM (Linear SVM) để phát hiện các khía cạnh và xác định cảm xúc của chúng. Qua phương pháp này họ đã đạt được F1-scores trên nhiệm vụ AD 77% và 69% trên hai bộ dữ liệu Restaurant domain và Hotel domain VLSP.

3. PHÁT BIỂU BÀI TOÁN

Đối với thử thách từ ABSA, chúng tôi mô tả bài toán như sau: Cho một câu văn bản $\boldsymbol{s}$ và một tập đầu ra $\{a, p\}_{(t=1)}^{T}$ với $\boldsymbol{p}$ là polarity (positive, neutral, negative) của một aspect $\boldsymbol{a}$. Aspect được xác định là một thực thể và một thuộc tính. Trong một câu văn bản có thể một hoặc nhiều hơn một tập **T**. Dưới đây là một ví dụ cho một câu văn bản trong bộ dữ liệu trên cả hai miền nhà hàng và khách sạn được minh họa trong Bảng 1.

4. THỰC NGHIỆM

4.1. Dữ liệu

Trong nghiên cứu này, chúng tôi đã đánh giá hiệu suất các mô hình trên bộ dữ liệu VSLP 2018 thu thập từ các trang thương mại điện tử và được chú thích thủ công. Dữ liệu được chia ra ba tập: Training, development và testing, bao gồm 12 và 34 khía cạnh trên 2 miền nhà hàng và khách sạn. Bảng 2 sẽ trình bày thống kê tóm chi tiết bộ dữ liệu.

Preprocessing: Do dữ liệu của VLSP được thu thập từ các trang thương mại điện tử nên dữ liệu chưa được làm sạch, dẫn đến một số vấn đề như: lỗi chính tả, lỗi khoảng cách, ký tự đặc biệt, hashtag, URL, từ viết tắt, và nhiều vấn đề khác. Những yếu tố này có thể gây khó khăn trong việc xử lý và phân tích dữ liệu. Vì vậy, cần phải tiến hành các bước tiền xử lý để làm sạch và chuẩn hóa dữ liệu trước khi áp dụng các mô hình. Các bước tiền xử lý bao gồm:

- **Bước 1**: HTML tags và những icon được loại bỏ khỏi câu.
- **Bước 2**: Chuẩn hóa bộ ký tự từ Windows-1252 sang UTF-8 và chuẩn hóa cách gõ tiếng Việt như chuẩn hóa kiểu gõ, chính tả và dấu câu.
- **Bước 3**: Các chữ viết tắt tiếng Việt thông dụng được chuẩn hóa cho từng miền trong tập dữ liệu như "khong", "ko", "khg", v.v., vì vậy những từ này sẽ được thay thế bằng từ được sử dụng thường xuyên nhất như là "không".
- **Bước 4**: Elongations (các từ có sự lặp lại của các chữ cái) đã được chuẩn hóa thành dạng từ thực sự của chúng (e.g., "ngooon" thì được chuẩn hóa thành "ngon").

Bước 5: Trong tiếng Việt, một từ có thể được tạo thành từ hai hoặc nhiều từ khác (từ ghép), chẳng hạn như "nhà hàng" (mỗi từ có nghĩa riêng khi đứng một mình và có nghĩa khác khi kết hợp). Vì lý do này, cần phải phân đoạn văn bản và ghép chúng lại thành dạng như "nhà_hàng". Trong bài này, chúng tôi sử dụng Vietnamese Core NLP Toolkit Pyvi[2] để phân đoạn từ.

Bảng 2. Tóm tắt thống kê cho bộ dữ liệu nhà hàng VSLP sau khi chúng tôi tiền xử lý dữ liệu. **N** là số lượng đánh giá, l biểu thị độ dài trung bình của đánh giá, **|V|** là kích thước của từ điển, **I** là số lượng từ vựng không nằm trong tập huấn luyện, **A** biểu thị số lượng aspect, (+, o, -) tương ứng với số lượng (positive, neutral, negative).

Domain	**Dataset**	**N**	***l***	**|V|**	**I**	**A**	**+**	**o**	**-**
Restaurant	Train	7028	71	6046	-	9458	5179	2533	1746
	Test	1938	71	3326	696	2629	1503	691	435
	Dev	771	71	2087	277	1053	579	278	196
Hotel	Train	7180	80	2911	-	11721	8964	654	2103
	Test	2030	80	1750	355	3260	2462	181	617
	Dev	795	80	1117	124	1309	997	78	234

[2] https://github.com/trungtv/pyvi

4.2. Phương pháp

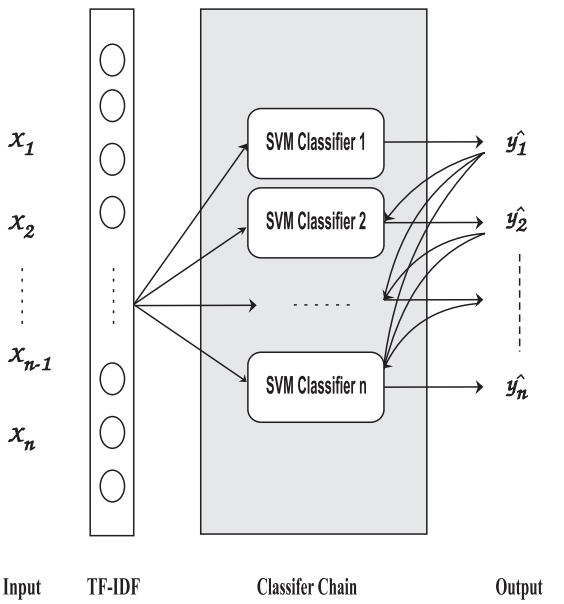
4.2.1. Multitask classification

Trong phần này, chúng tôi sẽ mô tả multi-task nhằm giải quyết hai nhiệm vụ con: AD và AP trong ABSA. Đã có nhiều nghiên cứu trước đây đề xuất các phương pháp giải quyết đồng thời cả hai nhiệm vụ này [12]. Chúng tôi dựa trên những ý tưởng này để chuyển đổi bài toán thành một bài toán phân loại nhiều nhãn, nhằm xử lý hiệu quả cả hai nhiệm vụ AD và AP trong cùng một mô hình.

Họ mô hình hóa đầu ra dưới dạng một danh sách các vector $\boldsymbol{A}$ trong đó $\boldsymbol{A}$ là số lượng khía cạnh của dữ liệu nhà hàng và khách sạn, với tổng cộng 12 và 34 khía cạnh. Mỗi vector v_i (với *i* từ 1 đến 12 hoặc 34) có bốn phần tử, được ký hiệu là $v_i = [v_{i1}, v_{i2}, v_{i3}, v_{i4}]$ trong đó chỉ có một phần tử được gán giá trị 1 và các phần tử còn lại được gán giá trị 0. Bốn trạng thái của vector được giải mã thành các giá trị: None, Positive, Neutral hoặc Negative cho một khía cạnh cụ thể. "None" chỉ ra rằng đầu vào không chứa khía cạnh đó. Như vậy, với mỗi aspect *i* vector v_i sẽ chỉ ra trạng thái cảm xúc tương ứng của khía cạnh đó trong dữ liệu đầu vào. Ví dụ, nếu $v_1 = [0, 1, 0, 0]$, điều đó có nghĩa là khía cạnh đầu tiên có cảm xúc positive. Nếu $v_2 = [1, 0, 0, 0]$, điều đó có nghĩa là khía cạnh thứ hai không xuất hiện trong đầu vào.

4.2.2. Phân loại đa tác vụ với SVM

Chúng tôi áp dụng hai mô hình phân loại đa nhiệm là bộ phân loại đa đầu ra (Multioutput Classifier) và chuỗi phân loại (Classifier Chain) từ thư viện Sklearn[3] để giải quyết bài toán phân loại đa nhiệm. Quy trình thực hiện bao gồm các bước sau: tiền xử lý dữ liệu (**Preprocessing**), xây dựng mô hình, huấn luyện mô hình, và cuối cùng là đánh giá mô hình trên tập dữ liệu kiểm tra. Hình 1 sẽ



Hình 1. Kết hợp mô hình SVM và Classifier Chain cho phương pháp phân loại đa mục tiêu

mô hình hóa toàn bộ quy trình của chúng tôi.

Trong quá trình xây dựng mô hình, chúng tôi sử dụng Linear SVM làm mô hình cơ sở. Để trích xuất đặc trưng cho bộ phân loại, chúng tôi áp dụng phương pháp TF-IDF với N-gram, bao gồm unigram, bigram và trigram. Phương pháp này giúp nắm bắt được ngữ cảnh và quan hệ giữa các từ trong văn bản. Chúng tôi sử dụng các tham số mặc định của tuyến tính SVM trong thư viện Sklearn.

4.2.3. Phân loại đa tác vụ với mô hình PhoBERT

Chúng tôi hiện thực lại phương pháp Multi-task với $\text{PhoBERT}_{\text{base}}$ bằng cách kết hợp từ bốn lớp cuối cùng của mô mình BERT lại với nhau từ công trình [8]. Kỹ thuật này đã được chứng minh bài báo báo gốc của

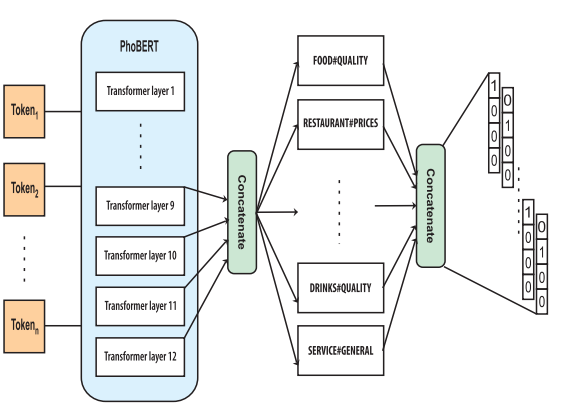
[3] https://scikit-learn.org/stable/modules/multiclass

BERT và mang lại kết quả tốt nhất trong các thử nghiệm của họ ở Bảng 3.

Bảng 3. Kết quả với các phương pháp tiếp cận dựa trên tính năng khác nhau của BERT

Feature-based approach $BERT_{base}$	**Dev F1**
Embeddings	91.0
Second-to-Last Hidden	95.6
Last Hidden	94.9
Weighted Sum Last Four Hidden	95.9
Concat Last Four Hidden	**96.1**
Weighted Sum All 12 Layers	95.5

Để tinh chỉnh được mô hình PhoBERT, chúng tôi thiết kế mô hình multitask để xử lý đồng thời hai nhiệm vụ AD và AP được minh họa ở Hình 2. Đầu vào là một câu văn bản được mã hóa thành một vector $x_i \in R^d$ với d là độ dài của một câu và $x_i \in X$, $X \in R^{l \times d}$ với l là số input có trong tập dữ liệu. Sau đó đưa qua mô hình PhoBERT và lấy ra 4 trạng thái ẩn cuối cùng $H = [h_{-4}, h_{-3}, h_{-2}, h_{-1}]$. Tại đây chúng tôi nối các trạng thái ẩn để tạo ra vector embedding kết hợp H_{pooled} và được truyền qua một lớp kết nối đầy đủ được tạo bằng cách ghép các lớp Dense A tương ứng với one-hot-vector A. Vì vậy chúng tôi sẽ thu



Hình 2. PhoBERT bốn lớp cuối cùng với Multi-task cho miền nhà hàng

được một lớp dense có 48 và 134 neuron cho tập dữ liệu nhà hàng (12×4) và (34×4) cho dữ liệu khách sạn, gọi là tính năng được tạo (g) và để có thể tính score cho giá trị dự đoán $\hat{y}^{(a)}$ của mô hình cho mỗi aspect $a \in A$ với hàm softmax:

$$\hat{y}^{(a)} = softmax\left(W^{(a)} \cdot g + b^{(a)}\right) \quad (1)$$

Vì vậy chúng tôi có thể dự đoán một aspect a cùng với polarity của nó chỉ trong một bước duy nhất:

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

Chúng tôi lấy Cross Entropy làm hàm Loss và nó được tính tổng trên tất cả các khía cạnh ở A.

$$Loss(y, \hat{y}) = -\sum_{a \in A}\sum_i y_i^{(a)} \cdot log\left(\hat{y}_i^{(a)}\right) \quad (3)$$

Trong quá trình huấn luyện chúng tôi sử dụng mô hình pre-trained $PhoBERT_{base}$ với batch size 32 và 16 cho quá trình xác thực. Độ dài cố định của input được đặt là 256 và số epoch cho quá trình huấn luyện là 10. Thuật toán tối ưu hóa AdamW [13] được sử dụng với learning rate khởi tạo là 1e-5.

4.2.4. ViT5

Chúng tôi xem bài toán ABSA như một bài toán tạo văn bản (text-generation) để có thể giải quyết đồng thời cả hai vấn đề AD và AP. Chiến lược của chúng tôi là từ một câu input, áp dụng mô hình ViT5 để tạo ra một câu văn bản khác phản ánh kết quả mong muốn. Ý tưởng này bắt nguồn từ công trình [14] họ tạo ra một khung hợp nhất (Unified Framework) chuyển đổi mọi vấn đề ngôn ngữ sang định dạng văn bản như tóm tắt văn bản, trả lời câu hỏi, phân loại văn bản, v.v. Chúng tôi muốn tạo ra một câu output có chứa tất cả các aspect và polarity tướng ứng $Y =$

$\{y_1, y_1, \dots, y_m\}$ với $Y \in R^{1\times m}$ và m là độ dài chưa được xác định, từ một câu input X có n từ với $X = \{x_1, x_2, \dots, x_n\}$. Để có thể xây dựng được mô hình Seq2Seq chúng tôi cần chuyển những nhãn mục tiêu thành một câu có thể hiểu theo ngôn ngữ tự nhiên. Vì nhãn của dữ liệu của chúng tôi đang có định dạng (e.g. Food#Quality, positive) vì vậy chúng tôi sẽ chuyển những output sang một câu hoàn chỉnh.

Bảng 4. Bảng thể hiện các danh mục aspect khác nhau cho dữ liệu nhà hàng

Enity#Attribute	**Output Form**	**Enity#Attribute**	**Output Form**
Restaurant#General	Nhà hàng nói chung	Drinks#Quality	Chất lượng đồ uống
Restaurant#Prices	Giá tiền nhà hàng	Drinks#Prices	Giá tiền đồ uống
Restaurant#Miscellaneous	Vấn đề khác	Drinks#Style&Options	Lựa chọn đồ uống
Food#Quality	Chất lượng đồ ăn	Location#General	Địa chỉ
Food#Prices	Giá tiền đồ ăn	Ambience#General	Không gian
Food#Style#Options	Lựa chọn đồ ăn	Service#General	Phục vụ

Điền vào chỗ trống (Slot-filling Teamplate): Chúng tôi sẽ tạo một câu hoàn chỉnh bằng cách sử dụng định dạng Output Teamplate Slot-filling (**SFT**) để lấy nhãn cho các aspect và polarity tương ứng. Ví dụ cho một câu "*Tuy giá hơi cao xíu nhưng chất lượng thì xứng đáng vô cùng.*" và chúng tôi sẽ thu được câu "*Đánh giá thể hiện ý kiến [negative] cho [Restaurant#Prices], [positive] cho [Restaurant #General]* ".

Mô hình Seq2Seq hợp nhất (US2S): Chúng tôi cũng sử dụng ý tưởng của các tác giả trong công trình [14] họ ánh xạ những aspect và polarity sang một từ điển. Họ xây dựng từ điển bằng cách thống kê bằng cách áp dụng kỹ thuật TF-IDF trích xuất những mẫu trên aspect trên tập training set và development set. *Bảng 4* là cách họ biểu diễn các aspect trên dữ liệu nhà hàng. Các polarity "positive", "neutral" và "negative" lần lượt được ánh xạ thành "tốt", "tạm" và "tệ". Phương pháp này cho phép mô hình tạo ra các câu văn bản tự nhiên chứa thông tin về khía cạnh và phân cực cảm xúc, thay vì chỉ các nhãn đơn giản. Ví dụ "Food#Quality, positive" thì sẽ được chuyển đổi thành "Chất lượng đồ ăn tốt" nếu có nhiều hơn một cặp aspect và polarity thì họ sẽ nối lại bằng một từ nối "và".

Trong bài này chúng tôi sẽ sử dụng lại kiến trúc ViT5. Một đầu vào chuỗi X được đưa vào encoder để thu được hidden states:

$$h^{encoder} = Encoder(X). \quad (1)$$

Tại mỗi bước thứ i của encoder mã thông báo được tạo y_i được tính toán dựa trên các hidden states và đầu ra $y_{0:i-1}$ trước đó để mang lại biểu diễn:

$$h_i^{encoder} = Decoder\left(h^{encoder}, y_{0:i-1}\right). \quad (2)$$

Đầu ra của decoder được áp dụng vào hàm softmax để tính xác suất phân bổ trên toàn bộ từ vựng cho mã thông báo tiếp theo. Hàm loss cho input-target được xây dựng:

$$Loss = \sum_{i}^{m} log\, p_\theta\,(y_i|h^{encoder}, y_{0:i-1}) \quad (3)$$

Với p_θ là phân phối xác xuất của token kế tiếp y_i, θ được khởi tạo với các trọng số tham số được huấn luyện trước đó và được tinh chỉnh trong quá trình huấn luyện.

Chúng tôi sử dụng mô hình $ViT5_{base}$ Để kiểm tra hiệu suất của mô hình. Do ViT5[4] có hai phiên bản được tung ra cho cộng đồng nghiên cứu là $ViT5_{base}$ và $ViT5_{large}$. Chúng tôi sử dụng độ dài dài nhất của input là 140 và output là 30 trên miền dữ liệu nhà hàng, 274 và 46 trên miền dữ liệu khách sạn. Batch size mà chúng tôi thử nghiệm trên tập training set 16 và 8 cho tập development set với epoch là 5. Mô hình được tối ưu hóa bằng AdamW optimizer với learning rate khởi tạo là 2e-5.

5. KẾT QUẢ THỰC NGHIỆM

Trong phần này, chúng tôi trình bày các kết quả của thí nghiệm phân tích bình luận dựa trên các mô hình khác nhau bao gồm SVM, PhoBERT, ViT5-SFT và ViT5-US2S trên hai miền dữ liệu là nhà hàng và khách sạn từ tập dữ liệu VSLP 2018 ABSA.

Bảng 5. Kết quả thử nghiệm trên miền nhà hàng

VSLP 2018 ABSA - Restaurant				
Task	**Model**	**Precssion**	**Recall**	**F1-score**
AD	SVM	80.50	63.56	69.14
	PhoBERT	**89.23**	79.39	83.86
	ViT5-SFT	85.71	82.60	83.98
	ViT5-US2S	84.91	**83.34**	**84.04**
AP	SVM	69.07	52.28	55.86
	PhoBERT	**77.58**	59.94	61.29
	ViT5-SFT	72.17	69.48	70.17
	ViT5-US2S	73.59	**71.68**	**72.30**

Bảng 5 cho thấy kết quả của nhiệm vụ AD và AP trên miền nhà hàng. Đối với nhiệm vụ AP, mô hình PhoBERT có precision cao nhất là 89.23%, trong khi mô hình ViT5-US2S có recall cao nhất là 83.34% và F1-score là 84.04%. Điều này cho thấy PhoBERT vượt trội trong việc xác định đúng các khía cạnh nhưng ViT5-US2S có khả năng phát hiện đầy đủ hơn các khía cạnh trong các bình luận. Đối với nhiệm vụ AP, PhoBERT vẫn duy trì precssion cao nhất là 77.58%, trong khi ViT5-US2S có recall và F1-score cao nhất, lần lượt là 71.68% và 72.30%. Sự khác biệt này gợi ý rằng mặc dù PhoBERT có khả năng xác định đúng các aspect và sentiment, ViT5-US2S lại có hiệu quả cao hơn trong việc phát hiện đầy đủ các khía cạnh kết hợp với cảm xúc trong bình luận.

Bảng 6. Kết quả thử nghiệm trên miền khách sạn

VSLP 2018 ABSA - Hotel				
Task	**Model**	**Precssion**	**Recall**	**F1-score**
AD	SVM	**86.58**	36.70	44.03
	PhoBERT	74.59	60.36	**65.39**
	ViT5-SFT	71.30	55.69	58.04
	ViT5-US2S	65.57	**61.20**	62.68
AP	SVM	88.47	43.02	46.29
	PhoBERT	**88.51**	46.31	48.10
	ViT5-SFT	71.67	56.55	56.15
	ViT5-US2S	76.05	**65.50**	**66.34**

Bảng 6 trình bày kết quả cho thấy kết quả của nhiệm vụ AD và AP trên miền khách sạn. Đối với nhiệm vụ AD, SVM đạt độ chính xác cao nhất là 86.58%, tuy nhiên recall chỉ đạt

[4] https://github.com/vietai/ViT5

36.70% và F1-score là 44.03%. Trong khi đó, PhoBERT mặc dù có độ chính xác thấp hơn (74.59%) nhưng lại đạt F1-score cao nhất là 65.39%. Mô hình ViT5-US2S có recall cao nhất là 61.20%, cho thấy khả năng phát hiện khía cạnh tốt hơn. Đối với nhiệm vụ AP, PhoBERT có precission cao nhất là 88.51%, trong khi ViT5-US2S tiếp tục đặt recall và F1-score cao nhất, lần lượt là 65.50% và 66.34%.

Kết quả trên hai miền dữ liệu nhà hàng và khách sạn cho thấy các mô hình học sâu như PhoBERT và ViT5 có hiệu quả cao hơn so với các mô hình truyền thống như SVM. Đặc biệt, mô hình ViT5-US2S thể hiện sự vượt trội trong việc phát hiện đầy đủ các khía cạnh và cực tính trong bình luận, qua đó ta thấy được việc chuẩn hóa dữ liệu đầu ra ảnh hưởng rất nhiều đến hiệu suất của mô hình. Với nhiệm vụ AD và AP thì ViT5-S2S có precission, recall, f1-score hơn ViT5-SFT lần lượt ($\pm$0.8%, $\pm$0.74%, $\pm$0.06%), ($\pm$1.42%, $\pm$2.2%, $\pm$2.13%) và ($\pm$5.73%, $\pm$5.51%, $\pm$4.64%), ($\pm$4.38%, $\pm$8.95%, $\pm$10.19%) trên hai miền dữ liệu tương ứng là nhà hàng và khách sạn. Điều này cho thấy tiềm năng ứng dụng của các mô hình học sâu trong lĩnh vực phân tích ngôn ngữ tự nhiên và phân tích ý kiến khách hàng.

6. KẾT LUẬN

Trong bài này chúng tôi đã trình bày các phương pháp để giải quyết vấn đề của VSLP cho bài toán ABSA bằng các phương pháp học máy truyền thống như SVM và sử dụng các mô hình pretrained như PhoBERT, ViT5. Qua đó ta thấy được rằng mô hình ViT5 cho được hiệu suất tốt nhất cho cả hai nhiệm vụ AD và AP, nhưng để cho mô hình T5 đạt hiệu suất tốt thì cũng phụ thuộc vào cách mà chúng tôi chuẩn hóa output. ViT5-US2S với output được chuẩn hóa theo phương pháp của tác giả [14] thì cho hiệu suất tốt hơn khi chúng tôi tự chuẩn bị output bằng phương pháp SFT. Tuy vậy, chúng tôi cũng phải đánh đổi về thời gian và chi phí tính toán cho mô hình ViT5. Trong tương lai, chúng tôi dự định thử nghiệm các kỹ thuật như Instruction và fine-tuning với các mô hình hiện đại như GPT. Large Language Model (LLM) hiện đang là một chủ đề nổi bật, và chúng tôi sẽ tiến hành khảo sát bằng các kỹ thuật few-shot và zero-shot để đánh giá hiệu suất của các mô hình này.

TÀI LIỆU THAM KHẢO

[1] Phan, Long and Tran, Hieu and Nguyen, Hieu and Trinh, Trieu H. (2022), "ViT5: Pretrained Text-to-Text Transformer for Vietnamese Language Generation", in *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Student Research Workshop*, Association for Computational Linguistics, 2022, pp. 136-142.

[2] Pontiki, Maria and Pontiki, Maria and Galanis, Dimitris and Pavlopoulos, John and Papageorgiou, Harris and Androutsopoulos, Ion and Manandhar, Suresh, (2024), "Aspect based sentiment analysis semeval-2014 task 4", *Asian Journal of Computer Science and Information Technology (AJCSIT) Vol,* vol. 4, no. Citeseer, pp. 27-35.

[3] Pontiki, Maria and Galanis, Dimitris and Papageorgiou, Haris and Manandhar,

Suresh and Androutsopoulos, Ion, (2015), "Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)", pp. 486-495.

[4] Pontiki, Maria and Galanis, Dimitris and Papageorgiou, Haris and Androutsopoulos, Ion and Manandhar, Suresh and Al-Smadi, Mohammed and Al-Ayyoub, Mahmoud and Zhao, Yanyan and Qin, Bing and De Clercq, Orph{\'e}e and others, (2016), "Semeval-2016 task 5: Aspect based sentiment analysis", pp. 19-30.

[5] Nguyen, Huyen and Nguyen, Hung and Ngo, Quyen and Vu, Luong and Mai, Vu and Xuan Bach, Ngo and Le, Cuong, (2019), "VLSP SHARED TASK: SENTIMENT ANALYSIS", *Journal of Computer Science and Cybernetics,* vol. 34, pp. 295-310.

[6] Jihan, Nadheesh and Senarath, Yasas and Tennekoon, Dulanjaya and Wickramarathne, Mithila and Ranathunga, Surangika, (2017), "Multi-domain aspect extraction using support vector machines", in *Proceedings of the 29th conference on computational linguistics and speech processing (ROCLING 2017)*, pp. 308-322.

[7] Wang, Yequan and Huang, Minlie and Zhu, Xiaoyan and Zhao, Li, (2016), "Attention-based {LSTM} for Aspect-level Sentiment Classification", *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing,* pp. 606-615.

[8] Dang, Hoang-Quan and Nguyen, Duc-Duy-Anh and Do, Trong-Hop, (2022), "Multi-task Solution for Aspect Category Sentiment Analysis on Vietnamese Datasets", in *2022 IEEE International Conference on Cybernetics and Computational Intelligence (CyberneticsCom)*, pp. 404-409.

[9] Nguyen, Dat Quoc and Tuan Nguyen, Anh, (2020), "{P}ho{BERT}: Pre-trained language models for {V}ietnamese", in *Findings of the Association for Computational Linguistics: EMNLP 2020*, Association for Computational Linguistics, pp. 1037-1042.

[10] Devlin, Jacob and Chang, Ming-Wei and Lee, Kenton and Toutanova, Kristina, (2019), "{BERT}: Pre-training of Deep Bidirectional Transformers for Language Understanding", in *Proceedings of the 2019 Conference of the North {A}merican Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, Association for Computational Linguistics, pp. 4171-4186.

[11] Dang, Thin and Nguyen, Duc-Vu and Nguyen, Kiet and Ngan, Nguyen, (2019), "A TRANSFORMATION METHOD FOR ASPECT-BASED SENTIMENT ANALYSIS", *Journal of Computer Science and Cybernetics,* vol. 34, pp. 323-333, 01.

[12] Thin, Dang and Nguyen, Duc-Vu and Nguyen, Kiet and Nguyen, Ngan and Hoang, Tu-Anh, (2020), "Multi-task Learning for Aspect and Polarity Recognition on Vietnamese Datasets", pp. 169-180.

[13] I. L. a. F. Hutter, (2017), "Decoupled Weight Decay Regularization", in *International Conference on Learning Representations*.

[14] Dang Thin and Ngan Nguyen, (2023), "Aspect-Category based Sentiment Analysis with Unified Sequence-To-Sequence Transfer Transformers", *VNU Journal of Science: Computer Science and Communication Engineering,* vol. 39.

Liên hệ:
Đặng Bá Quí
Trường Đại học Nông Lâm, thành phố Hồ Chí Minh
Địa chỉ: Phường Linh Trung, Thủ Đức, TP. Hồ Chí Minh
Email: 21130500@st.hcmuaf.edu.vn

Ngày nhận bài: 17/7/2024
Ngày gửi phản biện: 17/7/2024
Ngày duyệt đăng: 23/10/2024