

Phân tích cảm xúc dựa trên khía cạnh

Nguyễn Đức Duy Anh^{1,2,3}, Dương Quốc Lộc^{1,2,3}, Đặng Hoàng Quân^{1,2,3},
Võ Hồng Phúc Hạnh^{1,2,3}, and Nguyễn Hoàng Huy^{1,2,3}

¹ University of Information Technology, Ho Chi Minh City, Vietnam

² Vietnam National University, Ho Chi Minh City, Vietnam

³ {18520455, 18521006, 18520339, 18520275, 18520842 }@gm.uit.edu.vn

Abstract. Trong bài báo này, chúng tôi đã giải quyết hai nhiệm vụ trong bài toán phân tích cảm xúc dựa trên khía cạnh trong tiếng Việt đó là: Phát hiện khía cạnh, phân loại cảm xúc cho các khía cạnh đó. Bên cạnh đó, chúng tôi đề xuất các mô hình xử lý cùng lúc các nhiệm vụ trên với hai lĩnh vực (nhà hàng và khách sạn) của bộ dữ liệu VLSP 2018 Aspect Based Sentiment Analysis, đó là PhoBERT + multi-task và PhoBERT + multi-task hướng multi-branch. Cả hai mô hình đều cho kết quả rất tốt khi được áp dụng tiền xử lý. Đặc biệt mô hình multi-task đạt được kết quả state-of-the-art(SOTA) ở lĩnh vực khách sạn của bộ dữ liệu VLSP 2018 ABSA, cụ thể F1-score là 82.55% cho phát hiện khía cạnh và 77.32% cho phân loại cảm xúc. Còn lĩnh vực nhà hàng mô hình multi-task cũng đạt được SOTA tại nhiệm vụ phân loại cảm xúc với F1-score đạt được là 71.55%, ở nhiệm vụ phát hiện khía cạnh F1-score chúng tôi chỉ đạt được 83.29%.

Keywords: Aspect based sentiment analysis · PhoBERT · Aspect Category Detection · Sentiment Polarity Classification

1 Giới thiệu

Hiện nay, sự phổ biến của Internet đã dẫn đến sự bùng nổ về nguồn dữ liệu khổng lồ từ người dùng đặc biệt trong các lĩnh vực như thương mại điện tử, mạng xã hội, công cụ tìm kiếm. Chính điều này đã mang đến sự phát triển không ngừng trong lĩnh vực Trí tuệ nhân tạo (Artificial Intelligence - AI), đặc biệt là lĩnh vực Xử lý ngôn ngữ tự nhiên (Natural Language Processing - NLP). Trong đó nổi bật là bài toán Phân tích cảm xúc văn bản (Text sentiment analysis) ngày càng phổ biến và đạt nhiều thành công trong cả nghiên cứu và thương mại. Mục tiêu của Phân tích cảm xúc văn bản là hiểu được mức độ hài lòng của người dùng thông qua đánh giá của khách hàng đó. Tuy nhiên, nó chưa khai thác tối đa nguồn dữ liệu quý giá từ Internet bởi một đánh giá người dùng có thể chứa nhiều thông tin hữu ích cho các công ty hay các tổ chức nghiên cứu. Phân tích cảm xúc theo khía cạnh (Aspect based sentiment analysis - ABSA) là bài toán cải tiến của Phân tích cảm xúc văn bản sẽ giải quyết vấn đề đó.

Phân tích cảm xúc theo khía cạnh (Aspect based sentiment analysis) là một kỹ thuật phân tích văn bản, phân loại dữ liệu bằng khía cạnh và xác định cảm

xúc cho mỗi khía cạnh đó. Cụ thể hơn, nó được dùng để phân tích các đánh giá của khách hàng bằng cách liên kết các cảm xúc với các loại khía cạnh của sản phẩm hay dịch vụ. Khi nói đến khía cạnh, chúng có nghĩa là các thuộc tính hay thành phần của một sản phẩm hoặc dịch vụ. Ví dụ, trong miền khách sạn, các khía cạnh có thể là trải nghiệm của khách hàng về một dịch vụ, thời gian phản hồi với các phàn nàn của khách sạn, chất lượng của các đồ dùng trong phòng. Với sự phổ biến của internet, di động thông minh, mọi người giờ đây đều có thể dễ dàng đánh giá về khía cạnh của sản phẩm hay dịch vụ một cách nhanh chóng, tiện lợi. Nguồn dữ liệu phản hồi đánh giá vô cùng lớn, là một nguồn tài nguyên rất đáng giá cho các công ty. Các công ty thu thập dữ liệu và rút trích ra những thông tin giá trị để hiểu khách hàng muốn gì, cần gì qua đó định hướng phát triển. Để trích rút ra những thông tin đó thì bài toán Phân tích khía cạnh cảm xúc chính là điểm mấu chốt ở đây. Trong lĩnh vực ngôn ngữ tự nhiên, bài toán ABSA được đề xuất lần đầu tiên tại SemEval-2014 Task 4 [8] bởi Pontiki và các cộng sự, sau đó có nhiều cuộc thi lớn về Phân tích khía cạnh cảm xúc đã được tổ chức và đạt được những kết quả nổi bật như SemEval 2015 task 12[7], SemEval2016 task 5[6]. Thêm vào đó, với sự xuất hiện của các mô hình ngôn ngữ lớn, mạnh mẽ như BERT [1] giúp cho các bài toán trong xử lý ngôn ngữ tự nhiên có những bước tiến vượt bậc.

Trong đồ án này, chúng tôi thực nghiệm các mô hình máy học hiện đại, mạnh mẽ cho bài toán Phân tích khía cạnh cảm xúc trên bộ dữ liệu VLSP 2018 ABSA với hai lĩnh vực là nhà hàng và khách sạn. Bài toán Phân tích khía cạnh cảm xúc mà chúng tôi thực hiện xác định hai vấn đề, đó là: Phát hiện loại khía cạnh (Aspect Category Detection) và Phát hiện loại khía cạnh rồi Phân loại cảm xúc (Sentiment Polarity Classification) với loại khía cạnh đó. Chúng tôi dùng các mô hình học máy hiện đại để giải quyết hai vấn đề trên. Độ đo đánh giá được chúng tôi sử dụng cho cả ba bài toán con là F1-score micro.

Trong bài báo này, chúng tôi tập trung giới thiệu các thông tin liên quan đến bài toán phân tích cảm xúc theo khía cạnh trên ngôn ngữ tiếng Việt. Trong mục 2, chúng tôi sẽ trình bày về các bài toán con. Trong mục 3, chúng tôi sẽ trình bày về các công trình nghiên cứu liên quan. Trong mục 4, giới thiệu về bộ dữ liệu VLSP 2018 ABSA. Trong mục 5, chúng tôi sẽ giới thiệu các phương pháp được sử dụng để xử lý dữ liệu và huấn luyện mô hình. Quá trình thực nghiệm sẽ được mô tả trong mục 6 và mục 7 phân tích các kết quả. Cuối cùng là kết luận và hướng phát triển 8.

2 Mô tả bài toán

Bài toán chúng tôi thực hiện là Phân tích khía cạnh cảm xúc ở mức câu, tức là, với mỗi câu chọn ra hai mục tiêu đó là: phát hiện khía cạnh và phân loại cảm xúc các khía cạnh đã phát hiện. Chúng tôi chia bài toán Phân tích khía cạnh cảm xúc thành 2 bài toán con:

2.1 Phát hiện Loại Khía cạnh (Aspect Category Detection)

Nhận dạng thực thể E và thuộc tính A cặp với một ý kiến được diễn đạt trong một câu đã cho. E và A nên được chọn từ kho được xác định trước của các loại thực thể (ví dụ: "ROOMS", "HOTEL") và nhãn thuộc tính (ví dụ: "PRICE", "QUALITY").

2.2 Phân loại Cảm xúc (Sentiment Polarity Classification)

Mỗi cặp E#A xác định phải được chỉ định một trong những nhãn phân cực cảm xúc sau đây: "Positive", "Negative", "Neutral" (dương tính nhẹ hoặc âm tính nhẹ).

Một ví dụ cho bài toán như sau:

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

⇒ {cate: "ROOMS#DESIGN", pol: "positive"}.

⇒ {cate: "SERVICE#GENERAL", pol: "positive"}.

Trong đó, cate là Loại khía cạnh, pol là phân loại cảm xúc.

3 Công trình liên quan

Trong bài báo SemEval-2016 Task 5: Aspect Based Sentiment Analysis[6], Maria Pontiki và các cộng sự đã mô tả nhiệm vụ phân tích cảm xúc dựa trên khía cạnh trong SemEval 2016. Trong năm thứ ba, nhiệm vụ đã cung cấp 19 tập dữ liệu huấn luyện và 20 tập dữ liệu kiểm thử cho 8 ngôn ngữ và 7 lĩnh vực, cũng như một quy trình đánh giá chung. Từ các bộ dữ liệu này, 25 cho cấp độ câu và 14 cho ABSA cấp văn bản; cái sau được giới thiệu lần đầu tiên như một nhiệm vụ con trong SemEval. Nhiệm vụ thu hút 245 bài dự thi của 29 đội. Ba nhiệm vụ con trong bài báo là Phân tích cảm xúc dựa trên khía cạnh mức độ câu, Phân tích cảm xúc dựa trên khía cạnh mức độ văn bản và Phân tích cảm xúc dựa trên khía cạnh ở các lĩnh vực ngoài. SemEval ABSA 16 là lần đầu tiên bao gồm một nhiệm vụ phụ cấp văn bản.

Bài báo Deep Learning for Aspect-Based Sentiment Analysis: A Comparative Review[2] của Hai Ha Do và các cộng sự vào năm 2018 đã chỉ ra khối lượng ngày càng tăng của nội dung do người dùng tạo trên web đã làm cho phân tích cảm xúc trở thành một công cụ quan trọng để khai thác thông tin về trạng thái cảm xúc của con người. Bài viết này nhằm mục đích cung cấp một đánh giá so sánh về học sâu để phân tích tình cảm dựa trên khía cạnh để đặt các cách tiếp cận khác nhau trong ngữ cảnh.

Trong bài báo Aspect-based sentiment analysis of movie reviews on discussion boards[10], các tác giả đã thực hiện một phương pháp phân tích tình cảm tự động của các bài phê bình phim được đề xuất, thực hiện và đánh giá. Phương pháp này sử dụng phân tích chi tiết để xác định cả định hướng tình cảm và sức mạnh tình cảm của người đánh giá đối với các khía cạnh khác nhau của một bộ phim. Nó áp dụng một phương pháp tiếp cận ngôn ngữ để tính toán cảm xúc của một điều khoản từ điểm số cảm xúc trước đó.

Bài báo A Hierarchical Model of Reviews for Aspect-based Sentiment Analysis [9] của Sebastian Ruder và các cộng sự vào năm 2016 đã chỉ ra các lý thuyết khai thác ý kiến từ đánh giá của khách hàng đã trở nên phổ biến trong những năm gần đây. Các câu thường được phân loại độc lập, mặc dù chúng là một phần của cấu trúc lập luận của bài đánh giá. Họ cho thấy rằng LSTM hai chiều phân cấp hoạt động tốt hơn hai đường cơ sở không phân cấp và cạnh tranh với công nghệ tiên tiến nhất trên năm bộ dữ liệu đa ngôn ngữ, đa miền mà không có bất kỳ tính năng được điều chỉnh thủ công nào hoặc tài nguyên bên ngoài.

Aspect Based Sentiment Analysis with Gated Convolutional Networks[14] được xuất bản vào 2018 của các tác giả Wei Xue, Tao Li đã chỉ ra các Gated Tanh-ReLU Units có thể xuất ra một cách chọn lọc các đặc điểm tình cảm theo khía cạnh hoặc thực thể nhất định. Kiến trúc đơn giản hơn nhiều so với lớp chú ý được sử dụng trong các mô hình hiện có. Họ đề xuất một mô hình dựa trên mạng nơ-ron tích tụ và cơ chế định vị.

SemEval-2015 Task 12: Aspect Based Sentiment Analysis [7] của tác giả Maria Pontiki và các đồng sự đã mô tả nhiệm vụ SE-ABSA15 đã thu hút 93 bài nộp từ 16 đội được đánh giá theo ba vị trí: phân loại khía cạnh, biểu hiện mục tiêu quan điểm và phân loại cực. Công việc trong tương lai bao gồm việc áp dụng khuôn khổ mới cho các ngôn ngữ khác (ví dụ: tiếng Tây Ban Nha, tiếng Hy Lạp) và chú thích các chủ đề hoặc sự kiện. Trong nhiệm vụ này gồm hai nhiệm vụ phụ đó là Phân tích cảm xúc dựa trên khía cạnh ở các lĩnh vực trong, Phân tích cảm xúc dựa trên khía cạnh ở các lĩnh vực ngoài.

A BERT-based Hierarchical Model for Vietnamese Aspect Based Sentiment Analysis[11] của tác giả Oanh Thi Tran và cộng sự đã mô tả (ABSA) là nhiệm vụ xác định các cực của cảm xúc đối với các thực thể cụ thể và các khía cạnh của chúng được đề cập trong đánh giá của khách hàng. Bài báo này trình bày một mô hình phân cấp mới và hiệu quả bằng cách sử dụng mô hình ngôn ngữ được đào tạo trước, biểu diễn mã hóa hai chiều từ máy biến áp (BERT)

VLSP shared task: sentiment analysis [5] của Nguyen Thi Minh Huyen và các cộng sự. Trong đó đã trình bày chuỗi hội thảo VLSP về Phân tích Cảm xúc (SA) dành cho người Việt Nam đã được tổ chức nhằm cung cấp một phép đo đánh giá khách quan về hiệu suất (chất lượng) của các công cụ phân tích cảm xúc. Bài báo này mô tả các tập dữ liệu đã xây dựng cũng như kết quả đánh giá của các hệ thống tham gia vào các chiến dịch này.

Multi-task Learning for Aspect and Polarity Recognition on Vietnamese Datasets[12] là nghiên cứu do Thim Van Dang và các cộng sự đã thực hiện. Trong đó, nhà nghiên cứu Việt Nam đã phát triển một mô hình mạng nơ ron sâu để giải quyết hai nhiệm vụ trong phân tích cảm xúc dựa trên khía cạnh ở cấp độ tài liệu trong bộ dữ liệu của Việt Nam. Họ nhận thấy rằng mô hình của họ hoạt động tốt hơn so với các phương pháp tiếp cận đường cơ sở ổn định trên cả hai nhiệm vụ cho hai lĩnh vực và đạt được F1-score 64.78% cho lĩnh vực nhà hàng và 70.90% cho lĩnh vực khách sạn.

4 Bộ dữ liệu

Trong bài này chúng tôi đã sử dụng bộ dữ liệu VLSP 2018 Aspect based sentiment analysis⁴ với hai miền dữ liệu là nhà hàng và khách sạn.

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

Bảng 1: Thống kê bộ dữ liệu VLSP 2018 Aspect based sentiment analysis

Bảng 1 trình bày thống kê của bộ VLSP 2018 Aspect based sentiment analysis trên hai lĩnh vực là khách sạn và nhà hàng. Trong đó các cột thống kê lần lượt là lĩnh vực của các bình luận Domain, số lượng bình luận (#Reviews), số lượng khía cạnh (#Aspects), độ dài trung bình bình luận (AvgLength), kích thước bộ từ vựng (VocabSize), số lượng từ vựng trong tập Test hoặc Dev không xuất hiện trong tập Training (#DiffVocab).

Đánh giá	Nhãn
Rộng rãi KS mới nhưng rất vắng. Các dịch vụ chất lượng chưa cao và thiếu.	{HOTEL#DESIGN&FEATURES,positive}, {HOTEL#DESIGN&GENERAL,negative}
Địa điểm thuận tiện, trong vòng bán kính 1,5km nhiều quán ăn ngon	{LOCATION&GENERAL,positive}
Phục vụ, view đẹp, vị trí	{SERVICE#GENERAL,positive}, {HOTEL#GENERAL,positive}, {LOCATION#GENERAL,positive}

Bảng 2: Lĩnh vực khách sạn VLSP 2018 ABSA.

Bảng 2 và Bảng 3 lần lượt là các ví dụ của lĩnh vực khách sạn và lĩnh vực nhà hàng của bộ dữ liệu VLSP 2018.

5 Phương pháp

5.1 Tiền xử lý dữ liệu

Trong xử lý ngôn ngữ tự nhiên nói chung, và xử lý ngôn ngữ tiếng Việt nói riêng thì bước đầu tiên mà chúng ta cần làm là tiền xử lý dữ liệu. Việc xử lý này nếu

⁴ <https://vlsp.org.vn/vlsp2018/eval/sa>

Đánh giá	Nhãn
Ngon tuyệt vời luôn!	{FOOD#QUALITY, positive}
Gà rán giòn ướp vị chua cay rất ngon, bánh ăn kèm giòn và thơm mùi bơ	{FOOD#QUALITY, positive}, {FOOD#STYLE&OPTIONS, positive}
#lozi #lozisaigon #anchinh chả ngon !!	{FOOD#QUALITY, positive}

Bảng 3: Lĩnh vực nhà hàng VLSP 2018 ABSA.

phù hợp với bài toán cũng như là phù hợp với dữ liệu có thể làm tăng hiệu suất. Do đó, tiền xử lý dữ liệu là quá trình chuẩn hóa dữ liệu và loại bỏ các thành phần không có ý nghĩa cho việc xử lý bài toán. Vì vậy chúng tôi đã thiết kế một số bước để xử lý dữ liệu nhằm khai thác tốt hơn các thông tin so với dữ liệu gốc. Các bước được thực hiện lần lượt theo mô tả bên dưới: Đầu tiên, chúng tôi thực



Hình 1: Các bước tiền xử Lý

hiện xóa các HTML code trong bộ dữ liệu ban đầu. Sau đó chúng tôi thực hiện chuẩn hóa. Cụ thể là chuẩn hóa văn bản từ unicode tổ hợp về unicode dựng sẵn và chuẩn hóa các từ viết tắt thông dụng trong tiếng Việt. Trong tiếng Việt một từ có thể được cấu thành bởi hai hoặc nhiều từ khác ghép lại để tạo thành (gọi là từ ghép), chẳng hạn như “nhà_hàng” (nhà hàng mỗi từ có một nghĩa riêng khi đứng riêng lẻ nhưng khi ghép lại thì nó lại mang một nghĩa chung). Vì vậy, cần phải tách từ tiếng Việt trước khi bắt đầu các xử lý tiếp theo. Trong bài báo này chúng tôi sử dụng bộ công cụ tách từ VnCoreNLP [13]. Cuối cùng là xóa các ký tự không cần thiết, việc này giúp làm giảm số chiều đặc trưng của câu, tăng tốc độ xử lý, tránh làm ảnh hưởng tới kết quả của mô hình.

5.2 PhoBERT

PhoBERT[4] là các mô hình ngôn ngữ đơn ngữ cỡ lớn được tiền huấn luyện cho tiếng Việt. Các mô hình tiền huấn luyện PhoBERT đạt được các kết quả vượt trội trên bốn tác vụ Xử lý ngôn ngữ tự nhiên tiếng Việt, đó là: gán nhãn từ loại (Part-of-speech POS tagging), phân tích cú pháp phụ thuộc (Dependency parsing), và Nhận diện tên thực thể (Named entity recognition - NER) cấp độ từ (word-level); và một tác vụ hiểu ngôn ngữ của Suy luận Ngôn ngữ tự nhiên (Natural language inference) cấp độ từ hoặc âm tiết (syllable-level).

PhoBERT có hai phiên bản là PhoBERT_{base} và PhoBERT_{large}, Hướng tiếp cận tiền huấn luyện PhoBERT được dựa trên RoBERTa[3], mô hình được tối ưu hóa từ tiền huấn luyện BERT để có hiệu suất tốt hơn.

5.3 Kiến trúc mô hình

Nhóm sẽ triển khai cùng một kiến trúc model cho 2 cách tiếp cận: Multi-task và Multi-task theo hướng Multi-branch, để xử lý cả 2 tác vụ Phát hiện khía cạnh và Phân loại cảm xúc của khía cạnh đó cùng một lúc, sử dụng PhoBERT_{base} làm các mô hình ngôn ngữ tiền huấn luyện cho Tiếng Việt. Dựa theo bài báo gốc của BERT[1], mô hình đạt kết quả cao nhất khi nối 4 layers cuối cùng của BERT lại với nhau. Vì thế nhóm sẽ áp dụng phương pháp này cho kiến trúc mô hình với đầu vào được biến đổi thành 1 vector low-dimensional $x_I \in R^d$, trong đó d là chiều dài của vector.

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

Bảng 4: Kết quả với các phương pháp khác nhau của BERT

5.4 Multi-task

Mô hình sẽ được xây dựng với đầu ra được mô hình hóa thành một danh sách gồm C one-hot vectors với C là số lượng khía cạnh có trong bộ dữ liệu được sử dụng. Đối với bộ dữ liệu của VLSP, số lượng khía cạnh C sẽ lần lượt là 12 cho lĩnh vực Restaurant và 34 cho lĩnh vực Hotel. Mỗi vector có 4 thành phần tương ứng với 3 nhãn của Phân loại cảm xúc gồm Positive, Negative, Neutral và 1 nhãn None để chỉ ra rằng đầu vào không có khía cạnh này. Do đây là 1 one-hot vector nên loại cảm xúc nào tồn tại sẽ được gán là 1, còn lại sẽ là 0.

Các đặc trưng học được sẽ được đưa qua 1 lớp fully connected được tạo nên bằng việc nối C lớp fully connected tương ứng với C one-hot vectors lại. Như vậy, ta sẽ được 1 lớp fully connected gồm 48 neurons đối với bộ dữ liệu Restaurant (12 khía cạnh \times 4 loại cảm xúc), 136 neurons đối với bộ dữ liệu Hotel (34 khía cạnh \times 4 loại cảm xúc) và sử dụng hàm softmax để tính toán score cho giá trị dự đoán y_{pred} cho mỗi khía cạnh A :

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

Vì thế ta có thể dự đoán một khía cạnh a và loại cảm xúc của nó trong một bước bằng cách:

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

Đối với bài toán phân loại nhị phân, hàm loss sẽ có dạng:

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

Trong bài toán phân loại có C nhãn nhiều hơn 2. Đồng thời ta sử dụng hàm softmax để tính phân phối xác suất đầu ra thì hàm loss là một hàm cross entropy như sau:

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

Trong đó i là chỉ số mẫu, j là chỉ số từng tác vụ.

Trong thuật toán multitask learning, đối với mỗi tác vụ phân loại sẽ có giá trị hàm loss là phương trình (3). Vậy khi có C tác vụ phân loại khác nhau, hàm loss của multi-task learning sẽ là tổng các hàm loss (dạng binary cross entropy) của từng bài toán phân loại nhị phân ứng với mỗi tác vụ:

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

5.5 Multi-task Multi-branch

Mô hình chỉ khác mô hình trên ở chỗ là sẽ phân nhánh thành nhiều mô hình con bằng cách sử dụng C lớp fully connected nhưng không nối chúng lại thành 1 lớp fully connected duy nhất. Mỗi mô hình sẽ phụ trách dự báo cho mỗi tác vụ một cách độc lập.

Mô hình sử dụng hàm softmax để tính toán các score cho giá trị dự đoán và có thể dự đoán khía cạnh cùng với phân loại cảm xúc của nó cùng một lúc với cách tương tự như mô hình trên. Hàm loss tổng quát sẽ bằng tổng các hàm loss ở tất cả C nhánh:

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

6 Thực nghiệm

Nhóm sử dụng kích thước batch là 21 đối với bộ dữ liệu Restaurant và 20 đối với bộ dữ liệu Hotel với đầu vào được padding theo số lượng tokens tối đa được chấp nhận bởi PhoBERT, cụ thể là 256 tokens. Nhóm chọn số lượng epoch là 20 và sử dụng Early Stopping để dừng quá trình huấn luyện nếu loss của tập validation không có cải thiện gì sau 5 epochs.

Để fine-tuning, nhóm sử dụng cùng thuật toán tối ưu mà BERT pre-training đã được thiết lập: Adaptive Moments (Adam). Optimizer này giảm thiểu sự mất mát (loss) trong dự đoán và thực hiện regularization bằng cách phân rã hay giảm dần weight (không sử dụng moment), còn được gọi là AdamW. Đối với learning

rate, nhóm sử dụng cùng 1 learning rate schedule như BERT pre-training: phân rã tuyến tính (linear decay) của 1 learning rate ban đầu, được bắt đầu bằng 1 giai đoạn linear warm-up trong 10% đầu tiên của training steps. Ở đây, nhóm sẽ khởi tạo 1 learning rate có giá trị là $2e - 4$.

7 Phân tích kết quả

Trong phần phân tích kết quả ngoài hai phương pháp Multi-task và Multi-task theo hướng Multi-branch chúng tôi có trình bày các kết quả khác để có thể so sánh với kết quả của chúng tôi. Cụ thể VLSP best submission được tham khảo từ bài báo VLSP SHARED TASK: SENTIMENT ANALYSIS[5], Bi-LSTM+CNN được tham khảo từ Multi-task Learning for Aspect and Polarity Recognition on Vietnamese Datasets[12] và BERT-based Hierarchical được tham khảo từ A BERT-based Hierarchical Model for Vietnamese Aspect Based Sentiment Analysis [11].

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

Bảng 5: Kết quả đánh giá các mô hình trên bộ dữ liệu VLSP 2018 ABSA - Hotel, có sử dụng tiền xử lý và không tiền xử lý.

Bảng 5 biểu diễn kết quả của bộ dữ liệu VLSP 2018 ABSA lĩnh vực khách sạn khi không có tiền xử lý và khi có tiền xử lý. Khi áp dụng tiền xử lý phương pháp Multi-task cho kết quả tốt nhất trên cả hai nhiệm vụ là phát hiện khía cạnh(ACD) và phát hiện khía cạnh cùng với phân loại cảm xúc (ACD+APC) cao hơn SOTA hiện tại là model BERT-based Hierarchical. Cụ thể F1-score micro của ACD và ACD+APC là 82.55% và 77.32%. Đối với kết quả không áp dụng tiền xử lý, Multi-task cho kết quả cao nhất ở nhiệm vụ ACD+APC với F1-score micro là 75.45% và kết quả rất tốt ở nhiệm vụ ACD với F1-score micro là 81.56% chỉ thua BERT-based Hierarchical (F1-score micro-82.06%). Kết quả giữa Multi-task và Multi-task theo hướng Multi-branch chênh lệch nhau khá nhiều khoảng 20% trong cả hai trường hợp có tiền xử lý và không tiền xử lý ở cả hai nhiệm vụ ACD và ACD+APC.

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

Bảng 6: Kết quả đánh giá các mô hình trên bộ dữ liệu VLSP 2018 ABSA - Restaurant, có sử dụng tiền xử lý và không tiền xử lý.

Bảng 6 biểu diễn kết quả của bộ dữ liệu VLSP 2018 ABSA lĩnh vực nhà hàng khi không có tiền xử lý và khi có tiền xử lý. Khi áp dụng tiền xử lý phương pháp Multi-task và Multi-task hướng multi-branch cho kết quả tốt nhất trên

nhiệm vụ là phát hiện khía cạnh cùng với phân loại cảm xúc (ACD+APC) cao hơn SOTA hiện tại là model BERT-based Hierarchical. Cụ thể F1-score micro ACD+APC là 71.55% và 71.38%, còn nhiệm vụ phát hiện khía cạnh cả hai model trên F1-score micro chỉ đạt lần lượt là 83.29% và 83.97% chênh lệch với BERT-based Hierarchical là khoảng 1%-0.26%. Đối với kết quả không áp dụng tiền xử lý, Mô hình BERT-based Hierarchical cho kết quả tốt nhất ở cả hai nhiệm vụ ACD và ACD+APC, kết quả thu được từ Multi-task và Multi-task theo hướng Multi-branch cũng rất tốt. Cụ thể Multi-task là 80.57%, 69.10% và Multi-task theo hướng Multi-branch là 80.43%, 68.06% ở cả hai nhiệm vụ ACD và ACD+APC. Trong đó thì Multi-task có kết quả nhỉnh hơn Multi-task theo hướng Multi-branch khoảng 1% ở tiền xử lý và không tiền xử lý dữ liệu ở các nhiệm vụ trừ nhiệm vụ ACD với tiền xử lý dữ liệu.

Nhìn chung thì tiền xử giúp cải thiện độ chính xác của các mô hình trên cả hai lĩnh vực nhà hàng và khách sạn.

8 Kết luận và hướng phát triển

Trong đề án này chúng tôi đã giải quyết bài toán phân loại cảm xúc theo khía cạnh trên bộ dữ liệu VSLP 2018 ASBA với hai khía cạnh là nhà hàng và khách sạn. Chúng tôi đã thực hiện các mô hình là PhoBERT + multi-task và PhoBERT + multi-task hướng multi-branch. Trong đó, mô hình multi-task cho kết quả tốt nhất trên 2 nhiệm vụ của bộ dữ liệu VLSP 2018 ABSA lĩnh vực khách sạn và khách sạn. Mô hình multi-task theo hướng multi-branch cho kết quả tốt ở dữ liệu VLSP 2018 ABSA lĩnh vực nhà hàng. Ngoài ra, mô hình multi-task đạt được kết quả state-of-the-art ở lĩnh vực khách sạn, còn với lĩnh vực nhà hàng mô hình chỉ đạt state-of-the-art ở nhiệm vụ phát hiện khía cạnh cùng với phân loại cảm xúc của bộ dữ liệu VLSP 2018 ABSA. Thêm vào đó, thông qua quá trình thực nghiệm chúng tôi nhận thấy tiền xử lý giúp tăng hiệu suất mô hình.

Trong tương lai, chúng tôi dự định tìm hiểu hướng tiếp cận khác cho bài toán như: các mô hình về Graph, biến thể BERT. Thay đổi các lớp Embedding cho mô hình multi-task: PhoW2V, fastText, multi Embedding, vELECTRA. Xây dựng phương pháp tăng cường cho tập huấn luyện của VLSP nhằm tăng tính hiệu quả của mô hình như: Translate languages, Paraphrase, Word similarity.

References

1. Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.
2. Hai Ha Do, PWC Prasad, Angelika Maag, and Abeer Alsadoon. Deep learning for aspect-based sentiment analysis: a comparative review. *Expert Systems with Applications*, 118:272–299, 2019.
3. Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*, 2019.

4. Dat Quoc Nguyen and Anh Tuan Nguyen. Phobert: Pre-trained language models for vietnamese. *arXiv preprint arXiv:2003.00744*, 2020.
5. Huyen TM Nguyen, Hung V Nguyen, Quyen T Ngo, Luong X Vu, Vu Mai Tran, Bach X Ngo, and Cuong A Le. Vlsr shared task: sentiment analysis. *Journal of Computer Science and Cybernetics*, 34(4):295–310, 2018.
6. Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Mohammad AL-Smadi, Mahmoud Al-Ayyoub, Yanyan Zhao, Bing Qin, Orphée De Clercq, Véronique Hoste, Marianna Apidianaki, Xavier Tannier, Natalia Loukachevitch, Evgeniy Kotelnikov, Nuria Bel, Salud María Jiménez-Zafra, and Gülşen Eryigit. SemEval-2016 task 5: Aspect based sentiment analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 19–30, San Diego, California, June 2016. Association for Computational Linguistics.
7. Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Suresh Manandhar, and Ion Androutsopoulos. SemEval-2015 task 12: Aspect based sentiment analysis. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 486–495, Denver, Colorado, June 2015. Association for Computational Linguistics.
8. Maria Pontiki, Dimitris Galanis, John Pavlopoulos, Harris Papageorgiou, Ion Androutsopoulos, and Suresh Manandhar. SemEval-2014 task 4: Aspect based sentiment analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 27–35, Dublin, Ireland, August 2014. Association for Computational Linguistics.
9. Sebastian Ruder, Parsa Ghaffari, and John G Breslin. A hierarchical model of reviews for aspect-based sentiment analysis. *arXiv preprint arXiv:1609.02745*, 2016.
10. Tun Thura Thet, Jin-Cheon Na, and Christopher SG Khoo. Aspect-based sentiment analysis of movie reviews on discussion boards. *Journal of information science*, 36(6):823–848, 2010.
11. Oanh Thi Tran and Viet The Bui. A bert-based hierarchical model for vietnamese aspect based sentiment analysis. In *2020 12th International Conference on Knowledge and Systems Engineering (KSE)*, pages 269–274, 2020.
12. Dang Van Thin, Duc-Vu Nguyen, Kiet Van Nguyen, Ngan Luu-Thuy Nguyen, and Anh Hoang-Tu Nguyen. Multi-task learning for aspect and polarity recognition on vietnamese datasets. In *International Conference of the Pacific Association for Computational Linguistics*, pages 169–180. Springer, 2019.
13. Thanh Vu, Dat Quoc Nguyen, Dai Quoc Nguyen, Mark Dras, and Mark Johnson. VnCoreNLP: A Vietnamese natural language processing toolkit. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations*, pages 56–60, New Orleans, Louisiana, June 2018. Association for Computational Linguistics.
14. Wei Xue and Tao Li. Aspect based sentiment analysis with gated convolutional networks. *arXiv preprint arXiv:1805.07043*, 2018.