

Aspect-based Sentiment Analysis for Vietnamese Reviews: MinPhoBERT Application with Contrastive Learning

Phạm Đức Hoàng
Đại học Công Nghệ, VNU
Hà Nội, Việt Nam
22021200@vnu.edu.vn

Trần Đức Lương
Đại học Công Nghệ, VNU
Hà Nội, Việt Nam
22021148@vnu.edu.vn

Nguyễn Việt An
Đại học Công Nghệ, VNU
Hà Nội, Việt Nam
22021168@vnu.edu.vn

Tóm tắt nội dung—Sentiment Analysis đã trở nên rất phổ biến trong cả nghiên cứu và kinh doanh do lượng lớn văn bản mang tính ý kiến hiện đang được tạo ra bởi người dùng Internet. Aspect-based Sentiment Analysis (ABSA) là một nhiệm vụ phức tạp hơn, bao gồm việc xác định cả cảm xúc và các khía cạnh. Dự án này vận dụng tiềm năng của việc sử dụng các biểu diễn từ ngữ theo ngữ cảnh từ mô hình ngôn ngữ tiền huấn luyện BERT, cùng với phương pháp Contrastive Learning, để giải quyết vấn đề ABSA trên tập dữ liệu gồm các đánh giá Tiếng Việt. Kết quả cho tốt hơn các phương pháp truyền thống, khi đạt F1-score 0.87 cho tác vụ Aspect Detection và 0.82 cho tác vụ Sentiment Classification. Code của mô hình có thể truy cập từ tại GitHub repository¹.

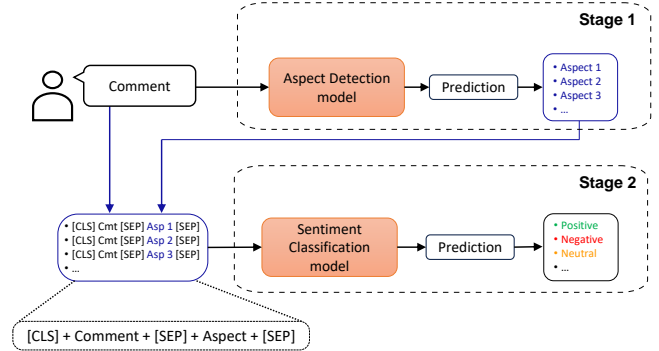
Index Terms—Aspect-Based Sentiment Analysis, Sentiment Analysis, BERT, Contrastive Learning

I. INTRODUCTION

Sentiment Analysis có hiệu quả đóng góp lớn trong thực tế, nó sử dụng các thuật toán phức tạp để xử lý ngôn ngữ tự nhiên của con người (NLP) và xác định các đặc điểm cảm xúc tiêu cực/tích cực tại một thời điểm thông qua văn bản hoặc lời nói. Nhưng nó gặp phải khó khăn khi trong một ngữ cảnh có thể xuất hiện nhiều ý kiến tại các khía cạnh khác nhau của một đối tượng. Do đó, Aspect-based Sentiment Analysis ra đời với hy vọng có thể giải quyết vấn đề của Sentiment Analysis cơ bản. Phân tích quan điểm dựa trên khía cạnh (ABSA) là một kỹ thuật phân tích văn bản giúp phân loại dữ liệu theo khía cạnh và xác định quan điểm được gán cho từng khía cạnh đó. ABSA có thể được sử dụng để phân tích phản hồi của khách hàng bằng cách liên kết quan điểm cụ thể với các khía cạnh khác nhau của sản phẩm hoặc dịch vụ.

Bài báo [10] sử dụng mô hình Bi-LSTM để giải quyết bài toán phân tích cảm xúc dựa trên khía cạnh. Tuy nhiên, việc sử dụng Bi-LSTM vẫn chỉ là kiểu học tuần tự, chưa thực sự biểu diễn được hết ngữ cảnh của câu. Transformer [19] cũng như mô hình BERT [3] ra đời đã giải quyết được vấn đề mất thông tin ngữ nghĩa. Trong bài báo này, chúng tôi đề xuất 3 giải pháp nhằm giải quyết nhược điểm của bài báo [10] cũng như đề xuất một số phương pháp hiệu quả khác như sau:

- **Đề xuất sử dụng pretrained model MinPhoBERT:** Lựa chọn MinPhoBERT, một mô hình pretrained mạnh mẽ trên dữ liệu tiếng Việt, làm nền tảng cơ sở để giải quyết bài toán đặt ra.
- **Tăng cường khả năng biểu diễn ngữ nghĩa bằng Contrastive Learning:** Áp dụng Contrastive Learning nhằm nâng cao chất lượng biểu diễn ngữ nghĩa của mô hình, giúp nhận diện và phân biệt các ngữ cảnh tốt hơn.
- **Giải quyết vấn đề unbalanced data:** Thực hiện upsampling dữ liệu bằng hai phương pháp: nhân bản dữ liệu nhân unbalanced hiện có (duplicate) và tạo dữ liệu mới bằng LLMs, cụ thể là Gemma2:2b [18].



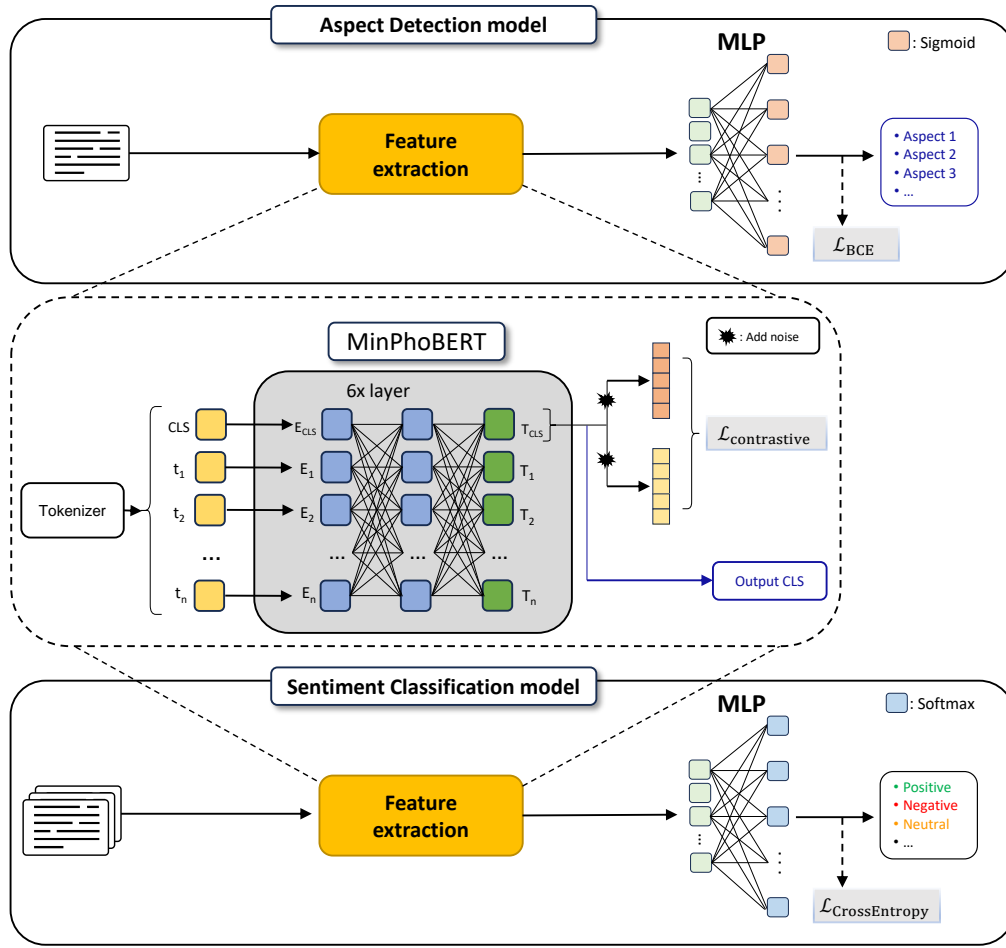
Hình 1. Tổng quan kiến trúc mô hình

II. RELATED WORK

Trong nhiều năm qua, Sentiment analysis là bài toán nhận được nhiều sự quan tâm từ giới nghiên cứu cho tới doanh nghiệp, bởi vì tính ứng dụng cao của nó trong thực tế. Trong phần này, chúng tôi sẽ cung cấp cái nhìn tổng quan về các kỹ thuật và mô hình sử dụng trong phần còn lại của bài báo, cũng như các nghiên cứu nổi bật liên quan hiện có.

a) **Aspect-based Sentiment Analysis:** ABSA (Aspect-Based Sentiment Analysis) là một nhiệm vụ phức tạp hơn so với phân tích cảm xúc ở mức độ văn bản truyền thống. Nhiệm vụ này tập trung vào việc xác định các thuộc tính

¹Code available at: <https://github.com/dluong1210/ABSA-model>



Hình 2. Chi tiết Aspect Detection model và Sentiment Classification model

hoặc khía cạnh được đề cập trong văn bản, cùng với cảm xúc được biểu đạt đối với từng khía cạnh đó. Nhiều hội thảo như SemEval2014 [14], SemEval2016 [15], đã được tổ chức nhằm tìm ra giải pháp tối ưu nhất cho vấn đề này.

Các phương pháp trong bài toán ABSA thường sử dụng các phương pháp học sâu như: RNN, Long Short-Term Memory (LSTM), Bidirectional LSTM (BiLSTM) [22, 6, 20, 16]. Bên cạnh đó, các phương pháp học máy truyền thống cũng được ứng dụng nhiều [22] như: Support Vector Machine (SVM) [8, 12], Conditional Random Field (CRF) [21], Naive Bayes, Random Forest, Decision Tree. Gần đây, một số nghiên cứu đề xuất phương pháp mới học tập theo ngữ cảnh (In-context learning) [23, 1, 17]. Các phương pháp này mở ra hướng nghiên cứu mới, nhưng vẫn đạt kết quả chưa tốt với các phương pháp non - ICL trước đó [22]. Nghiên cứu [10] trên bộ data UIT-ViFSD, một bộ dữ liệu bằng tiếng Việt dành cho việc đánh giá phân tích cảm xúc dựa trên khía cạnh (ABSA) trong lĩnh vực thương mại điện tử di động. Trong bài báo, nhóm tác giả đã sử dụng mô hình Bi-LSTM [5] kết hợp với bộ phân lớp để giải quyết bài toán phân tích cảm xúc dựa trên khía cạnh. Nghiên cứu cho thấy kết quả cải thiện hơn so với các phương pháp trước đó như: Naive Bayes, SVM, CNN, LSTM. Kết quả

của bài báo đạt được trên F1-Score với tác vụ Aspect Detection là 84,48%, và 63.06% với tác vụ Sentiment Classification. Mô hình này gặp hạn chế ở tác vụ Sentiment Classification do khó xử lý ngữ cảnh phức tạp và câu dài, dẫn đến hiệu suất phân loại cảm xúc còn chưa cao. Từ đó, nhóm đề xuất phương pháp dựa trên mô hình RoBERTa [9] kết hợp với kỹ thuật Contrastive Learning để giải quyết các tác vụ trong bài báo. Kết quả thực nghiệm cho thấy phương pháp này đạt hiệu năng vượt trội ở cả hai tác vụ.

b) Bidirectional Encoder Representations from Transformers: BERT là một mô hình học sâu đề xuất bởi Google được giới thiệu vào năm 2018. Nó đánh giá bước nhảy vọt về hiệu suất trong các tác vụ xử lý ngôn ngữ tự nhiên. Được xây dựng trên nền tảng là kiến trúc Transformer. [19], nó kế thừa sự mạnh mẽ từ cơ chế self-attention cho phép mô hình nắm bắt các mối quan hệ ngữ nghĩa trong văn bản. Nhiều công trình nghiên cứu đã lấy nền là BERT để phát triển lên những mô hình có hiệu suất tốt hơn ở những nhiệm vụ cụ thể. [9] mở rộng BERT bằng việc loại bỏ task Next Sentence Prediction (NSP) và tăng cường dữ liệu huấn luyện.

PhoBERT [13] là một pretrained model được huấn luyện trên monolingual language, tức là chỉ huấn luyện dành riêng

cho tiếng Việt. Việc huấn luyện dựa trên kiến trúc và cách tiếp cận giống RoBERTa của Facebook được Facebook giới thiệu giữa năm 2019. Trong bài báo của chúng tôi, chúng tôi sử dụng PhoBERT_base với 6-12 transformers block cùng với tinh chỉnh lựa chọn số layers, sau đó kết hợp với MLP để giải quyết tác vụ ABSA.

c) **Contrastive learning**: là một kỹ thuật tối ưu được sử dụng phổ biến. Nó đã được chứng minh trong các lĩnh vực như Computer Vision [2], NLP [4]. Ý tưởng cơ bản của contrastive learning là học cách phân biệt các cặp dữ liệu tương đồng và không tương đồng. Cụ thể là đưa các cặp dữ liệu tương tự lại gần nhau trong không gian biểu diễn, đồng thời đẩy các cặp không tương tự ra xa. Từ đó, nó giúp mô hình học được các đặc trưng có ý nghĩa để phân loại, so sánh dữ liệu. Trong lĩnh vực CV, [2] sử dụng các kỹ thuật augmentation như thêm noise, crop, xoay ngang, xoay dọc, ... để tạo ra 1 cặp tương đồng từ 1 input, sau đó sử dụng hàm mất mát contrastive [2] để tối ưu hóa sự tương đồng giữa các hình ảnh giống nhau và giảm thiểu sự tương đồng với các hình ảnh khác. Bài báo này cũng chứng minh một kết quả nền tảng đó là việc tối ưu nêu trên trong không gian embedding có thể tạo ra các biểu diễn chất lượng cao cho nhiều tác vụ khác nhau, ngay cả khi không có nhãn. Tiếp nối thành công của [2], vào năm 2021, [4] được đề xuất cho dữ liệu text. Tuy nhiên việc áp dụng các kỹ thuật bên augmentation khó khăn hơn bên ảnh rất nhiều. Vì vậy, nhóm tác giả [4] đã giới thiệu kỹ thuật dropout để sinh ra các cặp tương đồng. Hàm loss contrastive cũng được sử dụng để tối ưu hóa sự tương đồng giữa các biểu diễn. Với bài báo của chúng tôi, chúng tôi sử dụng hàm Contrastive Loss kết hợp với Cross-Entropy Loss nhằm mục đích tăng khả năng biểu diễn ngữ nghĩa của PhoBERT cũng như khả năng phân loại của mô hình.

III. DATASET

A. Thống kê dữ liệu

Mô hình được đánh giá trên bộ dữ liệu UIT-ViSFD² [10], bao gồm các bình luận tiếng Việt cùng thông tin về các khía cạnh (aspects) được đề cập và cảm xúc của người dùng trên từng khía cạnh. Với mục tiêu tập trung vào lĩnh vực điện thoại thông minh, một sản phẩm thương mại phổ biến và đang tiếp tục phát triển mạnh mẽ, bộ dữ liệu này khai thác tiềm năng từ lượng lớn phản hồi của người dùng về các sản phẩm điện thoại. Điều này được thực hiện dựa trên việc tham khảo và tổng hợp từ nhiều nghiên cứu liên quan trong lĩnh vực này. Bảng VI liệt kê các khía cạnh trong bộ dữ liệu cùng ví dụ minh họa. Thống kê chi tiết về các tập Train, Dev, và Test được trình bày trong Bảng I.

B. Tăng cường dữ liệu (Upsampling)

Qua phân tích hình 3, có thể thấy bộ dữ liệu UIT-ViSFD bị mất cân bằng nghiêm trọng, đặc biệt ở khía cạnh STORAGE có số lượng mẫu rất thấp so với các khía cạnh khác. Để khắc phục vấn đề này, nhóm đã thử nghiệm hai phương pháp tăng cường dữ liệu nhằm cải thiện hiệu quả mô hình:

Bảng I
THỐNG KÊ TRAIN/DEV/TEST SETS CỦA UIT-ViSFD DATASET

	Train	Dev	Test
Number of Comments	7,786	1,112	2,224
Number of Tokens	283,460	39,023	80,787
Number of Aspects	23,597	3,371	6,742
Average number of aspects per sentence	3.3	3.2	3.3
Average length per sentence	36.4	35.1	36.3

- **Duplicate**: Sao chép các câu chứa khía cạnh STORAGE để tăng số lượng mẫu thuộc nhãn này trong tập huấn luyện.
- **Sử dụng mô hình ngôn ngữ lớn (LLMs)**: Sử dụng Gemma2:2b [18] để tạo thêm dữ liệu chỉ chứa khía cạnh STORAGE.

Với phương pháp Duplicate, nhóm thực hiện nhân bản 10 lần các câu có nhãn chứa khía cạnh STORAGE nhằm mục đích giúp cho mô hình giảm bớt sự chú ý đến các nhãn khác trong quá trình huấn luyện.

Tiếp theo đó, nhằm mục đích gia tăng dữ liệu kết hợp với gia tăng về ngữ cảnh học của mô hình, nhóm sử dụng mô hình gemma2:2b để sinh thêm dữ liệu. Cụ thể về Prompt và kết quả được hiện thị trong bảng II

Bảng II
VÍ DỤ PROMPT DÙNG ĐỂ SINH THÊM MẪU DATA

Input Prompt	Trả lời bằng tiếng Việt. Độ dài có thể thoải mái từ 50 tới đa đến 70 từ. Sinh ra câu KHÔNG có các kí tự đặc biệt và emoji. Có thể viết văn phong vui vẻ, nghiêm túc hoặc liên quan nhiều đến cảm xúc. Hãy tạo 50 nhận xét của khách hàng về khía cạnh bộ nhớ (storage) của một chiếc điện thoại vừa mua. Chỉ sinh ra các câu nhận xét mà không sinh ra bất kì gì khác. Tập trung vào dung lượng, tốc độ, hoặc hiệu suất lưu trữ của chiếc điện thoại. Có thể là nhận xét tích cực hoặc tiêu cực. Các nhận xét đa dạng về văn phong, ngữ nghĩa, độ tuổi, giới tính, tính cách và chủ đề dành cho khả năng lưu trữ.
ExOutput	Để dàng lưu trữ tất cả những app yêu thích của mình. Bộ nhớ lớn, giúp sử dụng điện thoại hiệu quả hơn. Tối ưu hóa dung lượng, lưu trữ mọi thứ mà không lo đầy. Dung lượng lớn, để dàng tải app yêu thích.

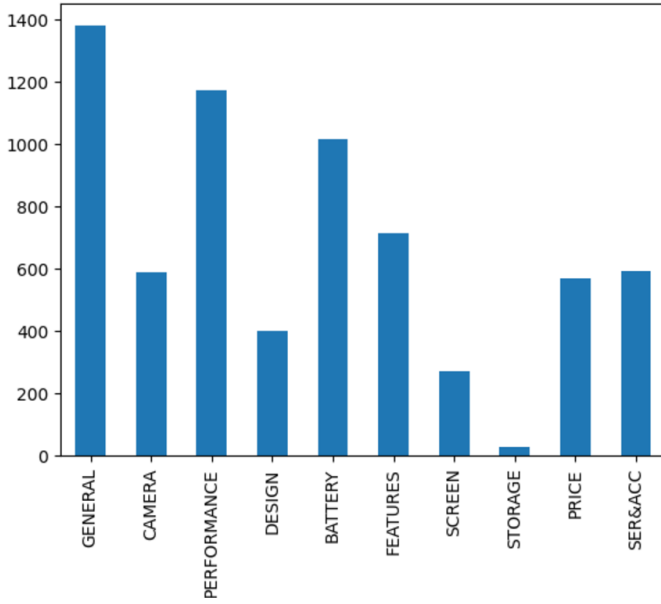
Chi tiết về tính hiệu quả của 2 phương pháp tăng cường dữ liệu được thể hiện ở bảng kết quả III.

IV. PROPOSED METHOD

Nhiệm vụ của Aspect-based Sentiment Analysis có thể đưa về các bài toán phân loại. Cho một câu đánh giá của người dùng, chúng ta cần tìm các khía cạnh được nói đến trong câu và phân loại nó vào một trong ba loại: positive (tích cực), negative (tiêu cực), neutral (trung lập). Từ đó, nhóm xây dựng 2 mô hình con:

- **Aspect Detection model**: cho tác vụ Nhận diện khía cạnh (Multi-label classification).
- **Sentiment Classification model**: cho tác vụ Phân loại cảm xúc trên mỗi khía cạnh (Multi-class classification).

²Data available at: <https://github.com/LuongPhan/UIT-ViSFD>



Hình 3. Phân phối số lượng dữ liệu theo nhân khía cạnh (aspect)

Hình 1 thể hiện tổng quan kiến trúc và luồng hoạt động của toàn bộ mô hình. Chi tiết về hai mô hình con được minh họa ở hình 2.

A. Mô hình MinPhoBERT

Mô hình MinPhoBERT được xây dựng trên nền tảng của PhoBERT [13], một mô hình RoBERTa đã được pre-trained trên bộ dữ liệu tiếng Việt. Để giảm độ phức tạp và chi phí tính toán, nhóm nghiên cứu đã thực hiện cải tiến bằng cách giảm số lớp encoder từ 12 lớp xuống còn 6 lớp. Việc này giúp cải thiện tốc độ của mô hình, mặc dù có thể dẫn đến một sự suy giảm nhỏ về độ chính xác.

MinPhoBERT là thành phần cốt lõi của hai mô hình: Aspect Detection model và Sentiment Classification model. Trong cả hai mô hình này, MinPhoBERT thực hiện vai trò trích xuất đặc trưng ngữ nghĩa từ các câu, tạo vector biểu diễn cho quá trình phân loại.

B. Học Tương Phản (Contrastive Learning)

Học tương phản giúp cải thiện sự khác biệt giữa các embedding bằng cách tăng tính tương đồng giữa các phiên bản khác nhau của cùng một câu và giảm tương đồng với các câu khác trong batch. Điều này tạo ra không gian embedding mạch lạc, phù hợp cho các tác vụ phân loại.

Quá trình huấn luyện MinPhoBERT với contrastive learning được kết hợp vào trong quá trình fine-tuning của mô hình Aspect Detection và Sentiment Classification như sau:

a) *Mô hình Aspect Detection*: Nhóm kết hợp cả Unsupervised Contrastive loss và Cross-entropy loss vào hàm mất mát tổng thể. Contrastive loss giúp cải thiện các embedding của MinPhoBERT, đảm bảo rằng các embedding của các câu tương đồng về mặt ngữ nghĩa sẽ có khoảng cách gần nhau trong không gian embedding, trong khi đó các câu khác nhau

sẽ được phân biệt rõ ràng. Cross-entropy loss được sử dụng để tối ưu hóa mô hình cho tác vụ phân loại khía cạnh.

Nhằm cải thiện khả năng biểu diễn của mô hình MinPhoBERT, nhóm áp dụng kỹ thuật học tương phản (contrastive learning). Cụ thể, đối với embedding của token [CLS], nhóm sử dụng một số kỹ thuật tăng cường (augmentation) để tạo các phiên bản augmented view, bao gồm:

- **Feature Noise**: thêm một lượng nhiễu ngẫu nhiên vào embedding
- **Feature Dropout**: ngẫu nhiên bỏ một số giá trị trên embedding

Hàm mất mát được sử dụng cho quá trình **Unsupervised Contrastive Learning** được định nghĩa như sau:

$$\mathcal{L}_{C_{asp}} = \frac{1}{|\mathcal{B}|} \sum_{i \in \mathcal{B}} -\log \frac{\exp(\cos(s'_i, s''_i)/\tau)}{\sum_{j \in \mathcal{B}} \exp(\cos(s'_i, s''_j)/\tau)} \quad (1)$$

Trong đó:

- s'_i, s''_i là dữ liệu tăng cường của embedding [CLS] sau khi đi dùng các phương pháp tăng cường
- \mathcal{B} là tập hợp các câu trong một batch
- \cos là hàm cosine similarity
- τ là siêu tham số temperature giúp điều chỉnh độ sắc nét của phân phối xác suất.

Hàm mất mát tổng thể trong giai đoạn này được tính theo công thức sau:

$$\mathcal{L}_{\text{Aspect Detection}} = \mathcal{L}_{\text{BCE}} + \lambda_{asp} \mathcal{L}_{C_{asp}} \quad (2)$$

Trong đó:

- \mathcal{L}_{BCE} là hàm mất mát binary cross-entropy cho tác vụ phân loại khía cạnh.
- λ_{asp} là siêu tham số cân bằng giữa các thành phần.

b) *Mô hình Sentiment Classification*: Sau khi huấn luyện Aspect Detection model, nhóm sử dụng MinPhoBERT đã học được từ giai đoạn trên và tiếp tục fine-tune mô hình cho tác vụ Sentiment Classification. Trong giai đoạn này, chúng tôi tiếp tục sử dụng Contrastive Loss trong bài toán Supervised[7].

C. Feature Extraction module

Module này giúp trích xuất các đặc trưng và trả về vector biểu diễn (representation) của câu đầu vào. Luồng hoạt động của module được minh họa ở hình 2 và chi tiết như sau:

a) *Pre-processing*: Câu đầu vào x được xử lý thông qua tokenizer của PhoBERT để chuyển đổi thành chuỗi các token đầu vào:

$$\mathbf{X} = \text{Tokenizer}(x), \quad (3)$$

với $\mathbf{X} = \{[\text{CLS}], t_1, t_2, \dots, t_n, [\text{SEP}]\}$, trong đó t_i là các token và [CLS], [SEP] là các token đặc biệt.

b) *Trích xuất đặc trưng*: Chuỗi token \mathbf{X} được đưa vào mô hình MinPhoBERT để trích xuất đặc trưng ngữ nghĩa. Đầu ra của mô hình là:

$$\mathbf{H} = \text{MinPhoBERT}(\mathbf{X}) \quad (4)$$

với $\mathbf{H} = \{\mathbf{h}_{[\text{CLS}]}, \mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_n\}$, trong đó $\mathbf{h}_{[\text{CLS}]}$ là vector biểu diễn của token [CLS] và d là kích thước của không gian ẩn (hidden size).

c) *Biểu diễn câu*: Vector $\mathbf{h}_{[\text{CLS}]}$ được chọn làm biểu diễn của toàn bộ câu đầu vào, với giả định rằng token [CLS] chứa thông tin tổng quát về nghĩa của câu. Do đó, biểu diễn của câu được xác định là:

$$\mathbf{v} = \mathbf{h}_{[\text{CLS}]} \quad (5)$$

Vector \mathbf{v} là output của module này, được sử dụng trong các tác vụ phân loại trong mô hình.

D. Aspect Detection model

Để trích xuất các khía cạnh (aspects) được đề cập trong câu đầu vào, nhóm xây dựng mô hình Aspect Detection dựa trên MinPhoBERT. Luồng hoạt động của mô hình được tóm tắt trong hình 2 và mô tả chi tiết như sau:

a) *Trích xuất đặc trưng*: Câu đầu vào x được xử lý thông qua Feature Extraction module được định nghĩa ở phần IV-C:

$$\mathbf{v} = \text{FeatureExtractor}(x) \quad (6)$$

b) *Dự đoán khía cạnh*: Vector biểu diễn \mathbf{v} được đưa qua một mạng Multi layer perceptron (MLP) với kiến trúc cụ thể như sau:

$$\begin{aligned} \mathbf{z}_1 &= \text{ReLU}(\mathbf{W}_1 \mathbf{v} + \mathbf{b}_1), \\ \mathbf{z}'_1 &= \text{Dropout}(\mathbf{z}_1), \\ \mathbf{z}_2 &= \mathbf{W}_2 \mathbf{z}'_1 + \mathbf{b}_2. \end{aligned} \quad (7)$$

Trong đó:

- $\mathbf{W}_1 \in \mathbb{R}^{h \times d}$ và $\mathbf{b}_1 \in \mathbb{R}^h$ là các tham số của lớp ẩn (hidden layer) với h là số chiều của lớp ẩn.
- $\mathbf{W}_2 \in \mathbb{R}^{k \times h}$ và $\mathbf{b}_2 \in \mathbb{R}^k$ là các tham số của lớp đầu ra, với k là số khía cạnh cần dự đoán

c) *Hàm sigmoid và quyết định dự đoán*: Đầu ra \mathbf{z}_2 của MLP được đưa qua hàm kích hoạt sigmoid để tính xác suất cho mỗi khía cạnh:

$$\mathbf{p} = \sigma(\mathbf{z}_2), \text{ với } \sigma(x) = \frac{1}{1 + e^{-x}} \quad (8)$$

Các khía cạnh được dự đoán dựa trên một ngưỡng (threshold):

$$a_i = \begin{cases} 1 & \text{nếu } p_i \geq t \\ 0 & \text{nếu } p_i < t \end{cases} \quad (9)$$

với a_i là nhãn dự đoán cho khía cạnh thứ i .

d) *Hàm mất mát*: Để huấn luyện mô hình, ta sử dụng hàm Binary Cross Entropy loss (BCELoss) kết hợp với hàm Contrastive loss 1, với BCE Loss được định nghĩa như sau:

$$\mathcal{L}_{BCE} = -\frac{1}{N} \sum_{i=1}^N [a_i \log(p_i) + (1 - a_i) \log(1 - p_i)] \quad (10)$$

với:

- N là số lượng khía cạnh (aspects),
- a_i là nhãn thực tế (ground truth) của khía cạnh thứ i , với giá trị $a_i \in \{0, 1\}$,
- p_i là xác suất dự đoán cho khía cạnh thứ i , được tính từ đầu ra của hàm sigmoid.

Hàm mất mát này đo lường sự khác biệt giữa xác suất dự đoán p_i và nhãn thực tế a_i . Mục tiêu của mô hình là tối thiểu hóa hàm mất mát tổng thể:

$$\mathcal{L}_{asp} = \mathcal{L}_{BCE} + \lambda_{asp} \mathcal{L}_{C_{asp}} \quad (11)$$

Trong đó:

- \mathcal{L}_{BCE} là hàm binary cross-entropy loss 10 cho tác vụ phân loại khía cạnh.
- \mathcal{L}_C là hàm Contrastive loss 1 cho tối ưu embedding của MinPhoBERT.
- λ_{asp} là siêu tham số cân bằng giữa hai thành phần.

E. Sentiment Classification model

Với mỗi khía cạnh (aspect) được dự đoán từ mô hình Aspect Detection. Để dự đoán cảm xúc (sentiment) của câu về khía cạnh đó, nhóm xây dựng mô hình Sentiment Classification dựa trên MinPhoBERT. Luồng hoạt động của mô hình được tóm tắt trong hình 2 và chi tiết như sau:

a) *Input model*: Đầu vào của mô hình bao gồm câu và aspect cần dự đoán sentiment. Hai thành phần này được nối lại tạo thành đầu vào cho mô hình:

$$x = [\text{CLS}] \parallel \text{sentence} \parallel [\text{SEP}] \parallel \text{aspect} \parallel [\text{SEP}] \quad (12)$$

b) *Trích xuất đặc trưng*: Câu đầu vào x được xử lý thông qua module trích xuất đặc trưng được định nghĩa ở phần IV-C:

$$\mathbf{v} = \text{FeatureExtractor}(x) \quad (13)$$

c) *Dự đoán sentiment*: Vector biểu diễn \mathbf{v} được đưa qua một mạng Multi Layer Perceptron (MLP) với kiến trúc cụ thể như sau:

$$\begin{aligned} \mathbf{z}_1 &= \text{ReLU}(\mathbf{W}_1 \mathbf{v} + \mathbf{b}_1), \\ \mathbf{z}'_1 &= \text{Dropout}(\mathbf{z}_1), \\ \mathbf{z}_2 &= \mathbf{W}_2 \mathbf{z}'_1 + \mathbf{b}_2. \end{aligned} \quad (14)$$

Trong đó:

- $\mathbf{W}_1 \in \mathbb{R}^{h \times d}$ và $\mathbf{b}_1 \in \mathbb{R}^h$ là các tham số của lớp ẩn (hidden layer) với h là số chiều của lớp ẩn.
- $\mathbf{W}_2 \in \mathbb{R}^{k \times h}$ và $\mathbf{b}_2 \in \mathbb{R}^k$ là các tham số của lớp đầu ra, với k là số lớp cảm xúc cần phân loại (3 lớp: Positive, Negative, Neutral).

d) *Hàm softmax*: Đầu ra \mathbf{z}_2 của MLP được đưa qua hàm kích hoạt Softmax để tính xác suất cho dự đoán sentiment của khía cạnh:

$$\hat{s} = \text{Softmax}(\mathbf{z}_2) \quad (15)$$

với $\hat{s} = \{\text{Positive}, \text{Negative}, \text{Neutral}\}$ là các cảm xúc cần phân loại, và hàm Softmax chuyển đổi đầu ra thành các xác suất cho từng cảm xúc.

Để xác định cảm xúc dự đoán cuối cùng cho khía cạnh, ta chọn cảm xúc có xác suất cao nhất từ kết quả của hàm Softmax. Cụ thể, cảm xúc dự đoán được tính như sau:

$$\hat{y} = \arg \max(\hat{s}) \quad (16)$$

Trong đó, \hat{y} là chỉ số của cảm xúc có xác suất lớn nhất trong tập $\{\text{Positive}, \text{Negative}, \text{Neutral}\}$. Dựa trên chỉ số này, ta xác định được cảm xúc dự đoán là Positive, Negative hoặc Neutral.

e) *Hàm mất mát Cross-Entropy*: Để huấn luyện mô hình, ta sử dụng hàm mất mát Cross-Entropy cho bài toán phân loại nhiều lớp, được định nghĩa như sau:

$$\mathcal{L}_{CE} = - \sum_{i=1}^K y_i \log(p_i) \quad (17)$$

Trong đó:

- K là số lớp (trong trường hợp này là 3: Positive, Negative, Neutral),
- y_i là nhãn thực tế cho cảm xúc của khía cạnh ($y_i \in \{0, 1\}$, với $y_i = 1$ là lớp đúng),
- p_i là xác suất dự đoán cho lớp cảm xúc thứ i , được tính từ hàm Softmax.

Tương tự như Aspect Detection model, chúng tôi cũng áp dụng contrastive learning để cải thiện cả năng học. Ý tưởng đơn giản đó là những câu cùng nhãn sẽ được kéo lại gần nhau, đồng thời đẩy những câu khác nhãn ra nhau. Ví dụ: những câu mang nhãn positive sẽ ở xa những câu có nhãn negative. Dựa theo [7], hàm loss được triển khai như sau:

$$\mathcal{L}_{C_{sen}} = \frac{1}{|\mathcal{B}|} \sum_{i \in \mathcal{B}} \frac{-1}{|P(i)|} \sum_{p \in P(i)} \log \frac{\exp(v_i \cdot v_p / \tau)}{\sum_{a \in A(i)} \exp(v_i \cdot v_a / \tau)} \quad (18)$$

Trong đó

- v_i là embedding [CLS] của câu thứ i trong batch
- $P(i)$ là tập hợp chỉ số của các điểm dữ liệu cùng lớp với i
- $A(i)$ là tập hợp các chỉ số khác của các điểm dữ liệu khác i
- Lưu ý: Cả 2 công thức của contrastive loss được viết theo dạng batch

Loss của tác vụ phân loại này là:

$$\mathcal{L}_{sen} = \mathcal{L}_{CE} + \lambda_{sen} \mathcal{L}_{C_{sen}} \quad (19)$$

Mô hình được huấn luyện bằng cách tối thiểu hóa hàm mất mát này.

V. EXPERIMENTS

A. Experimental Settings

Nhóm tiến hành thực nghiệm trên bộ dữ liệu UIT-VISFD [10], một tập dữ liệu chất lượng cao được xây dựng nhằm đánh giá bài toán ABSA (Aspect-Based Sentiment Analysis) trong tiếng Việt. Bên cạnh đó, nhóm sử dụng các độ đo phổ biến như Precision, Recall và F1-score để đánh giá hiệu quả của mô hình trên bài toán phân lớp và phân loại, đảm bảo các kết quả thu được mang tính toàn diện và đáng tin cậy.

1) *Compared Methods*: Vì đặc thù dữ liệu là tiếng Việt, nhóm tiến hành so sánh mô hình với một mô hình SA2SL[10] sử dụng cùng bộ dữ liệu UIT-VISFD. Mô hình này áp dụng phương pháp Bi-LSTM cho bài toán ABSA trong tiếng Việt, đạt hiệu quả cao so với các hệ thống sử dụng các thuật toán học máy cổ điển (Naive Bayes, SVM, và Random Forest) cũng như các mô hình học sâu khác (LSTM và CNN).

2) *Hyper-parameter Settings*: Nhóm đã thực hiện tìm kiếm các bộ siêu tham số tối ưu bằng phương pháp grid search, một cách tiếp cận hệ thống để thử nghiệm nhiều tổ hợp siêu tham số khác nhau nhằm xác định cấu hình hiệu quả nhất cho mô hình. Cụ thể, learning rate được kiểm tra với các giá trị $\{10^{-5}, 2 \times 10^{-5}, 3 \times 10^{-5}, 10^{-4}\}$. Kết quả cho thấy, giá trị 2×10^{-5} thường mang lại hiệu năng tốt nhất, phù hợp với đặc điểm của mô hình và tập dữ liệu huấn luyện. Điều này cho thấy đây là một giá trị ổn định và cân bằng giữa tốc độ hội tụ và độ chính xác.

Đối với các siêu tham số trong contrastive learning, nhóm tập trung vào việc tối ưu hóa các thành phần quan trọng như λ_{asp} , λ_{sen} , và τ . Cụ thể, λ_{asp} và λ_{sen} được tìm kiếm trong tập giá trị $\{0.01, 0.05, 0.1, 0.2, 0.5\}$, đại diện cho trọng số trong hàm mất mát contrastive. Trong khi đó, τ được thử nghiệm với các giá trị $\{0.2, 0.4, 1\}$, nhằm xác định nhiệt độ tối ưu để điều chỉnh mức độ mượt mà trong phân phối xác suất. Kết quả cho thấy các giá trị này có ảnh hưởng đáng kể đến việc cân bằng giữa các thành phần trong quá trình học và cải thiện hiệu năng của mô hình.

Bên cạnh việc điều chỉnh các siêu tham số học, nhóm cũng chú trọng vào quá trình tạo dữ liệu tăng cường nhằm nâng cao chất lượng huấn luyện cho contrastive learning. Trong đó, noise được thêm vào dữ liệu gốc với trọng số $\{0.01, 0.05, 0.1\}$ để tạo ra sự đa dạng trong biểu diễn dữ liệu, trong khi tỷ lệ dropout được thử nghiệm với các giá trị $\{0.1, 0.2, 0.3\}$ để tăng tính tổng quát của mô hình. Các kỹ thuật này không chỉ giúp cải thiện hiệu năng mà còn làm tăng tính mạnh mẽ trước các biến động trong dữ liệu.

Ngoài ra, nhóm đã thực hiện một phân tích chi tiết về sự thay đổi hiệu suất của mô hình khi điều chỉnh số lượng layer của BERT, một yếu tố quan trọng trong việc cân bằng giữa độ phức tạp của mô hình và khả năng tổng quát hóa. Những kết quả thu được từ quá trình thử nghiệm này được trình bày chi tiết trong phần Hyper-parameter Analysis, cung cấp cái nhìn sâu sắc về tác động của các siêu tham số đến hiệu năng tổng thể của mô hình.

Quá trình thử nghiệm không chỉ giúp nhóm xác định các cấu hình tối ưu mà còn đóng vai trò như một cơ sở khoa học để hiểu rõ hơn về cách các siêu tham số ảnh hưởng đến hiệu năng của mô hình trong từng kịch bản cụ thể.

VI. RESULT

A. Overall performance of the proposed methods

Bảng III trình bày kết quả thử nghiệm của các phương pháp được đề xuất trong bài cho hai nhiệm vụ Nhận diện Khía cạnh (Aspect Detection) và Phân loại Cảm xúc (Sentiment Classification). Kết quả cho thấy các phương pháp mà nhóm đề xuất đều mang lại hiệu quả vượt trội so với mô hình cơ sở PhoBERT.

Mô hình chính của nhóm, MinPhoBERT + Duplicate + Contrastive, đạt hiệu năng cạnh tranh trên cả hai nhiệm vụ. Đặc biệt, trên nhiệm vụ Nhận diện Khía cạnh, mô hình chính đạt điểm F1 là 0.87, cao hơn so với PhoBERT base và chỉ thấp hơn một chút so với phiên bản đầy đủ PhoBERT + Duplicate + Contrastive (0.90), nhưng lại sử dụng mô hình nhỏ gọn hơn

với chỉ 6 tầng. Điều này cho thấy phương pháp rút gọn mô hình và sử dụng học tương phản (Contrastive Learning) có thể duy trì hiệu quả trong khi giảm bớt độ phức tạp.

Ngoài ra, các phương pháp tăng cường dữ liệu bằng nhân bản (Duplicate) và kết hợp với học tương phản (Duplicate + Contrastive) đều giúp cải thiện đáng kể điểm F1 trên nhiệm vụ Nhận diện Khía cạnh, đạt mức 0.90, cao hơn so với mô hình cơ sở PhoBERT (0.82). Trên nhiệm vụ Phân loại Cảm xúc, hiệu quả của phương pháp cũng được cải thiện nhẹ so với mô hình cơ sở PhoBERT.

Các phiên bản rút gọn như TinyPhoBERT (4 tầng) và MinPhoBERT (6 tầng) cũng đạt được hiệu năng khá tốt, đặc biệt phù hợp với các hệ thống yêu cầu hiệu suất cao nhưng tài nguyên hạn chế.

Tóm lại, các kết quả cho thấy hiệu quả của việc kết hợp các chiến lược tăng cường dữ liệu và học tương phản trong việc cải thiện hiệu năng của các mô hình ngôn ngữ lớn, đồng thời chứng minh tính khả thi của việc rút gọn mô hình để giảm thiểu tài nguyên mà không làm suy giảm đáng kể chất lượng.

B. Overall Performance Comparison

Bảng IV trình bày chi tiết hiệu năng của mô hình mà chúng tôi đề xuất so với các mô hình baseline trong cả hai tác vụ Aspect Detection và Sentiment Classification. Dựa trên kết quả này, chúng tôi đưa ra một số quan sát và phân tích cụ thể như sau:

Thứ nhất, đối với tác vụ Aspect Detection, mô hình của chúng tôi đạt được hiệu quả cao hơn so với mô hình baseline tốt nhất, cho thấy một sự cải thiện nhất định. Tuy nhiên, mức độ cải thiện này không quá lớn, dẫn đến kết luận rằng hiệu năng của mô hình đề xuất là cạnh tranh với các phương pháp hiện có. Điều này có thể được giải thích bởi đặc điểm của tác vụ này, nơi mà các mô hình baseline đã đạt được một mức hiệu năng khá cao, khiến việc cải thiện trở nên khó khăn hơn. Dẫu vậy, việc mô hình của chúng tôi đạt được hiệu quả tốt hơn vẫn là một tín hiệu tích cực, chứng minh rằng phương pháp đề xuất có tiềm năng ứng dụng trong thực tế.

Thứ hai, về tác vụ Sentiment Classification, mô hình của chúng tôi thể hiện sự vượt trội rõ rệt so với các mô hình baseline. Điều này đặc biệt quan trọng vì tác vụ này yêu cầu khả năng nắm bắt tốt các ngữ cảnh phức tạp và mối quan hệ giữa các yếu tố ngữ nghĩa trong câu. Kết quả này cho thấy mô hình đề xuất không chỉ cải thiện khả năng nhận diện cảm xúc mà còn nắm bắt được các sắc thái ý nghĩa, từ đó nâng cao hiệu quả phân tích cảm xúc trong các trường hợp thực tế.

Tổng quan, so với các mô hình baseline, mô hình đề xuất của chúng tôi có thể được xem là ưu việt hơn trên cả hai tác vụ, đặc biệt là ở tác vụ Sentiment Classification. Sự cải thiện này không chỉ phản ánh khả năng mô hình hóa tốt hơn mà còn chứng minh tính hiệu quả của các cải tiến mà chúng tôi áp dụng trong kiến trúc mô hình và quá trình huấn luyện. Những kết quả này mở ra nhiều triển vọng cho việc áp dụng mô hình vào các bài toán thực tế, nơi mà sự chính xác và khả năng hiểu ngữ cảnh đóng vai trò then chốt.

C. Ablation Study

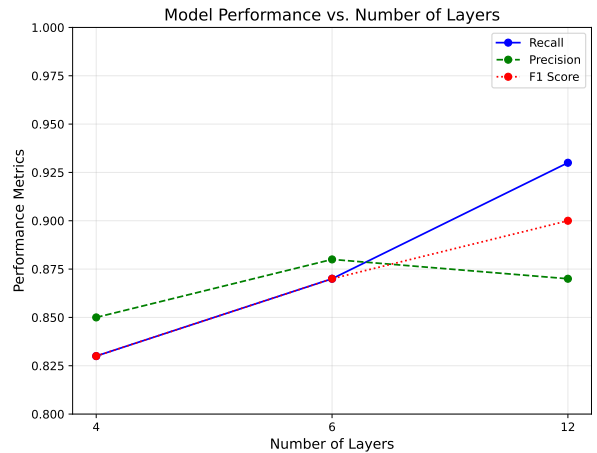
Để đánh giá chi tiết vai trò của các thành phần quan trọng trong mô hình, nhóm đã thực hiện thử nghiệm loại bỏ từng thành phần cụ thể và phân tích mức độ ảnh hưởng của chúng đến hiệu quả tổng thể. Cụ thể:

- w/o CL-Stage-1: Ở bước này, nhóm loại bỏ phương pháp contrastive learning tại giai đoạn Aspect Detection. Thay vì sử dụng các cặp mẫu đối lập để học các biểu diễn đặc trưng tốt hơn, mô hình chỉ được huấn luyện dựa trên các mục tiêu tiêu chuẩn của tác vụ.
- w/o CL-Stage-2: Ở bước này, nhóm loại bỏ thành phần unsupervised contrastive loss trong quá trình huấn luyện tác vụ Sentiment Classification.

Kết quả chi tiết của các thí nghiệm này được trình bày trong bảng V. Qua các phân tích, nhóm nhận thấy rằng việc sử dụng contrastive learning không chỉ cải thiện hiệu suất riêng lẻ trên từng tác vụ mà còn giúp mô hình học được các biểu diễn đặc trưng tốt hơn, từ đó nâng cao hiệu năng tổng thể. Điều này cho thấy contrastive learning đóng vai trò cốt lõi, giúp mô hình vượt trội hơn đáng kể so với các phương pháp truyền thống, đặc biệt trong các tình huống dữ liệu hạn chế hoặc không đồng nhất.

D. Model Study

1) *Hyper-parameter Analysis*: Để phân tích tầm quan trọng của số lớp trong mô hình PhoBERT, nhóm đã nghiên cứu sự thay đổi của các chỉ số đo lường. Trong phần này, nhóm lựa chọn tác vụ phát hiện khía cạnh (Aspect Detection) để làm ví dụ. Kết quả trình bày trong hình 4 cho thấy số lớp vẫn có ảnh hưởng đến hiệu suất của mô hình. Điều này cho thấy cần phải cân nhắc kỹ lưỡng sự đánh đổi giữa chi phí tính toán và kích thước mô hình, sao cho phù hợp với mục đích sử dụng.



Hình 4. Sự thay đổi qua số lượng lớp

2) *Impact Of Data Sampling*: Để nhận thức rõ hơn về tầm quan trọng của các phương pháp oversampling trong việc xử lý vấn đề mất cân bằng nhãn trong dữ liệu, chúng tôi đã tiến hành đánh giá hiệu năng của mô hình trước và sau khi áp

Bảng III
SO SÁNH HIỆU NĂNG CÁC PHƯƠNG PHÁP ĐỀ XUẤT

Model	Aspect Detection			Sentiment Classification		
	P	R	F	P	R	F
PhoBERT	0.89	0.84	0.84	0.86	0.86	0.86
PhoBERT + LLM gen	0.91	0.83	0.86	-	-	-
PhoBERT + Duplicate	0.89	0.87	0.88	-	-	-
PhoBERT + Duplicate + Contrastive	0.87	0.93	0.90	0.85	0.87	0.86
MinPhoBERT + Duplicate + Contrastive*	0.88	0.87	0.87	0.84	0.81	0.82
TinyPhoBERT + Duplicate + Contrastive	0.85	0.83	0.83	-	-	-

* Mô hình chính của nhóm.

Chú thích:

- PhoBERT: mô hình baseline.
- PhoBERT + LLM gen: tăng mẫu bằng mô hình gemma2:2b.
- PhoBERT + Duplicate: tăng mẫu bằng nhân bản.
- PhoBERT + Duplicate + Contrastive sử dụng thêm Contrastive Learning.
- MinPhoBERT và TinyPhoBERT là các phiên bản rút gọn của PhoBERT với lần lượt 6 và 4 layers.

Bảng IV
HIỆU NĂNG MÔ HÌNH VỚI CÁC NGHIÊN CỨU TRƯỚC ĐÓ

Model	Aspect Detection			Sentiment Classification		
	P	R	F	P	R	F
Naive Bayes	0.72	0.59	0.64	0.49	0.30	0.37
SVM	0.36	0.51	0.42	0.16	0.23	0.19
Random Forest	0.45	0.50	0.47	0.17	0.24	0.20
CNN	0.77	0.63	0.69	0.33	0.22	0.27
LSTM	0.82	0.78	0.80	0.56	0.48	0.52
SA2SL [10]	<u>0.87</u>	<u>0.83</u>	<u>0.84</u>	<u>0.65</u>	<u>0.60</u>	<u>0.63</u>
Ours	0.88	0.87	0.87	0.84	0.81	0.82
%Improve	1.15%	4.82%	3.57%	29.23%	35.00%	30.16%

P: Precision, R: Recall, F: F1-score

Các số liệu tính theo macro weight

Bảng V
ABLATION STUDY VỚI CONTRASTIVE LEARNING

Model	Aspect Detection			Sentiment Classification		
	P	R	F	P	R	F
w/o CL-Stage-1	<u>0.85</u>	<u>0.83</u>	<u>0.83</u>	-	-	-
w/o CL-Stage-2	-	-	-	0.83	0.76	0.77
Ours	0.88	0.87	0.87	0.84	0.81	0.82

P: Precision, R: Recall, F: F1-score

Các số liệu tính theo macro weight

dụng kỹ thuật này. Quan sát rõ hơn ở bảng III. Oversampling, trong đó các mẫu thuộc các nhãn thiểu số được nhân lên để cân bằng tỷ lệ giữa các nhãn, được sử dụng như một cách tiếp cận hiệu quả nhằm cải thiện khả năng học của mô hình trên các nhãn có ít dữ liệu.

Trước khi thực hiện oversampling, tập dữ liệu gốc có sự chênh lệch lớn về số lượng mẫu giữa các nhãn cụ thể là nhãn STORAGE với số lượng ít hơn nhiều so với các nhãn khác. Điều này dẫn đến hiện tượng mất cân bằng nhãn, khiến mô hình có xu hướng thiên vị các nhãn chiếm ưu thế. Các kết quả ban đầu cho thấy độ chính xác tổng thể của mô hình khá cao, nhưng khi phân tích kỹ hơn qua các chỉ số như Recall và F1-score trên nhãn thiểu số, hiệu năng lại rất thấp. Điều này cho thấy mô hình không đủ khả năng nhận diện tốt các trường hợp thuộc nhãn thiểu số, dẫn đến hiệu suất không đồng đều trên toàn bộ tập dữ liệu.

Sau khi áp dụng oversampling, tỷ lệ mẫu giữa các nhãn đã được cân bằng, giúp cải thiện đáng kể hiệu năng của mô hình. Quá trình này mang lại hai lợi ích chính: (1) mô hình có cơ hội tiếp xúc nhiều hơn với các mẫu thuộc nhãn thiểu số trong quá trình huấn luyện, và (2) giảm nguy cơ mất cân bằng trong việc tối ưu hóa, giúp mô hình học tốt hơn ở cả các nhãn chiếm ưu thế lẫn nhãn thiểu số. Kết quả cho thấy Recall của các nhãn thiểu số được cải thiện rõ rệt, điều này đồng nghĩa với việc mô hình có khả năng phát hiện chính xác hơn các trường hợp thuộc nhãn này. Đồng thời, F1-score cũng tăng lên đáng kể, phản ánh sự cân bằng tốt hơn giữa Precision và Recall.

Cụ thể, các thử nghiệm cho thấy hiệu quả của oversampling càng rõ ràng hơn khi các nhãn thiểu số có tỷ lệ cực thấp so với nhãn chiếm ưu thế. Hiệu năng cải thiện đồng đều ở cả các chỉ số chi tiết như Macro F1-score và Weighted F1-score, cho thấy mô hình không chỉ hoạt động tốt hơn trên các nhãn thiểu số mà còn duy trì hiệu quả trên toàn bộ tập dữ liệu.

Kỹ thuật oversampling cũng được sử dụng để tạo dữ liệu tăng cường, như nhân bản các mẫu gốc kèm theo noise nhỏ hoặc các biến đổi khác, ở đây là việc sử dụng LLMs để tăng cường dữ liệu, nhằm tăng độ đa dạng và tính tổng quát của dữ liệu huấn luyện. Điều này đặc biệt hiệu quả trong các trường hợp dữ liệu ban đầu không chỉ mất cân bằng mà còn thiếu sự đa dạng về ngữ cảnh.

VII. CONCLUSION AND FUTURE WORK

Trong bài báo này, chúng tôi đề xuất mô hình ABSA có khả năng dự đoán các khía cạnh trên bộ dữ liệu UIT-VISFD [10], bao gồm các bình luận Tiếng Việt và các thông tin về khía cạnh được đề cập và cảm xúc trên từng khía cạnh. Các đóng góp chính của chúng tôi gồm 4 phần. Thứ nhất, chúng tôi đề xuất sử dụng pretrained-model *PhoBERT_{base}* được huấn luyện trên dữ liệu tiếng Việt cùng với việc lựa chọn số layer để tăng cường độ hiệu quả của mô hình baseline. Thứ hai, chúng tôi sử dụng Contrastive Learning với mục đích tăng cường độ mạnh mẽ của các vector embedding. Thứ ba, với việc unbalanced data trên nhãn STORAGE, chúng tôi đưa ra 2 giải pháp là duplicate dữ liệu nhãn STORAGE và sử dụng LLM để tăng cường data, tránh bias trong quá trình training. Kết quả dựa trên các giải pháp của chúng tôi đưa ra

vượt trội hơn hẳn so với các cách tiếp cận của bài báo [10], cụ thể là đạt được 87% trên tác vụ Aspect Detection và 82% trên tác vụ Sentiment Classification dựa vào F1-score. Kết quả đạt được cho thấy sự cải thiện rõ rệt của hiệu năng mô hình, đồng thời chứng tỏ rằng việc áp dụng pretrained model *PhoBERT_{base}* mang lại hiệu quả to lớn trong việc hiểu ngữ cảnh tiếng Việt. Các nghiên cứu trong giải pháp không chỉ mang tính lý thuyết, mà còn mang đến một phương pháp giải quyết đầy hiệu quả cho bài toán ABSA khi tận dụng được pretrained model *PhoBERT* kết hợp với Contrastive Learning mang đến hiệu suất mạnh mẽ trên dữ liệu tiếng Việt. Mặc dù các kết quả đạt được đã có những đóng góp cụ thể như đã trình bày ở trên, song bên cạnh đó vẫn còn nhiều hạn chế cần khắc phục. Thứ nhất, khắc phục unbalanced data bằng LLMs còn có nhiều hạn chế. Việc hoàn toàn phụ thuộc vào LLMs và prompt sử dụng có thể đưa ra các bộ dữ liệu ít đa dạng về ngữ cảnh, cũng như văn phong chưa thực sự giống con người. Để khắc phục vấn đề này, chúng tôi mong muốn sẽ tự thu thập data trên Internet và đánh nhãn, từ đó sẽ mang lại hiệu quả cao hơn. Tiếp theo đó, hạn chế thứ hai mà chúng tôi nhận thấy là mô hình BERT nói chung cũng như *PhoBERT* nói riêng hạn chế trong việc học biểu diễn ngữ nghĩa dài. Khi đó, các văn bản dài thường phải cắt câu, dẫn đến thiếu biểu diễn ngữ cảnh toàn diện. Trong tương lai, nhóm mong muốn thử nghiệm thêm các mô hình khác như *bi-Bird* [11], một mô hình với kiến trúc Transformer giúp biểu diễn các câu dài với mong muốn cải thiện biểu diễn ngữ cảnh cho embedding.

REFERENCES

- [1] Mostafa M Amin et al. “A wide evaluation of ChatGPT on affective computing tasks”. In: *IEEE Transactions on Affective Computing* (2024).
- [2] Ting Chen et al. “A simple framework for contrastive learning of visual representations”. In: *International conference on machine learning*. PMLR, 2020, pp. 1597–1607.
- [3] Jacob Devlin. “Bert: Pre-training of deep bidirectional transformers for language understanding”. In: *arXiv preprint arXiv:1810.04805* (2018).
- [4] Tianyu Gao, Xingcheng Yao, and Danqi Chen. “Simcse: Simple contrastive learning of sentence embeddings”. In: *arXiv preprint arXiv:2104.08821* (2021).
- [5] S Hochreiter. “Long Short-term Memory”. In: *Neural Computation MIT-Press* (1997).
- [6] Bamba Kane et al. “CNN-LSTM-CRF for Aspect-Based Sentiment Analysis: A Joint Method Applied to French Reviews.” In: *ICAART (1)*. 2021, pp. 498–505.
- [7] Prannay Khosla et al. “Supervised contrastive learning”. In: *Advances in neural information processing systems* 33 (2020), pp. 18661–18673.
- [8] M Hari Krishna, K Rahamathulla, and Ali Akbar. “A feature based approach for sentiment analysis using SVM and coreference resolution”. In: *2017 International conference on inventive communication and computational technologies (ICICCT)*. IEEE, 2017, pp. 397–399.

Bảng VI
THE ANNOTATION GUIDELINES FOR LABELING THE ASPECTS AND THEIR SENTIMENT.

Aspect	Mean	User comments	Sentiment
SCREEN	User comments express screen quality, size, colors, and display technology.	màn hình đẹp (a nice screen)	Pos
CAMERA	The comments mention the quality of a camera, vibration, delay, focus, and image colors.	điện thoại chụp hình mờ (the phone took blur picture)	Neg
FEATURES	The users refer to features, fingerprint sensor, wifi connection, touch and face detection of the phone.	nhận diện khuôn mặt chậm (the face detection is slow)	Neg
BATTERY	The comment describes battery capacity and battery quality.	pín trâu (long battery life)	Pos
PERFORMANCE	The reviews describe ramming capacity, processor chip, performance using, and smoothness of the phone.	cấu hình có thể chấp nhận được (acceptable configuration)	Neu
STORAGE	The comment mentions storage capacity, the ability to expand capacity through memory cards.	bộ nhớ lớn (large storage)	Pos
DESIGN	The reviews refer to the style, design, and shell.	điện thoại thiết kế thô (rough design phone)	Neg
PRICE	The comments present the specific price of the phone.	giá cả ở mức trung bình (the price is at average)	Neu
GENERAL	The reviews of customers generally comment about the phone.	mọi thứ đều ok (everything is ok)	Pos
SER&ACC	The comments mention sales service, warranty, and review of accessories of the phone.	nhân viên tư vấn nhiệt tình (shop assistants advice enthusiastic)	Pos

- [9] Yinhan Liu. “Roberta: A robustly optimized bert pre-training approach”. In: *arXiv preprint arXiv:1907.11692* (2019).
- [10] Luong Luc Phan et al. “Sa2sl: From aspect-based sentiment analysis to social listening system for business intelligence”. In: *Knowledge Science, Engineering and Management: 14th International Conference, KSEM 2021, Tokyo, Japan, August 14–16, 2021, Proceedings, Part II 14*. Springer. 2021, pp. 647–658.
- [11] Guru Guruganesh Manzil Zaheer. “Big Bird: Transformers for Longer Sequences”. In: *arXiv preprint arXiv:2007.14062* (2021).
- [12] AL-Smadi Mohammad et al. “An enhanced framework for aspect-based sentiment analysis of Hotels’ reviews: Arabic reviews case study”. In: *2016 11th International conference for internet technology and secured transactions (ICITST)*. IEEE. 2016, pp. 98–103.
- [13] Dat Quoc Nguyen and Anh Tuan Nguyen. “PhoBERT: Pre-trained language models for Vietnamese”. In: *arXiv preprint arXiv:2003.00744* (2020).
- [14] Maria Pontiki and associates. “SemEval-2014 Task 4: Aspect Based Sentiment Analysis”. In: *Association for Computational Linguistics* 8 (2014).
- [15] Maria Pontiki and associates. “SemEval-2016 Task 5: Aspect Based Sentiment Analysis”. In: *Association for Computational Linguistics* (2016).
- [16] J Shobana and M Murali. “An efficient sentiment analysis methodology based on long short-term memory networks”. In: *Complex & Intelligent Systems* 7.5 (2021), pp. 2485–2501.
- [17] Huizhe Su et al. “Enhanced implicit sentiment understanding with prototype learning and demonstration for aspect-based sentiment analysis”. In: *IEEE Transactions on Computational Social Systems* (2024).
- [18] Gemma Team et al. “Gemma: Open models based on gemini research and technology”. In: *arXiv preprint arXiv:2403.08295* (2024).
- [19] A Vaswani. “Attention is all you need”. In: *Advances in Neural Information Processing Systems* (2017).
- [20] Xiaodi Wang et al. “A novel network with multiple attention mechanisms for aspect-level sentiment analysis”. In: *Knowledge-based systems* 227 (2021), p. 107196.
- [21] Hua Xu, Fan Zhang, and Wei Wang. “Implicit feature identification in Chinese reviews using explicit topic mining model”. In: *Knowledge-Based Systems* 76 (2015), pp. 166–175.
- [22] Wenxuan Zhang et al. “A Survey on Aspect-Based Sentiment Analysis: Tasks, Methods, and Challenges”. In: *IEEE Transactions on Knowledge and Data Engineering* 35.11 (2023), pp. 11019–11038.
- [23] Changzhi Zhou et al. “Span-Pair Interaction and Tagging for Dialogue-Level Aspect-Based Sentiment Quadruple Analysis”. In: *Proceedings of the ACM on Web Conference 2024*. 2024, pp. 3995–4005.