ĐẠI HỌC QUỐC GIA TP. HCM
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN

ĐẶNG VĂN THÌN

NGHIÊN CỨU CÁC KỸ THUẬT PHÂN TÍCH Ý KIẾN TRÊN BÌNH LUẬN PHẢN HỒI CỦA NGƯỜI DÙNG

Ngành: Khoa học máy tính
Mã số: 9 48 48 01 01

TÓM TẮT LUẬN ÁN TIẾN SĨ
NGÀNH KHOA HỌC MÁY TÍNH

TP HỒ CHÍ MINH – Năm 2025

Công trình được hoàn thành tại:

Trường Đại học Công nghệ Thông tin – ĐHQG TPHCM

Người hướng dẫn khoa học:

PGS.TS Nguyễn Lưu Thùy Ngân

Phản biện 1: .....

Phản biện 2: .....

Phản biện 3: .....

Phản biện độc lập 1: .....

Phản biện độc lập 2: .....

Luận án sẽ/đã được bảo vệ trước
Hội đồng chấm luận án cấp Trường tại :

.....

.....

vào lúc giờ ngày tháng năm

Có thể tìm hiểu luận án tại:

- Thư viện Quốc gia Việt Nam
- Thư viện ĐHQG-HCM
- Thư viện Trường Đại học Công nghệ Thông tin

MỤC LỤC

1. GIỚI THIỆU

1.1 LÝ DO THỰC HIỆN ĐỀ TÀI

Sự phát triển bùng nổ dữ liệu tạo ra trên các nền tảng trực tuyến dẫn đến nhu cầu phân tích tự động các bình luận một cách tự động. Luận án được thực hiện với các mục tiêu chính như sau:

1. Phân tích nghiên cứu liên quan đến xác định ưu điểm và nhược điểm trên bài toán phân tích ý kiến cho tiếng Việt.
2. Phát triển các phương pháp để nâng cao độ chính xác cho bài toán phân tích ý kiến theo mức độ và quan điểm.
3. Nghiên cứu các phương pháp cho vấn đề hạn chế dữ liệu huấn luyện ở bài toán phân tích ý kiến theo mức độ.

1.2 MỤC ĐÍCH, ĐỐI TƯỢNG VÀ PHẠM VI NGHIÊN CỨU

Luận án tập trung nghiên cứu các phương pháp cho bài toán phân tích ý kiến theo các mức độ và quan điểm cũng như vấn đề hạn chế dữ liệu huấn luyện. Đối tượng nghiên cứu là bình luận dưới dạng văn bản không qua các bước tiền xử lý thủ công. Phạm vi nghiên cứu bao gồm dữ liệu tiếng Việt và một số ngôn ngữ ít tài nguyên khác.

1.3 ĐÓNG GÓP CỦA LUẬN ÁN

- **Đóng góp #1:** Nghiên cứu phương pháp nâng cao độ chính xác cho bài toán phân tích ý kiến theo mức độ. Kết quả công bố tại công trình [CT.01, CT.04, CT.06, CT.08].
- **Đóng góp #2:** Nghiên cứu mô hình nâng cao độ chính xác cho bài toán phân tích ý kiến theo hai quan điểm: thông thường và so sánh. Kết quả công bố tại công trình [CT.03, CT.07].
- **Đóng góp #3:** Nghiên cứu các phương pháp cho vấn đề hạn chế dữ liệu huấn luyện trong phân tích ý kiến. Kết quả công bố tại công trình [CT.05, CT.09].

1.4 Ý NGHĨA KHOA HỌC VÀ THỰC TIỄN

Ý nghĩa khoa học: Kết quả nghiên cứu góp phần phát triển chủ đề và phương pháp nghiên cứu. Các phương pháp, mô hình được đề xuất nâng cao hiệu suất dự đoán trên bài toán. Kết quả luận án là nền tảng cho sự phát triển nghiên cứu tiếp theo.

Ý nghĩa thực tiễn: Kết quả nghiên cứu này có thể ứng dụng vào các phần mềm phân tích bình luận tự động, góp phần nâng cao chất lượng sản phẩm.

1.5 BỐ CỤC CỦA LUẬN ÁN

Luận án được tổ chức thành 6 chương khác nhau, không bao gồm các công trình nghiên cứu khoa học được công bố liên quan đến đề tài và tài liệu tham khảo. Các đóng góp chính của nghiên cứu được trình bày ở Chương 3, Chương 4, và Chương 5.

2. TỔNG QUAN

2.1 PHÂN TÍCH Ý KIẾN

Phân tích ý kiến được nghiên cứu với mục đích tự động trích xuất và phân loại ý kiến, cảm xúc, quan điểm liên quan đến các thực thể như là sản phẩm, dịch vụ, v.v. trên các thuộc tính liên quan [1]. Theo Bing Liu [1], ý kiến bình luận thuộc một trong hai dạng ý kiến quan điểm:

- **Ý kiến thông thường**: Là ý kiến thể hiện cảm xúc, đánh giá về một thực thể, sự kiện, đối tượng cụ thể hoặc một khía cạnh liên quan.
- **Ý kiến so sánh**: Là ý kiến thể hiện sự so sánh giữa các thực thể, sự kiện, đối tượng dựa trên các khía cạnh nào đó.

Bài toán phân tích ý kiến được nghiên cứu theo mức độ dữ liệu, bao gồm:

- **Mức độ tài liệu**: Mục tiêu của bài toán trên mức độ dữ liệu này là phân loại toàn bộ bình luận theo các mức độ ý kiến khác nhau.
- **Mức độ câu**: Mục tiêu bài toán ở mức độ này là phân loại các mức độ ý kiến cho các câu hoặc cụm từ bình luận.
- **Mức độ khía cạnh**: Mục tiêu bài toán xác định trạng thái ý kiến cho từng khía cạnh đề cập trong bình luận.

Trong luận án này, NCS tập trung nghiên cứu các mô hình, phương pháp cho các bài toán khác nhau trong chủ đề phân tích ý kiến, bao gồm: (1) Phân loại ý kiến theo mức độ: tài liệu, câu bình luận và khí cạnh; (2) Phân tích ý kiến theo quan điểm: thông thường và so sánh.

2.2 KHẢO SÁT NGHIÊN CỨU LIÊN QUAN

Luận án áp dụng phương pháp khảo sát [2] để khảo sát các nghiên cứu liên quan với ba bước (1) Phát triển giao thức khảo sát; (2) Áp dụng chiến lược tìm kiếm trên các nguồn dữ liệu; (3) Trích xuất nội dung. Sau đây là một số điểm hạn chế sau khi khảo sát về chủ đề phân tích ý kiến trên tiếng Việt:

- **Vấn đề mất cân bằng hoặc hạn chế dữ liệu**: Hầu hết các bộ dữ liệu công bố đều mất cân bằng về khía cạnh và nhãn ý kiến.
- **Lỗi văn phong và ngữ pháp**: Bình luận trực tuyến thường chứa lỗi như viết tắt, lỗi chính tả, v.v. gây khó khăn cho các công cụ xử lý ngôn ngữ tự nhiên, đặc biệt khi chúng được huấn luyện trên dữ liệu chuẩn.
- **Hạn chế về phương pháp và bài toán nghiên cứu**: Các nghiên cứu trước đây tập trung vào học máy, học sâu và mô hình transformer. Các nghiên cứu dùng BERT thường chỉ sử dụng một phiên bản duy nhất và chưa đánh giá trên nhiều bộ dữ liệu để kiểm chứng khả năng tổng quát. Với bài toán nghiên cứu, các công trình trước đây chủ yếu tập trung vào phân tích cơ bản, bỏ qua các bài toán phức tạp yêu cầu trích xuất thông tin theo quan điểm
- **Không nhất quán trong độ đo đánh giá và bộ dữ liệu toàn diện**: Các nghiên cứu dùng nhiều độ đo khác nhau (micro F1, macro F1, weighted F1), gây khó khăn trong việc so sánh và đánh giá khách quan hiệu suất mô hình. Các bộ dữ liệu tiếng Việt còn hạn chế, chủ yếu tập trung vào một số bài toán đơn lẻ và lĩnh vực cụ thể, gây cản trở cho sự phát triển toàn diện của các phương pháp và mô hình.

2.3 CÁC BỘ DỮ LIỆU SỬ DỤNG TRONG LUẬN ÁN

Luận án thử nghiệm phương pháp trên các bộ dữ liệu đã công bố với các mức độ và miền dữ liệu khác nhau, bao gồm HSA [3], UIT-VSFC [4], VLSP [5] và AiviVN cho mức độ tài liệu và câu, còn mức độ khía cạnh thì bộ dữ liệu miền nhà hàng [6], khách sạn [6] và điện thoại [7].

Đối với phân tích ý kiến theo quan điểm, luận án đánh giá các mô hình trên bộ dữ liệu Vi-AbSQA với bốn thuộc tính ý kiến được gán nhãn cho miền nhà hàng phát triển thử bộ dữ liệu trước [6] và bộ dữ liệu VCOM [8] được gán nhãn cho miền điện thoại với bộ năm thuộc tính cho quan điểm so sánh.

2.4 CÁC MÔ HÌNH NGÔN NGỮ SỬ DỤNG TRONG LUẬN ÁN

Dưới đây là một số mô hình ngôn ngữ (MHNN) huấn luyện sẵn được sử dụng để trích xuất đặc trưng và biểu diễn ngôn ngữ sử dụng trong luận án:

- MHNN theo kiến trúc mã hóa (encoder-only): Bao gồm mô hình đơn ngôn ngữ như PhoBERT [9], viBERT_FPT [10], vELECTRA_FPT [10], ViDeBERTa [11] và mô hình đa ngôn ngữ như là mBERT [12], InfoXLM [13], XLM-Align [14].
- MHNN theo kiến trúc mã hóa - giải mã (encoder-decoder): luận án sử dụng trọng số các mô hình viT5 [15], mT5 [16] và BARTPho [17] với các phiên bản tham số khác nhau.

3. PHÂN TÍCH Ý KIẾN THEO MỨC ĐỘ

3.1 PHƯƠNG PHÁP HỌC CHUYỂN TIẾP TRÊN MHNN

3.1.1 *Động lực nghiên cứu*

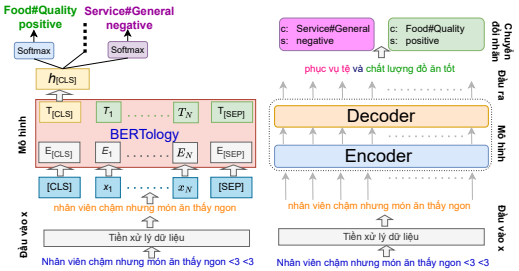
Qua khảo sát, luận án thấy rằng có nhiều mô hình ngôn ngữ khác nhau hỗ trợ tiếng Việt được huấn luyện trên các kiến trúc khác nhau và dữ liệu tiền huấn luyện và kỹ thuật tiền xử lý khác nhau. Điều này dẫn đến sự hiệu quả các mô hình cũng khác nhau với miền dữ liệu và bài toán thử nghiệm. Ngoài ra, NCS còn thấy một số hạn chế như sau:

- Nghiên cứu trước đây lựa chọn ngẫu nhiên một MHNN để huấn luyện và so sánh kết quả với các phương pháp truyền thống trên một số bộ dữ liệu.
- Thiếu thống nhất trong quá trình đánh giá so sánh kết quả giữa các phương pháp.
- Chưa có nghiên cứu nào khai thác tiềm năng của MHNN hướng tạo sinh văn bản cho bài toán trên tiếng Việt.

Vì vậy, NCS trình bày phương pháp học chuyển tiếp dựa trên các MHNN theo hai hướng tiếp cận: (1) phân loại và (2) tạo sinh văn bản trên các mức độ dữ liệu như sau:

3.1.2 *Mô hình theo hướng phân loại văn bản*

Hình 3.1 trình bày hai kiến trúc tổng quan cho phương pháp học chuyển tiếp dựa trên MHNN huấn luyện sẵn cho bài toán phân loại ý kiến theo các mức độ. Đầu tiên, các bình luận được chuẩn hóa thông qua các bước tiền xử lý dữ liệu. Bên cạnh đó, luận án cũng nhận thấy các MHNN đều áp dụng các bước tiền xử lý dữ liệu huấn luyện khác nhau. Ví dụ như PhoBERT [9] đã sử dụng kỹ thuật tách từ trong quá trình xây dựng dữ liệu tiền huấn luyện. Tuy nhiên, các mô hình khác như mBERT [12], XLM-R [18], viBERT_FPT [10] không áp dụng kỹ thuật này để chuẩn bị dữ liệu huấn luyện. Đặc điểm quan trọng này không được đề cập trong các nghiên



Hình 3.1. Hướng tiếp cận phân loại và tạo sinh văn bản dựa trên MHNN cho phân tích ý kiến theo mức độ khía cạnh.

cứu trước đây khi tinh chỉnh các MHNN [19, 20, 21, 22]. Do đó, điều quan trọng là phải thực hiện kỹ lưỡng các bước tiền xử lý tương ứng với các MHNN để tận dụng tối đa sức mạnh các mô hình. Các bước tiền xử lý được thiết kế dựa trên quan sát dữ liệu và mô hình tương ứng.

Sau khi trải qua các bước tiền xử lý, các bình luận đầu vào được tự động thêm hai ký tự đặc biệt là $[CLS]$ và $[SEP]$ trước khi được tách thành các tokens. Các tokens này sau đó được ánh xạ thành các chỉ mục tương ứng trong tập từ vựng của mô hình. Thông thường, mô hình sẽ chọn chuỗi bình luận có độ dài tokens lớn nhất để làm số chiều vectơ biểu diễn cho toàn bộ đầu vào, nếu độ dài bình luận nằm trong phạm vi mà mô hình xử lý. Sau khi các tokens được chuyển thành các vectơ dựa trên tập từ điển của mô hình, chúng được đưa vào các MHNN đã được huấn luyện sẵn với L lớp transformer để lấy các vectơ biểu diễn theo ngữ cảnh $H^L = \{h^L_{cls}, h^L_1, ..., h^L_T, h^L_{sep}\} \in \mathbb{R}^{T \times dim_h}$, với dim_h là số chiều của các vectơ biểu diễn và T là tổng số token trong chuỗi, bao gồm cả các token đặc biệt như $[CLS]$ và $[SEP]$. Mỗi lớp transformer cung cấp các biểu diễn ngữ cảnh, nghĩa là các vectơ biểu diễn cho từng token chứa thông tin ngữ cảnh xung quanh. Dựa trên nghiên cứu trước đây [12, 23, 24] ở các ngôn ngữ khác, luận án trích xuất h^L_{cls} của $[CLS]$ token tại lớp cuối cùng để làm đặc trưng biểu diễn ngữ cảnh tổng quát cho toàn bộ bình luận. Đối với mức độ dữ liệu tài liệu hoặc câu bình luận (Hình 3.1a), vectơ đặc trưng biểu diễn này được đưa trực tiếp vào lớp kết nối đầy đủ (fully connected layer)

với hàm kích hoạt Softmax hoặc Sigmoid để dự đoán xác suất nhãn ý kiến tương ứng.

Đối với bài toán phân loại ý kiến theo mức độ khía cạnh, NCS thiết kế mô hình theo hướng đa tác vụ (multi-task learning) để dự đoán tất cả các loại khía cạnh và trạng thái ý kiến tương ứng như mô tả ở Hình 3.1b. Tiếp cận theo hướng đa tác vụ cho phép mô hình học được mối quan hệ giữa các nhãn loại khía cạnh và nhãn ý kiến với nhau. Để rút trích tất cả các loại khía cạnh và nhãn ý kiến tương ứng, đầu ra được biểu diễn dưới dạng một danh sách các one-hot vectơ tương ứng với số lượng khía cạnh. Các one-hot vectơ có định dạng là một vectơ có bốn giá trị, trong đó chỉ có một phần tử được gán giá trị 1, các phần tử còn lại là 0. Bốn giá trị tương ứng của one-hot vectơ được giải mã với giá trị [“None”, “Positive”, “Neutral”, “Negative”] cho một loại khía cạnh cụ thể với “None” nghĩa là loại khía cạnh này không được đề cập trong bình luận. Danh sách các MHNN huấn luyện sẵn được sử dụng để tinh chỉnh trong luận án bao gồm mô hình đa ngôn ngữ như mBERT [12], XLM-R [18]; các mô hình đơn ngôn ngữ như viBERT4News, viBERT_FPT [10], viELECTRA_FPT [10], PhoBERT [9].

3.1.3 *Mô hình theo hướng tạo sinh văn bản*

Hình 3.1 trình bày phương pháp học chuyển tiếp dựa trên MHNN theo hướng tiếp cận tạo sinh văn bản theo các mức độ dữ liệu. Cho một bình luận đầu vào X gồm n từ vựng được đại diện bởi $X_{1:n} = \{x_1, x_2, ..., x_n\}$, nghiên cứu sinh mong muốn tạo ra một chuỗi đầu ra gồm $Y_{1:m} = \{y_1, y_2, ..., y_m\}$ với m nằm trong khoảng được định nghĩa trước. Đối với mức độ tài liệu hoặc câu bình luận thì đầu ra $Y_{1:m}$ chỉ chứa duy nhất thông tin liên quan đến nhãn ý kiến (như mô tả ở Hình 3.1a). Trong khi đó, với bài toán ở mức độ khía cạnh, đầu ra $Y_{1:m}$ phải chứa tất cả các thông tin về loại khía cạnh và nhãn ý kiến tương ứng được đề cập trong đầu vào. Sau đó, văn bản được tạo ra bởi mô hình được biến đổi thành các nhãn tương ứng để đánh giá và so sánh hiệu quả với các mô hình khác.

Một trong những đóng góp của luận án ở phương pháp này là kỹ thuật tuyến tính hóa các nhãn loại khía cạnh và nhãn ý kiến thành chuỗi dưới dạng ngôn ngữ tự nhiên $Y_{1:m}$ để tận dụng tối đa khả năng của MHNN tạo sinh. Nguyên nhân là do các nhãn được gán nhãn ở định dạng tiếng Anh (ví dụ như Food#Quality, positive). Do đó, nghiên cứu sinh áp dụng một tập từ điển để ánh xạ hai giá trị sang dạng ngôn ngữ tự nhiên. Để xây dựng tập từ điển ánh xạ, luận án áp dụng phương pháp thống kê dựa trên kỹ thuật TF-IDF để trích xuất các mẫu từ vựng cho các loại khía cạnh trên tập huấn luyện và tập phát triển. Hàm chuyển đổi khía cạnh thành từ vựng biểu diễn dưới dạng ngôn ngữ tự nhiên tiếng Việt được biểu diễn như sau:

$$\text{Cụm từ khía cạnh}(A) = \begin{cases} \text{“chất lượng đồ ăn”}, & \text{nếu } A = \text{Food\#Quality} \\ \vdots & \vdots \\ \text{“phục vụ”}, & \text{nếu } A = \text{Service\#General} \end{cases} \tag{1}$$

với $A \in \{\text{Food\#Quality}, \text{Drinks\#Quality}, \ldots \text{Service\#General}\}$ là các loại khía cạnh thuộc miền lĩnh vực nhà hàng.

Đối với các nhãn ý kiến, luận án cũng biến đổi thành các từ vựng mang tính tự nhiên như công thức sau:

$$\text{Nhãn ý kiến}(S) = \begin{cases} \text{“tốt”}, & \text{nếu } S = \text{positive} \\ \text{“tệ”}, & \text{nếu } S = \text{negative} \\ \text{“tạm”}, & \text{nếu } S = \text{neutral} \end{cases} \tag{2}$$

Trong quá trình phân tích dữ liệu, nghiên cứu sinh nhận thấy rằng người dùng có xu hướng bày tỏ ý kiến của họ một cách ngắn gọn về các loại khía cạnh và ý kiến của họ. Dựa trên quan sát này, câu ngôn ngữ tự nhiên cho một thể loại khía cạnh và nhãn ý kiến tương ứng (A, S) được biểu diễn dưới dạng định dạng đầu ra. Điều này có thể được biểu diễn dưới dạng hàm:

$$\text{Chuỗi đầu ra}(A, S) = \text{Cụm từ khía cạnh}(A) + \text{Nhãn ý kiến}(S) \tag{3}$$

Trường trường hợp nếu bình luận đầu vào chứa nhiều loại khía cạnh và trạng thái ý kiến khác nhau như $\{(A_1, S_1), (A_2, S_2), \ldots, (A_m, S_m)\}$ thì chuỗi ngôn ngữ tự nhiên đầu ra Y là kết quả của việc nối các biểu diễn cho từng loại khía cạnh (A_i, S_i) với liên từ "và". Chuỗi đầu ra Y được biểu diễn như sau:

$$Y = \text{Chuỗi đầu ra}(A_1, S_1) + \cdots + \text{ "và" } + \text{Chuỗi đầu ra}(A_m, S_m) \quad (4)$$

Để suy luận kết quả dự đoán của mô hình dưới dạng chuỗi ngôn ngữ tự nhiên Y thành các nhãn tương ứng thì nghiên cứu sinh thực hiện quá trình ánh xạ dựa trên tập từ điển khía cạnh và nhãn ý kiến như sau. Đầu tiên tách chuỗi đầu ra Y thành các diễn diễn cho từng loại cặp nhãn (A_i, S_i) dựa trên liên từ "và". Sau đó, từng chuỗi đầu ra Chuỗi đầu ra(A_1, S_1) sẽ xác định được kiểm tra sự trùng khớp với từ khóa của tập từ điển loại khía cạnh và từ điển nhãn ý kiến. Còn để huấn luyện mô hình, nghiên cứu sinh cài đặt các MHNN theo kiến trúc mã hóa - giải mã (encoder-decoder) với các trọng số được khởi tạo từ mô hình huấn luyện sẵn, bao gồm viT5 [15], mT5 [16] và BARTPho [17].

3.1.4 *Kết quả thử nghiệm*

Luận án đánh giá các phương pháp trên nhiều bộ dữ liệu ở các mức độ tài liệu và câu bình luận khác nhau, bao gồm HSA [3], VLSP [5], và UIT-VSFC [4]. Kết quả thử nghiệm cho thấy PhoBERT đạt hiệu suất cao nhất khi áp dụng học chuyển tiếp cho bài toán phân loại ý kiến theo hướng tiếp cận phân loại. Trong khi đó, viT5 hoạt động tốt trên các bộ dữ liệu mất cân bằng, đặc biệt khi đánh giá bằng các độ đo Balance Accuracy và Macro F_1. Những kết quả này chỉ ra rằng phương pháp tạo sinh dựa trên viT5 có thể là giải pháp hiệu quả cho các bộ dữ liệu mất cân bằng, nơi mà tầm quan trọng của các nhãn được coi là ngang nhau. Các kết quả thử nghiệm trong luận án phù hợp với những quan sát từ các nghiên cứu trước đó [9, 15] trong các bài toán XLNN-TN khác nhau.

Bảng 3.1. Kết quả thử nghiệm trên bộ dữ liệu UIT-VSFC [4].

Phương pháp	Mô hình	Accuracy	Balance Acc	Weighted F_1	Macro F_1	Micro F_1
Nghiên cứu trước đây	MaxEnt [4]	-	-	87.94	-	-
	LD-SVM [26]	90.74	-	90.20	-	-
	BiLSTM-CNN [27]	-	-	93.51	-	-
	CNN [28]	-	-	-	75.57	89.82
	CNN + Augmentation [28]	-	-	-	77.16	89.38
	Two-channel LSTM [29]	89.90	-	89.30	-	-
	Two-channel CNN [29]	89.60	-	88.90	-	-
	Ensemble [22]	-	-	-	-	92.79
	PhoBERT [19]	**94.28**	-	**93.92**	-	-
Hướng tiếp cận tạo sinh văn bản	mT5_{base}	85.75	60.46	83.47	59.07	85.75
	mT5_{large}	90.68	68.28	89.27	70.12	90.68
	viT5_{base}	92.40	**81.35**	92.19	82.07	92.40
	viT5_{large}	92.89	79.84	92.54	**82.98**	92.89
Hướng tiếp cận phân loại	viBERT4News	86.67	61.01	84.32	59.34	86.67
	viBERT_FPT	91.25	74.18	90.64	76.74	91.25
	viELECTRA_FPT	90.52	72.96	89.87	75.37	90.52
	mBERT	92.04	75.10	91.41	78.07	92.04
	XLM-R	93.08	77.25	92.55	80.35	93.08
	PhoBERT	93.87	79.57	93.45	82.83	**93.87**

Còn với mức độ khía cạnh, miền dữ liệu nhà hàng [6] và miền điện thoại [7] được sử dụng để đánh giá các phương pháp. Qua các bảng kết quả, luận án nhận thấy rằng phương pháp tiếp cận tạo sinh văn bản cải thiện đáng kể hiệu suất trên cả hai bộ dữ liệu so với cách tiếp cận dựa trên phân loại. Như được minh họa trong Bảng 3.2, mô hình viT5 cho kết quả vượt trội so với tất cả các phương pháp trước đó, bao gồm cả các MHNN khác. Cụ thể, viT5 đạt hiệu suất 75.53% về độ đo F_1, cao hơn so với kết quả tốt nhất từ sự kết hợp các mô hình BERT đã tinh chỉnh, dù mức chênh lệch không lớn. Tuy nhiên, phương pháp tạo sinh văn bản có ưu thế về độ phức tạp tính toán thấp hơn và tốc độ dự đoán nhanh hơn so với phương pháp kết hợp. So với hướng tiếp cận phân loại dựa trên MHNN BERT, viT5 cũng cho kết quả tốt hơn PhoBERT khoảng +0.65% trên độ đo F_1. Trong lĩnh vực điện thoại, mô hình viT5 tiếp tục thể hiện vượt trội khi đạt độ đo F_1 là 81.10%, cao hơn 2.34% so với kết quả nghiên cứu trước đây [25].

3.2 PHƯƠNG PHÁP TỔ HỢP DỰA TRÊN CÁC MÔ HÌNH NGÔN NGỮ

3.2.1 *Động lực nghiên cứu*

Nghiên cứu trước (Mục 3.1) của luận án đã chứng minh sự hiệu quả của phương pháp học chuyển tiếp dựa trên MHNN. Tuy nhiên, các MHNN lại có kiến trúc, áp dụng các bước tiền xử, và được huấn luyện trên các nguồn

Bảng 3.2. Kết quả ở mức độ khía cạnh trên miền nhà hàng [6] và điện thoại [7].

Phương pháp	Mô hình	Miền nhà hàng			Miền điện thoại		
		Độ Chính Xác	Độ phủ	Độ đo F_1	Độ Chính Xác	Độ phủ	Độ đo F_1
Nghiên cứu trước đây	SVM [6, 7]	59.35	60.37	59.85	16.09	23.35	19.69
	CNN [6, 7]	68.61	64.91	66.71	33.34	22.92	27.16
	BiLSTM-CNN [6, 7]	71.47	67.67	69.52	65.82	60.53	63.06
	PhoBERT + Processing [25]	-	-	-	80.21	79.46	78.76
Hướng tiếp cận phân loại	XLM-R	70.98	72.19	71.58	80.58	68.71	74.18
	XML-Align	72.08	72.08	72.08	79.84	69.24	74.16
	InfoXLM	72.41	72.77	72.59	80.88	69.12	74.02
	viBERT_FPT	64.94	60.94	62.87	77.18	65.80	71.04
	PhoBERT	73.86	73.18	73.52	**82.72**	67.66	74.44
Hướng tiếp cận tạo sinh văn bản	$\text{BARTpho}_{syllable+base}$	68.34	67.48	67.90	79.41	80.36	79.88
	$\text{BARTpho}_{word+base}$	69.01	66.57	67.76	78.86	77.40	78.12
	$\text{BARTpho}_{syllable+large}$	68.10	69.19	68.64	80.49	81.25	80.86
	$\text{BARTpho}_{word+large}$	71.25	69.57	70.40	79.26	80.62	79.93
	mT5_{base}	68.01	68.66	68.33	78.91	78.89	78.90
	viT5_{base}	73.60	73.79	73.69	78.90	**81.82**	80.33
	mT5_{large}	70.52	71.45	70.98	79.46	79.62	79.54
	viT5_{large}	**75.73**	**75.35**	**75.53**	81.06	81.20	**81.10**

dữ liệu khác nhau. Do đó, sử dụng các MHNN có khả năng tạo ra đặc trưng biểu diễn và dự đoán khác nhau cho từng bài toán và từng miền lĩnh vực. Do đó, việc kết hợp các đặc trưng hoặc khả năng dự đoán các mô hình sẽ giúp tăng cường sự khái quát hóa mô hình, từ đó mang lại những cải thiện đáng kể về hiệu suất so với mô hình đơn lẻ. Các nghiên cứu trước đây [30, 31, 32, 27, 22, 33] khi nghiên cứu về phương pháp học tổ hợp là dựa trên các luật định nghĩa, các mô hình máy học hoặc học sâu khác nhau để trích xuất đặc trưng trên dữ liệu đầu vào. Khác với các nghiên cứu trước đây, luận án nghiên cứu hai phương pháp kết hợp các đặc trưng biểu diễn và kết quả dự đoán các MHNN đã được huấn luyện riêng lẻ, bao gồm phương pháp (1) hợp nhất đặc trưng (Feature Fusion) và (2) Bình chọn mềm (Soft Voting). Để huấn luyện các mô hình cơ sở được sử dụng để tạo thành các đặc trưng hoặc tổng hợp kết quả, NCS cài đặt phương pháp học chuyển tiếp trên MHNN hướng phân loại để trích xuất các đặc trưng hoặc xác suất dự đoán. Chi tiết phương pháp huấn luyện mô hình cơ sở được trình bày ở Mục 3.1.

3.2.2 *Kỹ thuật hợp nhất đặc trưng*

Mục tiêu của kỹ thuật hợp nhất đặc trưng là khai thác sức mạnh của nhiều mô hình khác nhau nhằm biểu diễn dữ liệu đầu vào một cách toàn diện hơn so với việc chỉ sử dụng từng mô hình đơn lẻ. NCS áp dụng phép nối (concatenation) để hợp nhất đặc trưng sau khi đã thử nghiệm các phương

Bảng 3.3. Kết quả phương pháp tổ hợp tốt nhất so với các hướng tiếp cận trước đây trên hai bộ dữ liệu cho bài toán Phân loại ý kiến.

Mô hình	**Bộ dữ liệu UIT-VSFC**		
	Weighted F_1	**Macro F_1**	**Micro F_1**
MaxEnt [4]	87.94	77.00	87.72
LD-SVM [26]	90.20	-	-
BiLSTM-CNN [27]	93.51	-	-
CNN [28]	-	75.57	89.82
CNN+Augmentation [28]	-	77.16	89.38
Two-channel LSTM [29]	89.90	-	-
Two-channel CNN [29]	88.90	-	-
Ensemble Deep Models [22]	-	-	92.79
PhoBERT [19]	93.92	-	-
Kết quả luận án	**94.03**	**84.51**	**94.32**

Mô hình	**Bộ dữ liệu AiviVn**
	Weighted F_1
Hạng 1 AiviVn 2019	90.01
RCNN [34]	92.98
PhoBERT + Sentivec [35]	92.00
Gating Ensemble [36]	93.82
LIFA-SIGMOID [34]	95.20
Kết quả luận án	**95.65**

pháp thay thế khác như phép cộng, phép nhân và cơ chế chú ý (attention mechanism). Kết quả thử nghiệm cho thấy rằng phương pháp phép nối vẫn luôn cho kết quả tốt nhất. Hiệu suất vượt trội của chiến lược ghép nối có thể được giải thích bởi khả năng tăng cường sự đa dạng của đặc trưng.

3.2.3 *Kỹ thuật bình chọn mềm*

Luận án nghiên cứu áp dụng kỹ thuật bình chọn mềm để quyết định các dự đoán bằng các tính trung bình xác xuất cho mỗi lớp dự đoán và kết quả cuối cùng là nhãn dự đoán có xác xuất cao nhất. Mô hình bình chọn thường là sự kết hợp của hai hay nhiều mô hình cơ sở được tối ưu hóa riêng lẻ trên dữ liệu huấn luyện. Giả sử chúng ta có một tập hợp các đầu ra $\{y_1, y_2, ...y_T\}$ của T bộ phân loại cơ sở với $y_i = [y_i^1, y_i^2, ...y_i^l]$ với l là tập hợp các nhãn tiềm năng $\{c_1, c_2, ...c_l\}$, với $y_i^j \in [0, 1]$, có thể được coi là xác suất cho thực thể x. Luận án thiết lập trọng số của các mô hình cơ sở như nhau, vì thế phương pháp bỏ phiếu mềm được tính bằng trung bình cộng trên tất cả giá trị đầu ra của các mô hình cơ sở, và xác suất của lớp c_j được tính bằng công thức sau:

$$y^j(x) = \frac{1}{T}\sum_{i=1}^{T} y_i^j(x) \tag{5}$$

Bảng 3.4. Kết quả độ đo F_1 trên các bộ dữ liệu cho bài toán Phân tích ý kiến theo mức độ khía cạnh.

Loại	Mô hình	Miền nhà hàng	Miền khách sạn	Miền điện thoại
Mô hình trước đây	CNN	66.71	69.49	-
	LSTM [7]	-	-	52.13
	BiLSTM + Attention [6]	-	69.86	-
	BiLSTM-CNN [6]	69.52	70.72	63.06
	PhoBERT [6]	74.88	73.69	-
	HSUM-HC_4 [37]	-	74.88	-
	HSUM-HC_8 [37]	-	**75.25**	-
Mô hình đa ngôn ngữ	XLM-R	71.58	71.23	74.18
	InfoXLM	72.59	71.92	74.02
	XLM-Align	72.08	73.20	74.16
Mô hình đơn ngôn ngữ	viBERT_FPT	62.87	70.18	71.04
	PhoBERT	73.52	73.73	74.44
Hợp nhất đặc trưng (Feature Fusion)	Multi-models	75.12	73.21	75.94
	Mono-models	72.57	72.91	74.92
	All models	75.24	74.22	76.14
	Two-best models	75.32	74.54	75.89
Bình chọn mềm (Voting Ensemble)	Multi-models	74.55	73.81	75.87
	Mono-models	71.74	72.91	74.80
	Two-best models	75.09	74.79	75.67
	All models	**75.36**	74.15	**76.23**

3.2.4 *Kết quả thử nghiệm*

Bảng 3.3 trình bày kết quả thử nghiệm tốt nhất so với các phương pháp trước đó trên hai bộ dữ liệu UIT-VSFC và AiviVn. Có thể thấy rằng các mô hình tốt nhất của nghiên cứu hoạt động tốt hơn các cách tiếp cận trước đó. Trong đó, phương pháp bỏ phiếu mềm dựa trên dựa trên hai mô hình cơ sở mạnh cho hiệu suất tốt nhất là là 94,03%, 94,32% và 84,51% về chỉ số Weighted F_1, Micro F_1 và Macro F_1 cho bộ dữ liệu UIT-VSFC. Tương tự, trong bộ dữ liệu Aivivn, điểm số tốt nhất đạt được bởi phương pháp tổng hợp bỏ phiếu mềm dựa trên cả năm mô hình với 95,65%, 95,65% và 95,64% về Weighted F_1, Macro F_1 và Micro F_1. Lưu ý rằng Aivivn là một bộ dữ liệu gần như cân bằng, do đó điểm số của cả ba số liệu đều giống nhau. Lý do chính cho thành công của phương pháp bỏ phiếu là mỗi bộ phân loại cơ sở khám phá biểu diễn theo ngữ cảnh theo những cách khác nhau, góp phần vào hệ số phân loại trong bỏ phiếu mềm.

Với dữ liệu ở mức độ khía cạnh, Bảng 3.4 trình bày hiệu suất của các mô hình đối với miền nhà hàng, khách sạn và điện thoại. Nhất quán với các thử nghiệm trước đó trên các bộ dữ liệu SA, PhoBERT vượt trội hơn các mô hình cơ sở cá nhân khác vì độ dài tối đa của bộ dữ liệu phù hợp với trình

tự đầu vào của mô hình. Nói chung, các chiến lược kết hợp của luấn dựa trên hai mô hình tốt nhất đạt kết quả vượt trội so với các mô hình cá nhân khác và các phương pháp hiện có về điểm số F_1, cho thấy hiệu quả của các mô hình trong việc phân loại ý kiến theo khía cạnh. Cụ thể, điểm F_1 cao nhất được đạt được bởi chiến lược kết hợp bầu cử dựa trên tất cả các mô hình cá nhân cho lĩnh vực nhà hàng và điện thoại thông minh. Đối với lĩnh vực khách sạn, kết hợp bầu cử dựa trên hai bộ phân loại tốt nhất đạt hiệu suất về điểm số F_1 là 74,79%. Mặc dù kết quả thử nghiệm không thể vượt trội hơn HSUM-HCM-8 [37] trong thử nghiệm này, nhưng nó vẫn đạt được hiệu suất khá tốt. So sánh kết quả của việc hợp nhất tính năng so với kết hợp bầu cử mềm, kỹ thuật bình chọn mềm có thể đạt kết quả tốt hơn về điểm số F_1 trên hầu hết các chuẩn, nhưng sự khác biệt giữa hai chiến lược không đáng kể. Ngoài ra, mô hình đơn ngữ viBERT_FPT đạt hiệu suất kém nhất trong số các mô hình cá nhân ở ba lĩnh vực. Một trong những lý do cho hiệu suất kém của các mô hình cơ sở viBERT_FPT là từ điển từ vựng, được giữ nguyên từ mô hình mBERT, dẫn đến không đại diện đầy đủ cho các từ vựng khác nhau trong một lĩnh vực cụ thể [38].

3.3 KẾT CHƯƠNG

Một số nội dung và kết quả trong chương 3 đã được công bố tại các công trình nghiên cứu số bao gồm các tạp chí chuyên ngành quốc tế (CT.01, CT.04, CT.05), trong nước (CT.06), và hội nghị quốc tế (CT.08).

4. PHÂN TÍCH Ý KIẾN THEO QUAN ĐIỂM

4.1 ĐỘNG LỰC NGHIÊN CỨU

Phân tích ý kiến theo quan điểm là bài toán phức tạp nhất với mục tiêu là rút trích các bộ thuộc tính có mối quan hệ với nhau. Để giải quyết bài toán này, luận án đã xây dựng mô hình theo hướng tiếp cận tạo sinh văn bản dựa trên MHNN với các đóng góp và cải tiến như sau:

- (1) NCS thiết kế các khuôn mẫu để biểu diễn các thuộc tính của ý kiến thành văn bản dưới dạng ngôn ngữ tự nhiên có cấu trúc tiếng Việt.
- (2) NCS kết hợp các lời nhắc hướng dẫn (prompt instruction) với đầu vào để mô tả thông tin cho mô hình.
- (3) Để tận dụng mối liên hệ giữa các thuộc tính, luận án huấn luyện mô hình theo hướng đa tác vụ để nâng cao hiệu suất dự đoán.

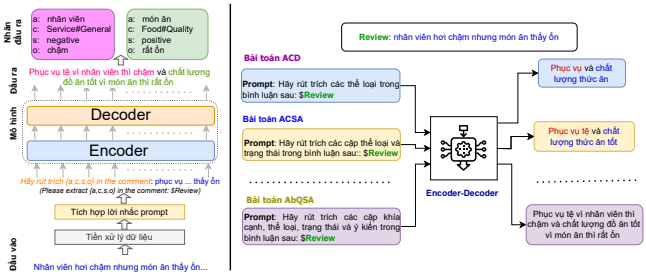
4.2 PHÂN TÍCH Ý KIẾN THEO QUAN ĐIỂM THÔNG THƯỜNG

4.2.1 *Mô hình tạo sinh văn bản kết hợp lời nhắc theo hướng đa tác vụ*

Hình 4.1 minh họa mô hình tạo sinh văn bản dựa trên kiến trúc mã hóa-giải mã với lời nhắc, và bên phải là cách huấn luyện đa tác vụ trên tổ hợp các thuộc tính. Các bình luận được tiền xử lý nhằm giảm thiểu nhiễu và tích hợp một lời nhắc hướng dẫn (prompt) để cung cấp thêm thông tin. Đầu vào này sau đó được đưa vào kiến trúc mã hóa-giải mã để sinh ra văn bản đầu ra. Văn bản đầu ra không chỉ chứa thông tin mà còn thể hiện mối quan hệ giữa các thuộc tính trong bộ bốn thuộc tính.

Sau khi tiền xử lý, bình luận được tự động thêm một lời nhắc prompt để giúp mô hình nâng cao khả năng phân tích thuộc tính cần rút trích. Ý tưởng của kỹ thuật này là dạy cho mô hình thực hiện các bài toán được mô tả thông qua lời nhắc hướng dẫn [39], điều này giúp mô hình có khả năng huấn luyện theo phương thức đa tác vụ như sau:

> Hãy rút trích các cặp khía cạnh, thể loại, trạng thái và ý kiến trong bình luận sau: {$ Review} (1)



Hình 4.1. Mô hình tạo sinh văn bản kết hợp lời nhắc hướng dẫn theo hướng đa tác vụ cho phân tích quan điểm thông thường.

Thiết kế lời nhắc như trên được sử dụng để hỗ trợ mô hình xác định và trích xuất bộ bốn thuộc tính từ đầu vào Review. Luận án thử nghiệm các thiết kế lời nhắc phức tạp [40] nhưng hiệu quả với các MHNN tiếng Việt. Nguyên nhân là lời nhắc phức tạp gây nhầm lẫn và giảm độ chính xác hoặc đầy đủ của đầu ra khiến hiệu suất tổng quan giảm.

Luận án giải quyết bài toán phân tích ý kiến theo quan điểm thông thường bằng cách tiếp cận tạo sinh văn bản. Vì vậy, cần chuyển đổi bộ thuộc tính thành một định dạng mà mô hình có thể dự đoán hiệu quả. Vì vậy, luận án đề xuất một kỹ thuật biểu diễn bộ thuộc tính thành câu ngôn ngữ tự nhiên tiếng Việt. Bằng cách chuyển đổi các thuộc tính thành đầu ra dưới dạng ngôn ngữ tự nhiên, nghiên cứu sinh không chỉ tận dụng được tri thức có sẵn của các MHNN mà còn dễ dàng tích hợp ngữ nghĩa của bộ thuộc tính. Chi tiết định dạng biểu diễn cho một bộ thuộc tính như sau:

$$\text{Chuỗi đầu ra}_{(a,o,c,s)} = P_c(c)\ P_s(s) \text{ vì } P_a(a) \text{ thì } P_o(o) \tag{2}$$

Trong đó $P_c(c)$ và $P_s(s)$ là các hàm chuyển đổi nhãn loại khía cạnh và nhãn ý kiến thành các cụm từ tiếng Việt tự nhiên và được biểu diễn như sau:

$$P_c(c) = \begin{cases} \text{“chất lượng đồ ăn”}, & \text{nếu } c = \text{Food\#Quality} \\ \vdots & \vdots \\ \text{“phục vụ”}, & \text{nếu } c = \text{Service\#General} \end{cases} \tag{3}$$

với $c \in \{\text{Food\#Quality}, \text{Drinks\#Quality}, \ldots \text{Service\#General}\}$ là các loại khía cạnh thuộc miền lĩnh vực nhà hàng.

Đối với các nhãn ý kiến, luận án cũng biểu diễn thành các từ vựng tự nhiên phù hợp với ngữ cảnh của miền dữ liệu thực nghiệm thông qua quá trình phân tích dữ liệu thủ công như trong công thức sau:

$$P_s(s) = \begin{cases} \text{“tốt”}, & \text{nếu } s = \text{positive} \\ \text{“tạm”}, & \text{nếu } s = \text{neutral} \\ \text{“tệ”}, & \text{nếu } s = \text{negative} \end{cases} \tag{4}$$

Trong dữ liệu thực nghiệm, mỗi bình luận được gán nhãn với ít nhất hai loại khía cạnh khác nhau với hai trạng thái ý kiến khác nhau. Do đó, luận án sử dụng liên từ “và” nối chuỗi biểu diễn của cặp bộ thuộc tính khác nhau thành chuỗi đầu ra của mô hình, tuân theo công thức sau:

$$Y = \text{Chuỗi đầu ra}_{(a_1,o_1,c_1,s_1)} + \cdots + \text{ “và” } + \text{Chuỗi đầu ra}_{(a_m,o_m,c_m,s_m)} \tag{5}$$

Trong đó, thứ tự của các chuỗi đầu ra cho từng bộ thuộc tính được sắp xếp theo thứ tự xuất hiện của chúng trong dữ liệu gốc. Điều này dựa trên quan sát cho thấy rằng các bộ thuộc tính đã được xác định trước thường xuất hiện ở đầu dữ liệu như cách con người đọc từ trái sang phải. Việc giữ nguyên thứ tự này không chỉ giúp mô hình học cách phân tích kết quả dự đoán theo đúng trật tự của các bộ thuộc tính trong dữ liệu đầu vào mà còn phù hợp với cách huấn luyện của các mô hình ngôn ngữ.

Đối với các cụm từ khía cạnh $P_a(a)$ và cụm từ ý kiến $P_o(o)$ thì được áp dụng các bước tiền xử lý tương tự tương tự như đối với bình luận đó để đảm bảo mô hình hiểu rằng thông tin thuộc tính này được trích xuất trong bình luận đầu vào. Điều này giúp mô hình liên kết chặt chẽ giữa các cụm từ này với ngữ cảnh của bình luận. Đối với những khía cạnh hoặc ý kiến không có cụm từ tương ứng rõ ràng hoặc tiềm ẩn trong bình luận, luận án sử dụng các cụm “khía cạnh tiềm ẩn” và “ý kiến tiềm ẩn” để biểu diễn trong chuỗi đầu ra của bộ thuộc tính tương ứng.

Bảng 4.1. Danh sách lời nhắc hướng dẫn tiếng Việt cho các bài toán khác nhau.

Bài toán	Lời nhắc hướng dẫn
ACD	Hãy rút trích các thể loại trong bình luận sau: {$Review}
ATE	Hãy rút trích các khía cạnh trong bình luận sau: {$Review}
OTE	Hãy rút trích các ý kiến trong bình luận sau: {$Review}
ATSA	Hãy rút trích các cặp khía cạnh và trạng thái trong bình luận sau: {$Review}
AOPE	Hãy rút trích các cặp khía cạnh và ý kiến trong bình luận sau: {$Review}
ACSA	Hãy rút trích các cặp thể loại và trạng thái trong bình luận sau: {$Review}
ASTE	Hãy rút trích các cặp khía cạnh, trạng thái và ý kiến trong bình luận sau: {$Review}
ACSD	Hãy rút trích các cặp khía cạnh, thể loại và trạng thái trong bình luận sau: {$Review}
AbSQA	Hãy rút trích các cặp khía cạnh, thể loại, trạng thái và ý kiến trong bình luận sau: {$Review}

Bảng 4.2. Mô hình hóa đầu ra của các bài toán thuộc bộ bốn thuộc tính.

Tiền tố	Tên bài toán	Đầu ra
ACD	Aspect Category Detection	$P_c(c)$
ATE	Aspect Term Extraction	$P_a(a)$
OTE	Opinion Term Extraction	$P_o(o)$
ATSA	Aspect-Term Sentiment Analysis	$P_a(a)\ P_s(s)$
AOPE	Aspect-Opinion Pair Extraction	$P_a(a)$ thì $P_o(o)$
ACSA	Aspect-Category Sentiment Analysis	$P_c(c)\ P_s(s)$
ASTE	Aspect Sentiment Triplet Extraction	$P_a(a)\ P_s(s)$ vì $P_o(o)$
ACSD	Aspect Category Sentiment Detection	$P_c(c)\ P_s(s)$ vì $P_a(a)$
AbSQA	Aspect-based Sentiment Quadruple Analysis	$P_c(c)\ P_s(s)$ vì $P_a(a)$ thì $P_o(o)$

Luận án huấn luyện mô hình theo hướng đa tác vụ để giúp mô hình nâng cao hiệu suất tổng thể khi trích xuất bộ thuộc tính. Điều này cho phép mô hình học các biểu diễn cho nhiều tác vụ của từng thuộc tính và mối liên hệ giữa các thuộc tính trên cùng một không gian biểu diễn [41]. Từ đó mô hình khai thác tối đa thông tin dữ liệu, đặc biệt là đối với các bộ dữ liệu bị hạn chế mẫu. Theo Zhang và cộng sự [42], từ bộ bốn thuộc tính được phát triển thành các bài toán riêng lẻ và phức tạp dựa trên sự tương tác giữa các thuộc tính. Vì vậy, luận án tạo thành các bài toán khác nhau từ bộ bốn thuộc tính với các thiết kế lời nhắc tương ứng ở Bảng 4.1 và định dạng biểu diễn đầu ra ở Bảng 4.2.

Luận án sử dụng phương pháp huấn luyện mô hình theo hướng tiếp cận tạo sinh văn bản dựa trên kiến trúc mã hóa-giải mã (encoder-decoder) của viT5 [15] và mT5 [16] với trọng số được khởi tạo từ các mô hình huấn luyện sẵn.

4.2.2 *Kết quả thử nghiệm*

Dựa vào Bảng 4.3, kết quả chỉ ra đề xuất tích hợp thêm lời nhắc hướng dẫn prompt giúp nâng cao hiệu suất. Kết quả này chứng minh rằng việc huấn

Bảng 4.3. Kết quả khi có và không tích hợp lời nhắc prompt theo hướng đơn tác vụ.

Bài toán	Single No-Prompt Tuning			Single Prompt Tuning		
	Độ chính xác	Độ phủ	Chỉ số F_1	Độ chính xác	Độ phủ	Chỉ số F_1
ACD	89.27	86.80	88.02	89.82	88.42	**89.12**
ATE	73.74	73.10	73.42	74.39	74.07	**74.23**
OTE	71.06	70.69	70.87	73.26	72.76	**73.01**
ACSA	74.91	72.53	73.70	77.86	75.94	**76.89**
ATSA	78.03	76.23	77.12	78.71	75.87	**77.26**
AOPE	70.72	66.72	68.66	71.60	67.93	**69.72**
ASTE	66.83	62.53	64.61	68.10	65.02	**66.53**
ACSD	69.85	64.72	67.19	71.77	69.92	**70.83**
AbSQA	63.19	60.37	61.75	63.54	61.09	**62.29**

Bảng 4.4. Kết quả khi có và không tích hợp lời nhắc prompt theo hướng đa tác vụ.

Bài toán	Multi-task No-Prompt Tuning			Multi-task Prompt Tuning		
	Độ chính xác	Độ phủ	Chỉ số F_1	Độ chính xác	Độ phủ	Chỉ số F_1
ACD	90.86	87.04	88.91	91.42	88.18	**89.77**
ATE	76.76	76.90	76.83	79.35	79.35	**79.35**
OTE	72.86	72.84	72.90	73.28	73.53	**73.40**
ACSA	78.25	75.77	76.99	79.07	76.13	**77.57**
ATSA	81.13	77.53	79.28	81.10	79.18	**80.13**
AOPE	74.04	69.81	71.86	74.76	70.65	**72.65**
ASTE	68.76	66.01	67.35	70.90	68.03	**69.44**
ACSD	71.85	69.81	70.82	72.65	70.08	**71.34**
AbSQA	65.24	62.51	63.85	66.11	62.72	**64.37**

luyện mô hình tạo sinh văn bản khi tích hợp lời nhắc hướng dẫn mang lại hiệu suất cao hơn khi không sử dụng. Việc tích hợp lời nhắc hướng dẫn cũng nâng cao hiệu suất khi huấn luyện mô hình theo hướng đa tác vụ ở Bảng 4.4.

Kết quả thí nghiệm cũng cho thấy huấn luyện mô hình theo hướng đa tác vụ cải thiện hiệu suất bài toán. Cụ thể, học đa tác vụ đã nâng cao hiệu suất bài toán AbSQA +2,1% về độ đo F_1 so với hướng đơn tác vụ khi không tích hợp lời nhắc hướng dẫn. Hơn nữa, chiến lược huấn luyện đa tác vụ cũng cải thiện độ đo F_1 cho bài toán AbSQA +2,08% khi tích hợp lời nhắc hướng dẫn. Còn với các bài toán khác, huấn luyện mô hình đa tác vụ đều giúp nâng cao hiệu suất mô hình (khoảng +0,89% đến +3,63% và khoảng +0,39% đến +5,12% khi có và không tích hợp lời nhắc hướng dẫn). Những kết quả này chứng minh sự hiệu quả của chiến lược huấn luyện mô hình tạo sinh văn bản theo hướng đa tác vụ.

Bảng 4.5 trình bày kết quả của phương pháp đề xuất so sánh với các phương pháp khác nhau. Nhìn chung, mô hình tiếp cận theo hướng tạo sinh văn bản (như mô hình luận án, GAS, và Paraphrase) cho hiệu suất tốt hơn các mô

Bảng 4.5. Kết quả các phương pháp trên bài toán phân tích ý kiến theo quan điểm thông thường.

Phương pháp	**Độ chính xác**	**Độ phủ**	**Chỉ số F_1**
Extract-Classify-BERTs ([43])	37.68	48.90	42.58
Two-stages BERTs	48.51	53.37	50.82
GAS-Annotation [44]	57.63	56.84	57.23
GAS-Extraction [44]	62.04	59.45	60.72
Paraphrase [45]	63.19	60.37	61.75
Mô hình luận án	**66.11**	**62.72**	**64.37**

hình hướng trích xuất và phân loại (như Extract-Classify hay Two-Stage BERTs). Điều này chứng minh rằng hướng tiệp cận trên các mô hình tạo sinh giúp chúng ta học mối quan hệ ngữ nghĩa giữa các bộ thuộc tính hiệu quả [45, 44]. Nhìn vào Bảng 4.5, các phương pháp dựa trên BERT chỉ trích xuất các đặc trưng từ văn bản rồi phân loại thuộc tính. Điều này có thể gây ra vấn đề thiếu thông tin về mối quan hệ và ngữ cảnh giữa các từ vựng trong bình luận đầu vào. Đối với các mô hình theo hướng sinh văn bản, kết quả cho thấy mô hình luận án đạt hiệu suất tốt nhất. Mô hình đề xuất cho kết quả vượt trội hơn Parapharase [45] +2,66% và GAS-Extraction [44] +3,65% về độ đo F_1. Kết quả này chứng minh rằng các thành phần được đề xuất cách biểu diễn khuôn mẫu nhãn đầu ra với việc tích hợp lời nhắc hướng dẫn và huấn luyện mô hình theo hướng đa tác vụ giúp nâng cao hiệu suất mô hình.

4.3 PHÂN TÍCH Ý KIẾN THEO QUAN ĐIỂM SO SÁNH

Mục tiêu bài toán là xác định ra các bình luận thể hiện sự so sánh và rút trích bộ năm thuộc tính liên quan sự so sánh. Để giải quyết bài toán, luận án đưa ra một kiến trúc hai giai đoạn: (1) Xác định bình luận so sánh; (2) Rút trích bộ năm thuộc tính.

4.3.1 *Xác định câu bình luận so sánh*

Để xác định bình luận có tính chất so sánh hay không, luận án giải quyết bài toán này như là một bài toán phân loại văn bản nhị phân. Bình luận sau khi tiền xử lý sẽ được tách thành tokens và đưa vào BERT để trích xuất đặc trưng biểu diễn $H = \{h_{cls}, h_1, ..., h_n, h_{sep}\}$ với h_i là đặc trưng biểu diễn

của từ vựng thứ i. Sau đó, đặc trưng biểu diễn của CLS ở lớp cuối cùng là h_{cls} được sử dụng làm vectơ đặc trưng biểu diễn. Đặc trưng này được đưa vào một lớp phân loại với hàm sigmoid để tính xác suất dự đoán:

$$\hat{y} = sigmoid(W \cdot h_{cls} + b) \tag{6}$$

trong đó W, b là ma trận trọng số và độ lệch của lớp phân loại. Hàm mất mát Weighted Binary Cross-Entropy (BCE) được sử dụng để tối ưu.

4.3.2 *Trích xuất bộ năm thuộc tính so sánh*

Mục đích của thành phần này là trích xuất bộ năm thuộc tính so sánh, bao gồm: chủ thể s, đối tượng o, khía cạnh a, vị từ so sánh p, và nhãn so sánh l. Để giải quyết bài toán này, luận án áp dụng mô hình đa tác vụ kết hợp lời nhắc huấn luyện theo hướng đa tác vụ như mô tả ở Mục 4.2.1. Tuy nhiên, nhãn đầu ra được biểu diễn cho bộ năm thuộc tính so sánh dưới dạng một cấu trúc thay vì dưới dạng ngôn ngữ tự nhiên. Lý do là vì bài toán này tập trung trích xuất các thuộc tính từ bình luận đầu vào, do đó, cấu trúc cố định sẽ giúp chuyển đổi và đánh giá mô hình hiệu quả. Ngoài ra, kích thước dữ liệu cũng hạn chế, việc biểu diễn dưới dạng tự nhiên khiến cho mô hình thiếu khả năng khái quát. Do đó, sử dụng cấu trúc biểu diễn như dưới đây không chỉ làm tăng độ chính xác, hiệu quả mà còn giảm thiểu lỗi trong xử lý dữ liệu. Vì vậy, đầu ra bộ năm thuộc tính được biểu diễn theo một cấu trúc cố định với thứ tự như sau:

$$Q = \{s, o, a, p, l\} \tag{7}$$

Các giá trị s, o, a, p là các chuỗi từ vựng hoặc cụm từ được đề cập tường minh hoặc không tường minh (giá trị rỗng) từ trong bình luận đầu vào. Nếu giá trị của thuộc tính rỗng thì sẽ được biểu diễn bằng từ vựng đại diện. Các chỉ số tương ứng của từng phần tử được xác định thông qua thuật toán so khớp chuỗi mờ. Các bình luận có nhiều hơn hai bộ thuộc tính được biểu diễn như sau:

$$\text{Chuỗi đầu ra} = \{(s_1, o_1, a_1, p_1, l_1)\ ;\ ...\ ;\ (s_r, o_r, a_r, p_r, l_r)\} \tag{8}$$

Bảng 4.6. Kết quả mô hình luận án so sánh với các mô hình khác nhau .

Phương pháp	Macro Average			Micro Average		
	Độ chính xác	Độ phủ	Chỉ số F_1	Độ chính xác	Độ phủ	Chỉ số F_1
BERT-LSTM-CRF [8]	5.25	9.98	6.68	9.82	20.93	13.37
BERT2BERT [8]	10.41	8.46	9.23	20.26	16.85	18.40
Three-Stage BERTs [46]	9.68	10.65	9.97	16.75	18.94	17.78
Multi-Stage Framework [47]	9.64	13.75	11.19	17.09	26.98	20.92
Seq2Seq T5	20.93	21.99	21.31	**29.41**	29.41	29.41
Ensemble Framework (GNN,viT5,PhoBERT)	20.21	27.18	23.00	22.34	**33.59**	26.84
Mô hình luận án	**22.16**	**28.62**	**23.73**	28.80	30.29	**29.52**

Thay vì tạo ra các bài toán khác nhau dựa trên sự kết hợp ngẫu nhiên giữa các phần tử, NCS tập trung vào các bài toán mà các phần tử có mối quan hệ tương quan. Điều này giúp mô hình tận dụng kiến thức từ các bài toán cũng như tăng số lượng dữ liệu cho bài toán chính, từ đó giúp cải thiện hiệu suất mô hình. Các lời nhắc hướng dẫn chỉ rõ cho mô hình về bài toán thực hiện và giúp mô hình xác định mối liên hệ giữa các phần tử.

Luận án cũng sử dụng các MHNN mã hóa-giải mã (encoder-decoder) để để huấn luyện theo hướng tạo sinh văn bản. Để hạn chế sự nhạy cảm về vị trí các thuộc tính, luận án áp dụng kỹ thuật tăng cường dữ liệu để tăng kích thước mẫu huấn luyện thông qua nhân bản ngẫu nhiên thuộc tính như sau:

- **Tạo tập từ điển**: Trích xuất các giá trị khác rỗng cho từng phần tử trong bộ năm để tạo thành các tập từ điển.
- **Tạo bình luận với bộ năm thuộc tính**:NCS ngẫu nhiên thay thế chủ thể, đối tượng, khía cạnh và vị ngữ bằng giá trị từ bộ từ điển, đảm bảo nhãn so sánh nhất quán.
- **Cân bằng và chọn lọc dữ liệu**: Tập dữ liệu tăng cường được cân bằng theo nhãn so sánh, tự động loại bỏ mẫu trùng lặp (độ tương đồng > 0.8) hoặc sai thứ tự nhãn bộ năm ý kiến.

4.3.3 *Kết quả thử nghiệm*

Kết quả so sánh hiệu suất mô hình luận án với các hướng tiếp cận khác nhau trên tập dữ liệu kiểm tra ở Bảng 4.6. Các mô hình BERT-LSTM-CRF và BERT2BERT cho kết quả độ đo F_1 macro thấp nhất. Điều này chỉ ra sự không hiệu quả của các kiến trúc này đối với một bài toán này. Các mô hình trên không đủ mạnh mẽ để rút trích được các đặc trưng cần thiết nên

Bảng 4.7. Kết quả chi tiết các biến thể mô hình khác nhau trong luận án.

Phương pháp	Macro Average			Micro Average		
	Độ chính xác	**Độ phủ**	**Chỉ số F_1**	**Độ chính xác**	**Độ phủ**	**Chỉ số F_1**
Đơn tác vụ	17.57	15.79	16.61	27.39	25.22	26.26
Đơn tác vụ + DA	15.95	14.41	15.10	28.15	26.10	27.09
Đơn tác vụ + Prompt	17.21	16.03	16.59	29.86	28.85	29.34
Đơn tác vụ + Prompt + DA	17.15	16.38	16.72	**30.09**	28.30	29.17
Đa tác vụ	20.55	15.63	17.64	27.58	21.81	24.35
Đa tác vụ + DA	21.11	16.10	18.17	27.96	22.36	24.85
Đa tác vụ + Prompt	24.17	21.80	22.51	28.13	29.41	28.76
Đa tác vụ + Prompt + DA	**28.62**	**22.16**	**23.73**	28.80	**30.29**	**29.52**

kém hiệu quả trên tập kiểm tra. Đối với các phương pháp đa giai đoạn như Three-Stage BERTs [46] hay Multi-Stage Framework [47] cũng cải thiện kết quả tổng quan. Một điểm đặc biệt là hướng tiếp cận Seq2Seq T5 đạt kết quả tương đương với luận án về độ đo micro F_1. Tuy nhiên, xét về kết quả trên độ đo F_1, phương pháp của luận án vượt trội hơn Seq2Seq T5 khoảng +2% về độ đo macro F_1. NCS cũng nhận thấy rằng phương pháp kết hợp - Ensemble trên GNN, viT5, và PhoBERT có hiệu suất cạnh tranh với luận án về độ đo macro F_1. Tuy nhiên, phương pháp này đòi hỏi một lượng lớn tài nguyên tính toán để huấn luyện và thực thi.

Bảng 4.7 trình bày hiệu suất của các biến thể từ các thành phần của mô hình luận án. Kết quả cho thấy sử dụng lời nhắc hướng dẫn nâng cao đáng kể hiệu suất tổng thể của mô hình, đặc biệt là đối với hướng tiếp cận đa tác vụ, so với các mô hình không tích hợp lời nhắc hướng dẫn. Bên cạnh đó, kỹ thuật tăng cường dữ liệu cũng góp phần cải thiện hiệu suất trên độ đo micro F_1 nhưng không đáng kể. Nguyên nhân là do kỹ thuật tăng cường còn khá đơn giản khi chỉ tạo ra các mẫu dựa trên các phần tử ý kiến xoay quanh tập huấn luyện nên không đa dạng được các mẫu ở tập kiểm tra. Kết quả cũng chứng minh rằng huấn luyện theo hướng đa tác vụ giúp cải thiện hiệu suất so với hướng đơn tác vụ, đặc biệt là độ đo macro F_1.

4.4 KẾT CHƯƠNG

Nội dung nghiên cứu đã được công bố tại một tạp chí quốc tế ở CT.03 và tạp chí trong nước CT.07.

5. PHÂN TÍCH Ý KIẾN VỚI DỮ LIỆU HẠN CHẾ

5.1 ĐỘNG LỰC NGHIÊN CỨU

Trong chương này, luận án nghiên cứu các phương pháp khác nhau trong vấn đề hạn chế dữ liệu huấn luyện và nâng cao kết quả dự đoán như sau:

- Kết hợp mô hình đa ngôn ngữ trên ngôn ngữ giàu tài nguyên với kịch bản: (1) Học chuyển tiếp chéo ngôn ngữ không dữ liệu, (2) Học chuyển tiếp chéo ngôn ngữ huấn luyện chung.
- Nghiên cứu kỹ thuật thiết kế lời nhắc (prompt engineering) với các mô hình ngôn ngữ lớn trong kịch bản không sử dụng dữ liệu huấn luyện.

5.2 PP1: MÔ HÌNH ĐA NGÔN NGỮ VỚI NGÔN NGỮ GIÀU TÀI NGUYÊN

5.2.1 *Học chuyển tiếp chéo ngôn ngữ không dữ liệu*

Trong lĩnh vực XLNN-TN, thuật ngữ học chuyển tiếp chéo ngôn ngữ không dữ liệu (zero-shot cross-lingual transfer learning) có nghĩa là chuyển tiếp mô hình đã huấn luyện trên ngôn ngữ nguồn được sử dụng để dự đoán trên ngôn ngữ khác mà không cần dữ liệu huấn luyện ở ngôn ngữ đích [48, 49, 50, 51]. Vì vậy, các bước thực hiện như sau: (1) Huấn luyện mô hình đa ngôn ngữ trên dữ liệu gán nhãn ở ngôn ngữ giàu tài nguyên (nguồn); (2) Chuyển tiếp trọng số tri thức của mô hình để dự đoán kết quả trên ngôn ngữ ít tài nguyên (đích). Khác với nghiên cứu trước chỉ sử dụng tiếng Anh là ngôn ngữ chính để chuyển tiếp thì luận án tận dụng nguồn tài nguyên ở nhiều ngôn ngữ.

5.2.2 *Học chuyển tiếp chéo ngôn ngữ huấn luyện chung*

Huấn luyện chung chéo ngôn ngữ (joint training cross-lingual learning) là mô hình được huấn luyện trên sự kết hợp dữ liệu của hai hoặc nhiều ngôn ngữ khác nhau. Ví dụ, với hai dữ liệu được gán nhãn trên tiếng Anh và tiếng Pháp có ký hiệu là L_{en} và L_{fr}. Theo phương pháp này, NCS kết hợp L_{en}

Bảng 5.1. Kết quả học chuyển tiếp chéo ngôn ngữ không dữ liệu khi sử dụng một ngôn ngữ nguồn.

Ngôn ngữ nguồn	**Ngôn ngữ đích**				
	Tiếng Anh	Tiếng Pháp	Tiếng Hà Lan	Tiếng Tây Ban Nha	Tiếng Nga
Tiếng Anh	-	**53.74**	**57.52**	**61.88**	60.00
Tiếng Pháp	58.41	-	56.86	59.85	57.52
Tiếng Hà Lan	**62.81**	53.66	-	61.32	60.47
Tiếng Tây Ban Nha	60.10	53.35	56.50	-	**61.25**
Tiếng Nga	56.35	44.74	51.35	55.56	-

và L_{fr} để huấn luyện mô hình đa ngôn ngữ. Sau đó, mô hình này dự đoán trên dữ liệu ở ngôn ngữ ít tài nguyên như L_{fr}. Các nghiên cứu trước đây [52, 53, 54, 55] đã chứng minh hiệu quả phương pháp này ở các bài toán khác nhau. Các bước thực hiện như sau: (1) Kết hợp bộ dữ liệu huấn luyện của nhiều ngôn ngữ tạo thành tập huấn luyện mới; (2) Áp dụng phương pháp học chuyển tiếp dựa trên mô hình đa ngôn ngữ, (3) Đánh giá hiệu quả phương pháp trên bộ dữ liệu kiểm tra của từng ngôn ngữ.

5.2.3 *Kết quả thử nghiệm*

Học chuyển tiếp ngữ chéo ngôn ngữ không dữ liệu: Dựa vào bảng 5.1 khi sử dụng một ngôn ngữ nguồn, mô hình huấn luyện trên tiếng Anh cho hiệu suất tốt nhất trên tiếng Pháp, Hà Lan và Tây Ban Nha. Hơn nữa, NCS cũng thấy rằng kết quả của mô hình XLM-Align khác nhau cho một số cặp ngôn ngữ. Vì vậy, lựa chọn ngôn ngữ nguồn có tác động đáng kể đến hiệu quả của phương pháp này. Cụ thể, mô hình huấn luyện trên dữ liệu tiếng Anh (thuộc nhóm Germanic) đạt kết quả vượt trội với các ngôn ngữ đích có sự tương đồng về hình thái hoặc nguồn gốc từ, như tiếng Pháp và Tây Ban Nha (thuộc nhóm Romance), và đặc biệt là tiếng Hà Lan (cùng nhóm Germanic). Trong đó, tiếng Hà Lan đạt hiệu suất cao nhất trên tiếng Anh nhờ sự tương đồng lớn về cú pháp và từ vựng giữa hai ngôn ngữ. Ngược lại, khi sử dụng tiếng Nga (thuộc nhóm Slavic) làm ngôn ngữ nguồn, kết quả giảm đáng kể đối với tất cả các ngôn ngữ đích, đặc biệt là các ngôn ngữ Romance như tiếng Pháp và tiếng Tây Ban Nha. Những quan sát này củng cố giả thuyết rằng tính tương đồng ngôn ngữ có vai trò quan trọng trong phương pháp chuyển tiếp chéo ngôn ngữ.

Bảng 5.2. Kết quả học chuyển tiếp không dữ liệu trên nhiều ngôn ngữ nguồn. KQ@1 là kết quả của một ngôn ngữ nguồn.

Nguồn ⇒ Đích	**Độ chính xác**	**Độ phủ**	**Micro F_1**	**KQ@1**
fr + ru + nl + es ⇒ en	71.14	67.49	69.27 (+6.46)	62.81
en + nl + es + ru ⇒ fr	64.08	58.96	61.41 (+7.67)	53.74
en + fr + es + ru ⇒ nl	63.62	63.98	62.81 (+5.29)	57.52
en + fr + nl + ru ⇒ es	66.09	68.23	67.14 (+5.26)	61.88
en + fr + nl + es ⇒ ru	65.96	63.29	64.60 (+3.35)	61.25

Luận án cũng thử nghiệm kết quả khi kết hợp nhiều ngôn ngữ nguồn khác nhau. Cụ thể, Bảng 5.2 trình bày kết quả khi sử dụng dữ liệu của thử nghiệm này. Kết quả cho thấy rằng việc kết hợp dữ liệu từ nhiều ngôn ngữ nguồn cải thiện hiệu suất dự đoán so với duy nhất một ngôn ngữ nguồn. Kết quả này cho thấy rằng việc tận dụng dữ liệu đa ngôn ngữ nguồn giúp mô hình học được nhiều đặc trưng ngữ nghĩa và hình thái hơn, từ đó cải thiện khả năng tổng quát hóa sang các ngôn ngữ đích. Bên cạnh đó đối với bài toán phân tích theo khía cạnh thì điều này làm tăng số lượng mẫu huấn luyện ở một số khía cạnh ít dữ liệu. Bảng số liệu 5.2 cũng chỉ ra sự cải thiện thấp hơn ở tiếng Nga cho thấy rằng khi có sự khác biệt lớn giữa ngôn ngữ nguồn và các ngôn ngữ đích (Slavic so với Romance hoặc Germanic), do đó, mô hình không cải thiện hiệu suất lớn như các ngôn ngữ khác.

Học chuyển tiếp ngữ chéo ngôn ngữ huấn luyện chung: Bảng 5.3 trình bày kết quả của việc huấn luyện chung cho các cặp ngôn ngữ. Đường chéo chính là kết quả khi huấn luyện và đánh giá mô hình trên cùng dữ liệu đích. Còn các giá trị còn lại trong mỗi cột thể hiện hiệu suất khi kết hợp dữ liệu của một cặp ngôn ngữ nguồn và đích. Nhìn chung, phương pháp cải thiện hiệu suất của mô hình nhưng sự cải thiện phụ thuộc vào cặp ngôn ngữ. Luận án nhận thấy rằng việc huấn luyện các cặp ngôn ngữ trong cùng một nhóm ngôn ngữ liên quan đến mối quan hệ ngôn ngữ như tiếng Anh với tiếng Hà Lan (cặp ngôn ngữ Germanic), tiếng Pháp và tiếng Tây Ban Nha (cặp ngôn ngữ Romance) mang lại lợi ích hơn so với các cặp khác.

Bảng 5.4 trình bày kết quả khi huấn luyện trên bộ dữ liệu tổng hợp năm ngôn ngữ. Vì mô hình đa ngôn ngữ sử dụng chung một không gian biểu

Bảng 5.3. Kết quả phương pháp học chuyển tiếp chéo ngôn ngữ huấn luyện chung trên cặp ngôn ngữ nguồn và ngôn ngữ đích.

Ngôn ngữ nguồn	**Ngôn ngữ đích**				
	Tiếng Anh	Tiếng Pháp	Tiếng Hà Lan	Tiếng Tây Ban Nha	Tiếng Nga
Tiếng Anh	69.76	64.74	**68.51**	68.68	71.38
Tiếng Pháp	71.23	61.86	67.92	**69.98**	70.29
Tiếng Hà Lan	**73.35**	63.59	63.79	69.02	69.33
Tiếng Tây Ban Nha	72.32	**65.50**	67.80	68.65	**72.20**
Tiếng Nga	73.03	60.78	65.27	69.84	68.26

Bảng 5.4. Kết quả khi huấn luyện chung nhiều ngôn ngữ. KQ@0 và KQ@1 là kết quả tốt nhất khi chỉ huấn luyện trên ngôn ngữ đích và dữ liệu kết hợp một cặp ngôn ngữ nguồn và đích.

Nguồn ⇒ Đích	**Độ chính xác**	**Độ phủ**	**Micro F_1**	**KQ@0**	**KQ@1**
en + fr + nl + es + ru ⇒ en	76.13	73.95	75.02 (+5.26)	69.76	73.35 (+3.59)
en + fr + nl + es + ru ⇒ fr	68.50	65.04	66.73 (+4.87)	61.86	65.50 (+3.64)
en + fr + nl + es + ru ⇒ nl	70.70	72.42	71.55 (+7.76)	63.79	68.51 (+4.72)
en + fr + nl + es + ru ⇒ es	70.63	71.77	71.20 (+2.55)	68.65	69.98 (+1.33)
en + fr + nl + es + ru ⇒ ru	74.22	72.41	73.31 (+5.05)	68.26	72.20 (+3.94)

diễn cho các token ở nhiều ngôn ngữ khác nhau. Do đó, các từ mang ý nghĩa tương tự sẽ có vị trí gần nhau trong không gian, giúp mô hình đa ngôn ngữ có khả năng liên kết các ý nghĩa tương đồng. Nhìn vào Bảng 5.4 thấy rằng kết quả cải thiện một cách đáng kể khi huấn luyện mô hình trên dữ liệu nhiều ngôn ngữ khác nhau. Điều này chứng tỏ việc tận dụng sức mạnh của mô hình đa ngôn ngữ với các bộ dữ liệu đã được gán nhãn ở các ngôn ngữ giàu tài nguyên có thể cải thiện hiệu suất ở các ngôn ngữ ít tài nguyên hoặc hạn chế dữ liệu.

5.3 PP2: THIẾT KẾ LỜI NHẮC VỚI MÔ HÌNH NGÔN NGỮ LỚN

Các mô hình ngôn ngữ lớn (LLM) đã giúp máy tính hiểu ngôn ngữ con người với hiệu suất ngày càng cao [56, 57]. Khi được huấn luyện trên ngữ lớn lớn, các mô hình này hiểu các thông tin ngữ nghĩa, cấu trúc theo ngữ cảnh văn bản hiệu quả. Đặc biệt, một số nghiên cứu gần đây đã tập trung nghiên cứu giải pháp tận dụng sức mạnh của LLM cho các bài toán phân tích ý kiến, tuy nhiên chỉ tập trung tiếng Anh [58, 59, 60, 61, 62, 63, 64]. Vì vậy, nghiên cứu khả năng LLM cho các ngôn ngữ ít tài nguyên là một chủ đề nghiên cứu cần thiết. Vì vậy, luận án nghiên cứu các kỹ thuật thiết kế lời nhắc trên các LLM mã nguồn mở (Llama, SeaLLM) và mã nguồn

đóng (GPT 3.5, GPT 4, GPT 4o) cho bài toán phân tích ý kiến theo mức độ.

5.3.1 *Phương pháp thiết kế lời nhắc - Prompt Engineering*

Các LLM tạo ra các phản hồi khác nhau tùy thuộc vào thông tin được cung cấp thông qua lời nhắc prompt. Do đó, cần thiết kế prompt hiệu quả để tận dung trí thức có sẵn của các mô hình LLM [65]. Một prompt được thiết kế tốt giúp các LLM tạo ra phản hồi chính xác. Luận án nghiên cứu ba cách thiết kế mẫu lời nhắc prompt trên tiếng Anh và tiếng Việt như sau:

- **Câu hỏi trực tiếp (Direct Question)**: Lời nhắc promt này hiệu quả cho nhiệm vụ yêu cầu câu trả lời cụ thể, giảm mơ hồ và phù hợp với các tình huống đơn giản cần rõ ràng.
- **Hướng dẫn gán nhãn (Labeling Instruction)**: Hướng dẫn rõ ràng giúp mô hình hiểu yêu cầu, đảm bảo nhất quán và chính xác trong phản hồi.
- **Nhập vai (Role-playing)**: Gán vai trò cụ thể, như chuyên gia phân tích, để khai thác khả năng của mô hình ngôn ngữ lớn.

Bên cạnh đó, NCS cũng nghiên cứu các chiến lược thiết kế lời nhắc, bao gồm không mẫu (zero-shot prompting) [39, 66] và ít mẫu (few-shot prompting) [67] như sau:

- **Nhắc không mẫu (Zero-shot Prompting)**: Yêu cầu mô hình thực hiện nhiệm vụ dựa trên hướng dẫn, không cần ví dụ mẫu bài toán.
- **Nhắc ít mẫu (Few-shot Prompting)**: Kỹ thuật này sử dụng các ví dụ k-shot trong prompt để cải thiện khả năng học theo ngữ cảnh, với các ví dụ ngẫu nhiên từ dữ liệu huấn luyện cho mỗi nhãn ý kiến.

5.3.2 *Kết quả thử nghiệm*

Luận án sử dụng các mô hình LLM: mã nguồn đóng (GPT 3.5, GPT 4 và GPT 4o) và mã nguồn mở (Llama-3 8B [68], SeaLLM v3 7B [69]). Kết quả thử nghiệm trên các mức độ dữ liệu khác nhau cho thấy thiết kế prompt

theo dạng "Role-playing" cho kết quả các độ đo cao hơn các cách thiết kế prompt còn lại trên cả hai ngôn ngữ. Kết quả này chỉ ra rằng thiết kế prompt theo dạng "Role-playing" khuyến khích mô hình hiểu được nhiệm vụ thực hiện tốt hơn. Hơn nữa, thiết kế "Role-playing" giúp chúng ta tương tác với các mô hình LLM một cách tự nhiên, từ đó dẫn đến hiệu suất các bài toán XLNN-TN tốt hơn [70]. Kết quả cho thấy prompt tiếng Anh thường hiệu quả hơn tiếng Việt, đặc biệt với các mô hình lớn như GPT-4. Tuy nhiên, sự chênh lệch không lớn, và với GPT-4, prompt tiếng Việt cho kết quả tốt hơn trên bộ dữ liệu VLSP. Điều này do các mô hình được huấn luyện chủ yếu trên dữ liệu tiếng Anh.

Kết quả thử nghiệm cho thấy GPT-4 cho kết quả tốt hơn so với GPT-3.5 và GPT-4o trên hầu hết các bộ dữ liệu. So sánh trên kết quả trung bình, GPT-4 luôn vượt trội hơn hai mô hình còn lại trong cả ba mức độ dữ liệu. Bên cạnh đó, kết quả thử nghiệm cho thấy rằng sự ảnh hưởng của thiết kế prompt không ảnh hưởng lớn đến hiệu suất khi sử dụng các LLM như GPT-4 và GPT-4o. Lý do là do hai mô hình này được cải thiện khả năng hiểu ngôn ngữ một cách hiệu quả. Còn so sánh kết quả với hai mô hình LLM mã nguồn mở (Llama-3 8B Instruct và Seallm v3 7B), các mô hình mã nguồn đóng kết quả vượt trội trong kịch bản không dữ liệu (xem Bảng 5.5).

Từ kết quả Bảng 5.5, áp dụng chiến lược k-shot cải thiện hiệu suất đáng kể so với chiến lược zero-shot trên hầu hết các mô hình. Tuy nhiên, luận án cũng quan sát thấy xu hướng giảm hiệu suất ở một số mô hình có kích thước tham số lớn như GPT-4 và GPT-4o trên bộ dữ liệu HSA khi tăng số lượng mẫu ví dụ. Điều này có thể được giải thích do hiện tượng overfitting trong cách chiến lược thiết kế prompt, khi mô hình dựa quá nhiều vào các ví dụ được cung cấp thay vì hiệu một cách tổng quan bài toán cần thực hiện. Đối với VLSP, chiến lược few-shot cũng cải thiện kết quả nhưng sự khác biệt không đáng kể trên ba mô hình LLM. Nguyên nhân là do bộ dữ

Bảng 5.5. Kết quả của các kỹ thuật thiết kế prompt Role-playing chiến lược few-shot khác nhau.

Model	Phân tích ý kiến ở mức độ tài liệu hoặc câu						Phân tích ý kiến ở mức độ khía cạnh					
	UIT-VSFC		HSA		VLSP		Nhà hàng		Khách sạn		Điện thoại	
	Macro F_1	Micro F_1	Macro F_1	Micro F_1	Macro F_1	Micro F_1	Macro F_1	Micro F_1	Macro F_1	Micro F_1	Macro F_1	Micro F_1
Llama-3 8B Instruct												
0-Shot	59.78	68.41	66.71	69.59	65.80	65.90	56.08	60.50	70.21	82.23	67.71	77.80
1-Shot	70.01	84.30	73.19	79.43	68.76	68.95	54.33	60.33	71.39	84.00	65.53	77.39
3-Shot	70.88	84.14	71.58	79.12	66.93	68.10	55.30	61.17	71.59	84.16	65.58	77.39
5-Shot	75.96	87.05	70.19	80.03	69.39	69.90	-	-	-	-	-	-
Sea-LLM v3 7B												
0-Shot	63.89	75.36	63.44	69.44	46.08	52.10	36.94	46.00	56.13	76.05	51.64	67.64
1-Shot	65.76	78.49	70.09	77.31	46.70	50.38	45.93	54.50	65.94	82.88	59.46	74.95
3-Shot	71.72	83.86	70.75	75.64	49.82	52.38	45.29	54.50	64.49	82.80	59.86	74.34
5-Shot	71.76	85.06	70.86	77.76	56.14	57.14	-	-	-	-	-	-
GPT 3.5												
0-Shot	71.69	84.65	63.60	80.79	56.14	63.24	56.68	64.50	69.50	82.13	60.16	76.99
1-Shot	74.30	87.21	70.11	81.45	69.52	71.05	63.75	66.00	71.06	83.76	58.39	74.54
3-Shot	72.97	85.79	71.73	80.94	67.33	69.71	64.56	68.17	69.89	81.35	61.35	74.95
5-Shot	73.69	86.83	71.01	81.54	69.10	71.14	-	-	-	-	-	-
GPT 4o												
0-Shot	68.96	81.21	74.74	80.33	72.01	72.67	71.90	74.33	73.36	84.89	72.53	83.80
1-Shot	74.72	86.77	76.46	81.85	77.22	77.14	71.84	74.17	73.88	85.23	73.11	**84.26**
3-Shot	76.09	88.66	75.29	80.03	76.20	76.38	74.74	76.67	72.32	82.80	**75.07**	82.28
5-Shot	77.41	89.86	75.38	79.73	77.70	77.62	-	-	-	-	-	-
GPT 4												
0-Shot	69.31	82.93	76.74	**83.06**	74.15	74.76	71.42	74.83	72.71	85.93	70.26	81.87
1-Shot	76.46	89.01	**76.93**	82.45	75.68	76.10	72.34	75.00	**74.99**	**86.50**	70.58	82.08
3-Shot	77.77	89.51	75.36	80.79	75.62	76.48	**75.66**	**77.67**	73.71	85.13	74.86	83.71
5-Shot	**80.41**	**91.25**	75.16	80.18	**77.57**	**77.71**	-	-	-	-	-	-

liệu VLSP là một bộ dữ liệu thách thức, được gán nhãn ở mức tài liệu và chứa nhiều lỗi về từ vựng, cú pháp và ngữ pháp. Llama-3 8B Instruct cũng cho kết quả tốt hơn GPT-3.5 ở miền dữ liệu khách sạn và điện thoại. Hơn nữa, kết quả thử nghiệm cho thấy rằng việc tăng số lượng ví dụ k-shot cải thiện hiệu suất trên hầu hết ở các bộ dữ liệu khác nhau đối với các mô hình LLM khác nhau. Kết quả của luận án tương đồng với các nghiên cứu trước đó [58] trong ngôn ngữ tiếng Anh.

So sánh với phương pháo học có giám sát, phương pháp thiết kế prompt với LLM cho kết quả cạnh tranh ở mức độ dữ liệu tài liệu và câu bình luận. Còn đối với mức độ khía cạnh thì kết quả vẫn còn khoảng cách lớn. Tuy nhiên, kết quả đánh giá vẫn tưởng đối khả quan trong kịch bản không dữ liệu huấn luyện. Điều này cho phép mở rộng ra các miền dữ liệu bình luận khác.

5.4 KẾT LUẬN CHƯƠNG

Các kết quả nghiên cứu trong chương này được trình bày ở CT.05 và CT.09.

6. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

6.1 KẾT LUẬN

Luận án có các đóng góp chính như sau:

- **Đóng góp #1:** Nghiên cứu phương pháp cải thiện độ chính xác cho bài toán phân tích ý kiến theo mức độ với(1) phương pháp học chuyển tiếp dựa trên MHNN theo hướng phân loại; (2) phương pháp học chuyển tiếp dựa trên MHNN theo hướng tạo sinh văn bản; (3) phương pháp tổ hợp dựa trên các MHNN hướng phân loại. Kết quả nghiên cứu được công bố tại các công trình CT.01, CT.04, CT.06, và CT.08.
- **Đóng góp #2:** Xây dựng mô hình hiệu quả cho bài toán phân tích ý kiến theo quan điểm thông thường và so sánh. Kết quả nghiên cứu được công bố ở công trình CT.03 và CT.07.
- **Đóng góp #3:** Nghiên cứu các phương pháp cho vấn đề hạn chế dữ liệu huấn luyện trong bài toán phân tích ý kiến. Các kết quả nghiên cứu được trình bày ở CT.05 và CT.09.

6.2 HƯỚNG PHÁT TRIỂN

Hướng phát triển tiếp theo của luận án như sau:

- Tích hợp các quy tắc cảm xúc theo đặc trưng ngôn ngữ như đặc trưng về cú pháp, đặc trưng từ loại, cấu trúc ngữ nghĩa hoặc tri thức ngoài vào mô hình.
- Tận dụng sức mạnh của các mô hình ngôn ngữ lớn như là Llama-3 [71], Vina-Llama [72] để nâng cao hiệu quả dự đoán.
- Các bình luận mỉa mai hoặc cần suy luận để xác định được ý nghĩa ý kiến cũng là một trong những vấn đề thách thức cần được xử lý trong tiếng Việt.