HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

Trương Tuấn Anh

ỨNG DỤNG HỌC TĂNG CƯỜNG TRONG PHÂN TÍCH CẢM XÚC THEO CHỦ ĐỀ

CHUYÊN NGÀNH : KHOA HỌC MÁY TÍNH

MÃ SỐ: 8.48.01.01

TÓM TẮT ĐỀ ÁN TỐT NGHIỆP THẠC SĨ

HÀ NỘI – 2025

Đề án tốt nghiệp được hoàn thành tại:

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

Người hướng dẫn khoa học: TS. Nguyễn Thị Mai Trang

Phản biện 1: ……

Phản biện 2: ...……..

Đề án tốt nghiệp sẽ được bảo vệ trước Hội đồng chấm đề án tốt nghiệp thạc sĩ tại Học viện Công nghệ Bưu chính Viễn thông

Vào lúc: ..... giờ ..... ngày ..... tháng .... năm .....

Có thể tìm hiểu đề án tốt nghiệp tại:

- Thư viện của Học viện Công nghệ Bưu chính Viễn thông.

MỞ ĐẦU

1. Lý do chọn đề tài

Trong thời đại chuyển đổi số và bùng nổ dữ liệu, đặc biệt là sự gia tăng mạnh mẽ của nội dung do người dùng tạo ra (User-Generated Content), ý kiến khách hàng trực tuyến được xem là một nguồn tài nguyên vô giá. Tại Việt Nam, sự tăng trưởng vũ bão của các nền tảng thương mại điện tử như Shopee, Lazada, Tiki cùng với sự phổ biến của mạng xã hội như Facebook, TikTok đã tạo ra một lượng dữ liệu đánh giá, bình luận khổng lồ từ người dùng. Nguồn dữ liệu này, nếu được khai thác hiệu quả, sẽ mang lại thông tin thiết thực để doanh nghiệp cải tiến sản phẩm, dịch vụ, tối ưu hóa chiến lược marketing và nâng cao trải nghiệm người dùng, từ đó tạo lợi thế cạnh tranh bền vững.

Tuy nhiên, dữ liệu ý kiến khách hàng thường có dạng phi cấu trúc, phức tạp, chứa đựng nhiều yếu tố cảm xúc đa dạng và đôi khi khó nắm bắt. Đặc thù của tiếng Việt với tính đa nghĩa, cấu trúc ngữ pháp linh hoạt và sự phong phú trong cách diễn đạt cảm xúc (bao gồm cả tiếng lóng, sắc thái mỉa mai) càng làm tăng độ phức tạp cho bài toán phân tích tự động, khiến việc phân tích thủ công trở nên không chỉ khó khăn, tốn kém thời gian mà còn dễ gặp phải sai sót chủ quan. Chính vì vậy, việc ứng dụng các công nghệ trí tuệ nhân tạo (AI), đặc biệt là học máy (Machine Learning), và sâu hơn nữa là học tăng cường (Reinforcement Learning), đang trở thành một xu hướng quan trọng và tất yếu trong nghiên cứu và phát triển các giải pháp phân tích dữ liệu văn bản thông minh.

Học tăng cường là một phương pháp học tự động dựa trên cơ chế thử - sai và tương tác liên tục với môi trường để học một chính sách hành động tối ưu. Trong đó, mô hình không chỉ học cách phân loại một cách tĩnh mà còn liên tục điều chỉnh chiến lược của mình để đạt được hiệu quả cao nhất thông qua các tín hiệu phản hồi (phần thưởng). Khác với các phương pháp học máy truyền thống thường yêu cầu một lượng lớn dữ liệu đã được gắn nhãn đầy đủ và chi tiết, học tăng cường cho phép hệ thống học hiệu quả ngay cả từ dữ liệu bán giám sát hoặc dữ liệu tự nhiên chưa qua xử lý sâu. Điều này đặc biệt phù hợp với bối cảnh phân tích ý kiến khách hàng, nơi dữ liệu đầu vào có thể vô cùng đa dạng, thiếu cấu trúc, và việc gắn nhãn chính xác từng khía cạnh, từng sắc thái cảm xúc là một công việc đòi hỏi nhiều nguồn lực.

Ngoài ra, việc phân loại ý kiến khách hàng không chỉ dừng lại ở việc xác định cảm xúc tổng thể (tích cực, tiêu cực hay trung tính) mà còn cần mở rộng sang nhận diện và phân tích cảm xúc đối với các khía cạnh (aspects) cụ thể liên quan, chẳng hạn như chất lượng sản phẩm,

dịch vụ chăm sóc khách hàng, giá cả, hay thậm chí là các tính năng chi tiết. Việc nhận diện được các khía cạnh cụ thể này và cảm xúc đi kèm giúp doanh nghiệp hiểu rõ hơn về nhu cầu, mong muốn, cũng như những điểm hài lòng và chưa hài lòng của khách hàng, từ đó đưa ra các chiến lược cải tiến trúng đích. Học tăng cường, với khả năng tối ưu hóa qua từng bước học tập và ra quyết định tuần tự, sẽ là công cụ hữu hiệu để giải quyết bài toán phức tạp này, có khả năng vượt qua những giới hạn của các phương pháp học có giám sát hoặc không giám sát truyền thống vốn có thể gặp khó khăn trong việc mô hình hóa mối quan hệ động giữa các yếu tố trong văn bản.

Với tất cả các lý do trên, đề tài "Ứng dụng học tăng cường trong phân tích cảm xúc theo chủ đề" không chỉ mang tính cấp thiết trong bối cảnh hiện tại mà còn tạo ra cơ hội để ứng dụng các công nghệ AI hiện đại vào giải quyết một bài toán thực tế có ý nghĩa, đồng thời khẳng định và khám phá sâu hơn vai trò của học tăng cường trong việc giải quyết các bài toán phức tạp trong lĩnh vực xử lý dữ liệu ngôn ngữ tự nhiên, đặc biệt với tiếng Việt.

2. Mục đích nghiên cứu

Mục đích của nghiên cứu này là phát triển và ứng dụng phương pháp học tăng cường (Reinforcement Learning - RL) để xây dựng một hệ thống phân tích ý kiến khách hàng. Hệ thống này có khả năng tự động phân loại cảm xúc trong ý kiến khách hàng (tích cực, tiêu cực, trung tính) và nhận diện các khía cạnh cụ thể như chất lượng sản phẩm, dịch vụ, giá cả hay tính năng.

3. Đối tượng và phạm vi nghiên cứu

Đối tượng nghiên cứu

Đề tài tập trung vào dữ liệu ý kiến khách hàng dưới dạng văn bản phi cấu trúc, chứa thông tin về cảm xúc và các khía cạnh liên quan đến sản phẩm hoặc dịch vụ. Ngoài ra, đối tượng còn bao gồm các thuật toán học tăng cường, như Q-Learning, Deep Q-Network (DQN), và Actor-Critic, được áp dụng để xây dựng hệ thống phân tích cảm xúc và nhận diện khía cạnh.

Phạm vi nghiên cứu

Phạm vi nghiên cứu tập trung vào việc phát triển và thử nghiệm mô hình học tăng cường trên dữ liệu tiếng Việt và tiếng Anh từ các nền tảng trực tuyến như sàn thương mại điện tử hoặc mạng xã hội. Đề tài tập trung vào phân tích cảm xúc (tích cực, tiêu cực, trung

tính) và các khía cạnh cụ thể (sản phẩm, dịch vụ, giá cả), với ứng dụng chính trong thương mại điện tử và chăm sóc khách hàng.

4. Phương pháp nghiên cứu

Nghiên cứu lý thuyết

- Tìm hiểu nền tảng về học tăng cường (Reinforcement Learning): Nghiên cứu các thuật toán học tăng cường, bao gồm Q-Learning, Deep Q-Network (DQN), và các phương pháp nâng cao như Actor-Critic. Phân tích các ưu điểm và hạn chế của các thuật toán này khi áp dụng vào bài toán phân tích dữ liệu ngôn ngữ tự nhiên.

- Phân tích ý kiến khách hàng: Nghiên cứu các phương pháp truyền thống trong xử lý ngôn ngữ tự nhiên (NLP) như Word2Vec, GloVe, hoặc BERT để trích xuất đặc trưng văn bản. Tìm hiểu các kỹ thuật phân loại cảm xúc và nhận diện khía cạnh dựa trên học có giám sát, không giám sát và bán giám sát.

Thu thập và xử lý dữ liệu

- Thu thập dữ liệu: Lựa chọn và thu thập tập dữ liệu phản hồi khách hàng từ các nguồn khác nhau như sàn thương mại điện tử và các khảo sát trực tuyến.

- Tiền xử lý dữ liệu: Thực hiện các bước tiền xử lý để chuẩn hóa dữ liệu, bao gồm loại bỏ từ dư thừa, xử lý dấu câu, từ lóng và biểu tượng cảm xúc

Xây dựng mô hình nghiên cứu

- Thiết kế mô hình học tăng cường: Xây dựng hệ thống học tăng cường với môi trường mô phỏng, trong đó mỗi hành động tương ứng với việc phân loại cảm xúc hoặc nhận diện một khía cạnh cụ thể. Sử dụng phần thưởng (reward) để đánh giá hiệu quả của từng hành động.

- Tích hợp với xử lý ngôn ngữ tự nhiên: Kết hợp các mô hình NLP như BERT hoặc các biểu diễn từ (word embeddings) với học tăng cường để trích xuất và xử lý đặc trưng từ dữ liệu văn bản.

Thực nghiệm và đánh giá

- Thiết lập môi trường thực nghiệm: Chia tập dữ liệu thành các tập huấn luyện, kiểm tra và đánh giá.

- So sánh với các phương pháp truyền thống: Đánh giá hiệu quả của mô hình học tăng cường so với các phương pháp học máy truyền thống.

- Phân tích kết quả: Sử dụng các công cụ trực quan hóa để phân tích kết quả thực nghiệm, xác định các yếu tố ảnh hưởng đến hiệu suất của mô hình và đề xuất hướng cải tiến.

CHƯƠNG 1: TỔNG QUAN VỀ BÀI TOÁN

1.1. Giới thiệu về bài toán

Phân tích cảm xúc theo chủ đề (Aspect-Based Sentiment Analysis – ABSA) là một lĩnh vực quan trọng trong xử lý ngôn ngữ tự nhiên (NLP), tập trung vào việc xác định và đánh giá cảm xúc của người dùng đối với các khía cạnh cụ thể của sản phẩm hoặc dịch vụ được đề cập trong văn bản, chẳng hạn như "giá cả", "chất lượng dịch vụ" hay "hỗ trợ khách hàng". Khác với các phương pháp phân tích cảm xúc truyền thống chỉ đánh giá cảm xúc tổng thể của một câu hoặc đoạn văn, ABSA cung cấp cái nhìn chi tiết hơn bằng cách liên kết cảm xúc (tích cực, tiêu cực, trung tính) với từng khía cạnh cụ thể. Cách tiếp cận này mang lại giá trị lớn trong các ứng dụng như phân tích đánh giá khách hàng, hệ thống tư vấn thông minh và tối ưu hóa trải nghiệm người dùng.

Hiện nay, các phương pháp giải quyết bài toán ABSA bao gồm:

- Phương pháp dựa trên luật (rule-based): Sử dụng từ điển cảm xúc và quy tắc ngôn ngữ để nhận diện khía cạnh và cảm xúc, nhưng thường thiếu linh hoạt với dữ liệu phức tạp.

- Học máy truyền thống: Áp dụng các thuật toán như Naive Bayes hoặc SVM, hiệu quả trong một số trường hợp nhưng hạn chế trong việc nắm bắt ngữ cảnh sâu.

- Học sâu (deep learning): Sử dụng các mô hình tiên tiến như LSTM, BERT hoặc cơ chế Attention để khai thác ngữ nghĩa và ngữ cảnh văn bản một cách hiệu quả hơn.

Trong đề án này, chúng tôi đề xuất một hướng tiếp cận mới bằng cách tích hợp học tăng cường (Reinforcement Learning – RL) vào ABSA. Học tăng cường cho phép mô hình tự động tối ưu hóa quá trình nhận diện khía cạnh và phân loại cảm xúc thông qua tương tác với môi trường dữ liệu, dựa trên phần thưởng từ độ chính xác hoặc phản hồi người dùng. Phương pháp này hứa hẹn cải thiện độ chính xác và khả năng thích nghi so với các kỹ thuật truyền thống, đặc biệt trong các ngữ cảnh văn bản đa dạng và phức tạp.

1.2. Phương pháp giải quyết bài toán

Phương pháp dựa trên quy tắc (rule-based) là một trong những hướng tiếp cận truyền thống trong phân tích cảm xúc theo khía cạnh. Phương pháp này chủ yếu dựa vào tập hợp các quy tắc ngữ pháp và từ điển cảm xúc – nơi lưu trữ các từ hoặc cụm từ được gán nhãn mức độ cảm xúc tích cực hay tiêu cực, kèm theo cường độ thể hiện. Ví dụ, từ "tuyệt vời" thường thể

hiện cảm xúc tích cực mạnh, trong khi "khá tốt" chỉ mang cảm xúc tích cực nhẹ. Các quy tắc ngôn ngữ trong phương pháp này giúp xác định mối liên hệ giữa các danh từ (thường là khía cạnh) và các tính từ hoặc trạng từ (thường là biểu hiện cảm xúc), từ đó phân tích được thái độ của người viết đối với từng khía cạnh được đề cập. Phương pháp này đặc biệt phù hợp trong những bài toán đơn giản hoặc khi dữ liệu huấn luyện bị hạn chế, vì không cần đến quá trình học phức tạp như trong các mô hình học máy. Tuy nhiên, nhược điểm lớn của nó là thiếu tính linh hoạt khi xử lý các ngữ cảnh đa dạng hoặc cấu trúc câu phức tạp, nơi mà sắc thái cảm xúc có thể thay đổi tùy vào cách diễn đạt. Ngoài ra, việc mở rộng hoặc cập nhật hệ thống khi dữ liệu thay đổi thường đòi hỏi phải sửa đổi hoặc xây dựng lại tập quy tắc, gây tốn kém thời gian và công sức trong quá trình bảo trì. Vì vậy, rule-based thường chỉ được áp dụng hiệu quả trong các hệ thống nhỏ, dữ liệu ổn định và ngôn ngữ không quá phức tạp.

Các phương pháp dựa trên học máy truyền thống (Machine Learning-based methods) áp dụng các thuật toán như Naive Bayes, SVM (Support Vector Machine), hoặc Logistic Regression để thực hiện phân loại cảm xúc. Quy trình xử lý thường bao gồm ba giai đoạn chính: tiền xử lý văn bản, trích xuất đặc trưng và huấn luyện mô hình. Ở bước đầu tiên, dữ liệu văn bản sẽ được làm sạch – loại bỏ các ký tự dư thừa, chuẩn hóa từ ngữ, chuyển đổi chữ viết, và xử lý ngôn ngữ để chuẩn bị cho quá trình phân tích. Tiếp đó, đặc trưng của văn bản được trích xuất, thường thông qua các kỹ thuật như n-gram, TF-IDF hoặc tích hợp đặc trưng cảm xúc từ các từ điển chuyên dụng. Những đặc trưng này đóng vai trò quan trọng trong việc giúp mô hình học máy phát hiện các mẫu xuất hiện trong dữ liệu văn bản để từ đó đưa ra dự đoán cảm xúc tương ứng. Sau khi trích xuất đặc trưng, mô hình sẽ được huấn luyện trên tập dữ liệu đã gán nhãn để học cách phân biệt cảm xúc theo từng khía cạnh cụ thể. So với phương pháp dựa trên quy tắc, học máy truyền thống thường có khả năng tổng quát tốt hơn và xử lý được nhiều tình huống phức tạp hơn. Tuy vậy, hiệu quả của phương pháp này phụ thuộc chặt chẽ vào chất lượng dữ liệu đầu vào và quá trình trích xuất đặc trưng. Nếu đặc trưng không đủ biểu đạt hoặc dữ liệu gán nhãn kém chính xác, hiệu suất phân loại có thể bị ảnh hưởng đáng kể.

Phương pháp học sâu (Deep Learning-based methods) ứng dụng các mô hình mạng nơ-ron sâu như LSTM (Long Short-Term Memory), BiLSTM (Bidirectional LSTM) và Transformer để tự động học và trích xuất các đặc trưng ngữ nghĩa phức tạp từ văn bản. Ưu điểm nổi bật của các mô hình này là khả năng nắm bắt ngữ cảnh và mối liên kết giữa các từ

trong câu, từ đó cải thiện độ chính xác trong việc phân tích cảm xúc gắn với từng khía cạnh. Trong đó, LSTM và BiLSTM tỏ ra hiệu quả với các dạng dữ liệu tuần tự, bởi chúng có thể ghi nhớ và duy trì thông tin quan trọng trong chuỗi thời gian. BiLSTM đặc biệt nổi bật nhờ khả năng xử lý thông tin theo cả hai chiều, giúp mô hình hiểu rõ hơn các cấu trúc câu phức tạp. Mặt khác, Transformer – với cơ chế hoạt động không tuần tự và cấu trúc tự chú ý (attention) – cho phép mô hình học được mối quan hệ giữa các từ dù cách nhau xa trong câu, mang lại khả năng hiểu ngữ nghĩa vượt trội. Đặc biệt, việc tích hợp các mô hình ngôn ngữ tiên tiến như BERT, RoBERTa hoặc multilingual BERT đã mở rộng khả năng áp dụng của học sâu, giúp mô hình thích ứng với nhiều loại ngữ cảnh và dữ liệu khác nhau. Những mô hình này tận dụng kỹ thuật attention và multi-head attention để tập trung vào những từ hoặc cụm từ quan trọng, từ đó nâng cao hiệu quả phân tích cảm xúc theo khía cạnh. Tuy nhiên, chi phí tính toán cao và yêu cầu về một lượng dữ liệu lớn để huấn luyện vẫn là những rào cản khi triển khai học sâu trong môi trường có tài nguyên hạn chế.

Phương pháp học tăng cường (Reinforcement Learning - RL) tiếp cận bài toán phân tích cảm xúc theo chủ đề bằng cách mô hình hóa quá trình dự đoán như một chuỗi hành động trong môi trường tương tác. Thay vì học từ dữ liệu gán nhãn theo cách truyền thống, RL cho phép mô hình học thông qua phản hồi (feedback) từ môi trường, thông qua cơ chế phần thưởng (reward). Trong ngữ cảnh phân tích cảm xúc, tác nhân (agent) có thể được huấn luyện để lựa chọn hành động phù hợp như xác định khía cạnh hoặc cảm xúc trong một câu, nhằm tối đa hóa tổng phần thưởng nhận được theo thời gian. Quá trình huấn luyện thường dựa trên các kỹ thuật như Q-learning hoặc Deep Q-Network (DQN), nơi mô hình học cách đưa ra các hành động tối ưu dựa trên trạng thái hiện tại – ví dụ như biểu diễn của câu, đặc trưng ngữ nghĩa, hoặc vị trí của từ trong văn bản. Việc kết hợp học tăng cường với các mô hình ngôn ngữ như BERT giúp tăng khả năng hiểu ngữ cảnh, đồng thời cho phép mô hình thích ứng tốt hơn với các tình huống chưa từng gặp trong dữ liệu huấn luyện. So với các phương pháp học sâu truyền thống vốn học từ dữ liệu tĩnh, RL có ưu điểm trong việc học liên tục và cải thiện hiệu suất thông qua trải nghiệm. Ngoài ra, khả năng phản hồi theo chuỗi hành động cũng giúp mô hình RL xử lý tốt hơn các câu văn dài hoặc chứa nhiều khía cạnh cảm xúc đan xen. Tuy nhiên, học tăng cường cũng gặp không ít thách thức, như yêu cầu thiết kế phần thưởng phù hợp, thời gian huấn luyện dài, và khó khăn trong việc đảm bảo ổn định khi áp dụng vào dữ liệu ngôn ngữ tự nhiên. Dẫu vậy, với tiềm năng thích ứng cao và khả năng học chính sách tối ưu trong

môi trường phức tạp, học tăng cường đang trở thành hướng tiếp cận mới đầy triển vọng trong bài toán phân tích cảm xúc theo chủ đề.

1.3. Thách thức hạn chế gặp phả

- Đặc thù ngôn ngữ tiếng Việt
- Đặc điểm phân tích ý kiến theo khía cạnh
- Thách thức kỹ thuật
- Xử lý các trường hợp phức tạp

1.4. Mục tiêu

Mục tiêu chính của đề án là xây dựng một mô hình có khả năng phân tích cảm xúc theo chủ đề cho dữ liệu văn bản tiếng Việt, giúp nhận diện và phân loại chính xác các cảm xúc của người dùng đối với các chủ đề khác nhau của sản phẩm, dịch vụ hoặc sự kiện

CHƯƠNG 2: CƠ SỞ KIẾN THỨC

2.1. Tổng quan về học tăng cường (Reinfrcement Learning)

2.1.1 Giới thiệu

Học tăng cường (Reinforcement Learning – RL) là một nhánh của học máy, tập trung vào cách một tác nhân (agent) học cách tương tác với môi trường nhằm đạt được phần thưởng tối ưu thông qua quá trình thử - sai (trial-and-error). Khác với học có giám sát (supervised learning), nơi các cặp đầu vào - đầu ra được cung cấp rõ ràng, hay học không giám sát (unsupervised learning) tìm kiếm cấu trúc tiềm ẩn trong dữ liệu, RL không yêu cầu nhãn trực tiếp cho mỗi hành động. Thay vào đó, tác nhân học thông qua phản hồi dưới dạng phần thưởng (reward) hoặc trừng phạt (penalty) từ môi trường sau mỗi hành động, từ đó dần cải thiện chính sách hành động (policy) để tối đa hóa tổng phần thưởng tích lũy theo thời gian. Điểm đặc trưng của RL là sự cân bằng giữa khám phá (exploration) các hành động mới để tìm ra chiến lược tốt hơn và khai thác (exploitation) những hành động đã biết là mang lại phần thưởng cao. Khái niệm RL có nguồn gốc sâu xa từ tâm lý học hành vi, đặc biệt là lý thuyết về điều kiện hóa qua thử và sai của động vật (Sutton & Barto, 2018). Trên thực tế, quá trình huấn luyện một tác nhân RL giống như cách một sinh vật học từ tương tác với môi trường xung quanh: thử một hành động, quan sát kết quả, ghi nhớ phản hồi và điều chỉnh hành vi trong tương lai. Do đó, RL đặc biệt phù hợp với các bài toán cần ra quyết định liên tục trong một chuỗi các trạng thái phức tạp, nơi mỗi quyết định không chỉ ảnh hưởng đến phần thưởng tức thời mà còn tác động đến các trạng thái và phần thưởng trong tương lai. Các ví dụ điển hình bao gồm robot học cách di chuyển, game AI tự học chơi cờ, hay các hệ thống đề xuất thích ứng theo phản hồi của người dùng.

Trong những năm gần đây, học tăng cường đã ghi nhận những bước tiến vượt bậc nhờ vào sự kết hợp với học sâu (Deep Learning), hình thành nên học tăng cường sâu (Deep Reinforcement Learning – DRL). Những mô hình như Deep Q-Network (DQN) và Proximal Policy Optimization (PPO) đã mang lại thành công ấn tượng trong các môi trường phức tạp có không gian trạng thái và hành động rất lớn, nơi các phương pháp RL truyền thống gặp khó khăn. Những thành tựu đó đã khẳng định vai trò ngày càng quan trọng của RL trong trí tuệ nhân tạo hiện đại, mở ra tiềm năng ứng dụng trong nhiều lĩnh vực, bao gồm cả Phân tích Cảm xúc theo Chủ đề (Aspect-Based Sentiment Analysis – ABSA). Trong ABSA, việc xác định

khía cạnh và sau đó là phân loại cảm xúc cho khía cạnh đó có thể được xem như một quá trình ra quyết định tuần tự, phụ thuộc vào ngữ cảnh, và có thể được tối ưu hóa thông qua quá trình học từ tương tác với dữ liệu văn bản

2.1.2 Các thành phần chính

- Tác nhân (Agent)
- Môi trường (Environment)
- Trạng thái (State)
- Hành động (Action)
- Phần thưởng (Reward)
- Chính sách (Policy)
- Hàm giá trị (Value function)
- Hàm Q (Q-function)

2.1.3 Mô hình toán học

Học tăng cường thường được mô hình hóa thông qua quá trình quyết định Markov (Markov Decision Process - MDP), bao gồm năm thành phần chính:

- S: Tập hợp các trạng thái (states), ví dụ: các đặc trưng biểu diễn câu văn, từ ngữ, hoặc trạng thái ngữ cảnh hiện tại trong văn bản.
- A: Tập hợp các hành động (actions), chẳng hạn như: gán nhãn cảm xúc (tích cực, tiêu cực, trung tính) hoặc xác định chủ đề của một từ/cụm từ.
- P (s′|s, a): Hàm xác suất chuyển trạng thái, thể hiện xác suất chuyển từ trạng thái hiện tại s sang trạng thái mới s′ khi thực hiện hành động a.
- R (s, a, s′): Hàm phần thưởng (reward function), ví dụ: nhận +1 nếu gán nhãn cảm xúc đúng, -1 nếu sai hoặc không phù hợp với ngữ cảnh.
- $\gamma \in [0,1]$: Hệ số chiết khấu (discount factor), quyết định mức độ quan trọng của phần thưởng tương lai so với phần thưởng hiện tại.

2.1.4 Phân loại thuật toán RL trong ABSA

Bảng 2.1 Bảng so sánh các phương pháp học tăng cường

Nhóm thuật toán	**Ưu điểm**	**Hạn chế**	**Ứng dụng trong ABSA**
Value-Based (Q-learning, DQN)	Dễ triển khai, phù hợp không gian hành động rời rạc	Không phù hợp với văn bản dài, trạng thái phức tạp	Gán nhãn cảm xúc đơn giản
Policy-Based (REINFORCE, PPO)	Hỗ trợ hành động liên tục, dễ tùy biến chính sách	Cần dữ liệu lớn, dễ dao động trong huấn luyện	Nhận diện chủ đề phức tạp, định hướng đa chiều
Actor-Critic (A3C, SAC)	Kết hợp Value & Policy-Based, hiệu quả cao	Triển khai phức tạp, khó điều chỉnh	Hệ thống ABSA end-to-end đa nhiệm

2.2 Phân loại ý kiến và nhận diện khía cạnh

Trong hệ thống Phân tích cảm xúc theo chủ đề (Aspect-Based Sentiment Analysis - ABSA), hai thành phần trung tâm là nhận diện khía cạnh (Aspect Extraction) và phân loại ý kiến (Sentiment Classification). Hai bước này thường được xây dựng nối tiếp hoặc học đồng thời trong một hệ thống, nhằm tách riêng các khía cạnh được đề cập trong câu văn và gán nhãn cảm xúc cho từng khía cạnh đó. Tuy nhiên, trong thực tế triển khai, việc xây dựng một mô hình hiệu quả để đảm nhiệm hai nhiệm vụ này một cách tối ưu không hề dễ dàng, đặc biệt khi phải xử lý các văn bản tự nhiên có ngữ nghĩa phức tạp, mơ hồ, hoặc nhiều tầng nghĩa.

Việc ứng dụng Học tăng cường (Reinforcement Learning - RL) mang lại một hướng tiếp cận mới, nơi hệ thống có thể tự động học chiến lược tốt nhất thông qua tương tác với dữ liệu và phần thưởng phản hồi, giúp mô hình thích ứng tốt hơn với sự đa dạng và biến đổi trong ngôn ngữ tự nhiên.

- Nhận diện khía cạnh (Aspect Extraction)
- Phân loại ý kiến (Sentiment Classification)
- Hướng tiếp cận phối hợp hai nhiệm vụ với RL

2.3 Ứng dụng học tăng cường

- Trí tuệ nhân tạo trong trò chơi (AI Gaming)
- Điều khiển Robot (Robotics Control)
- Xe tự hành (Autonomous Vehicles)
- Quản lý tài chính và giao dịch tự động (Financial Trading)
- Tối ưu hóa mạng lưới và tài nguyên (Network Optimization)
- Hệ thống đề xuất (Recommendation Systems)

CHƯƠNG 3: THỰC NGHIỆM, ĐÁNH GIÁ KẾT QUẢ

3.1. Mô tả bộ dữ liệu

Trong đề án này, hai bộ dữ liệu tiếng Việt được sử dụng để huấn luyện và đánh giá mô hình học tăng cường trong bài toán phân tích cảm xúc theo chủ đề (Aspect-Based Sentiment Analysis – ABSA), bao gồm: UIT-VSFC và UIT-ViFSD. Việc sử dụng kết hợp cả dữ liệu thực tế và dữ liệu có cấu trúc sẵn giúp mô hình học được đa dạng các biểu hiện cảm xúc cũng như khía cạnh trong phản hồi của người dùng.

Bộ dữ liệu UIT-VSFC (Vietnamese Students' Feedback Corpus)

UIT-VSFC (UIT Vietnamese Students' Feedback Corpus) là một tập dữ liệu tiếng Việt được xây dựng bởi nhóm nghiên cứu UIT NLP thuộc Trường Đại học Công nghệ Thông tin – ĐHQG TP.HCM. Mục tiêu của bộ dữ liệu là phục vụ cho các nghiên cứu về phân tích cảm xúc và nhận diện chủ đề trong các phản hồi của sinh viên về môi trường học tập.

Bộ dữ liệu UIT-ViFSD (UIT Vietnamese Feedback Sentiment Dataset)

UIT-ViFSD (UIT Vietnamese Feedback Sentiment Dataset) là một tập dữ liệu tiếng Việt được xây dựng bởi nhóm nghiên cứu UIT NLP tại Trường Đại học Công nghệ Thông tin – ĐHQG TP.HCM. Bộ dữ liệu này được phát triển với mục tiêu phục vụ cho các nghiên cứu về phân tích cảm xúc theo chủ đề (Aspect-Based Sentiment Analysis – ABSA) trong lĩnh vực đánh giá sản phẩm, đặc biệt là các thiết bị công nghệ như điện thoại di động.

3.2 Xây dựng mô hình

3.2.1 Mô hình RL_DQN (Deep Q-Learning Network)

Bước 1: Cài đặt và chuẩn bị môi trường

Bước 2: Chuẩn bị và xử lý dữ liệu đầu vào

Bước 3: Trích xuất biểu diễn ngữ nghĩa với PhoBERT

Bước 4: Thiết kế kiến trúc mạng DQN và Replay Buffer

Bước 5: Tham số huấn luyện (Hyperparameters)

Bước 6: Quy trình huấn luyện thuật toán DQN

Bước 7: Đánh giá và đo lường hiệu năng

Bước 8: Kết quả thực nghiệm

Bước 9: Dự đoán kết quả

3.2.1 Mô hình RL_PPO (Proximal Policy Optimization)

Bước 1: Cài đặt và chuẩn bị môi trường

Bước 2: Tiền xử lý dữ liệu

Bước 3: Mã hóa văn bản

Bước 4: Mô hình PPO Agent

Bước 5: Tham số huấn luyện (Hyperparameters)

Bước 6: Đánh giá mô hình

Bước 7: Kết quả huấn luyện và đánh giá

Bước 8: Dự đoán kết quả

3.2.1 Mô hình RL_A3C (Asynchronous Advantage Actor-Critic)

Bước 1: Cài đặt và chuẩn bị môi trường

Bước 2: Tiền xử lý dữ liệu

Bước 3: Môi trường tác nghiệp cho bài toán ABSA

Bước 4: Mô hình học tăng cường A3C

Bước 5: Quá trình huấn luyện

Bước 6: Tham số huấn luyện

Bước 7: Đánh giá hiệu năng mô hình

Bước 8: Dự đoán kết quả

3.3. Đánh giá kết quả

Kết quả trên Bộ dữ liệu UIT-VSFC

Bảng 3.1 Kết quả phân loại Sentiment trên UIT-VSFC

Mô hình	Accuracy (%)	Precision	Recall	F1-score	Loss	Time
LSTM	89.10	0.7653	0.7106	0.7284	0.9031	14.09s
Multi BERT	89.26	0.7996	0.6910	0.7130	0.6953	384.42s
RL_DQN	90.15	0.8951	0.8961	0.9050	0.1124	102.53s
Double_RL_DQN	87.87	0.7453	0.7001	0.7156	0.356	411.83s
RL_PPO	85.52	0.841	0.837	0.839	0.462	302.56s
RL_A3C	78.23	0.7814	0.7751	0.7782	0.503	230.51s

Bảng 3.2 Kết quả phân loại Topic trên UIT-VSFC

Mô hình	Accuracy (%)	Precision	Recall	F1-score	Loss	Time
LSTM	86.13	0.7471	0.7504	0.7483	0.9031	14.09s
Multi BERT	87.37	0.8029	0.7174	0.7463	0.6953	384.42s
RL_DQN	85.36	0.8448	0.8453	0.8536	0.1586	102.52s
Double_RL_DQN	81.74	0.6892	0.6852	0.6850	0.378	411.35s
RL_PPO	78.75	0.796	0.784	0.790	0.551	303.78s
RL_A3C	69.45	0.6932	0.6890	0.6911	0.654	180.32s

Kết quả trên bộ dữ liệu UIT-ViFSD

Bảng 3.3 Kết quả phân loại Sentiment trên UIT-ViFSD

Mô hình	Accuracy (%)	Precision	Recall	F1-score	Loss	Time
LSTM	74.2	0.746	0.735	0.740	0.561	496
Multi BERT	83.8	0.828	0.843	0.835	0.346	1238
RL_DQN	85.5	0.849	0.859	0.854	0.292	2143
Double_RL_DQN	86.4	0.856	0.867	0.861	0.275	2305
RL_PPO	87.6	0.871	0.879	0.870	0.239	2724
RL_A3C	86.8	0.864	0.867	0.863	0.251	2503

Bảng 3.4 Kết quả phân loại Topic trên UIT-ViFSD

Mô hình	Accuracy (%)	Precision	Recall	F1-score	Loss	Time
LSTM	69.1	0.694	0.689	0.692	0.614	502
Multi BERT	80.2	0.797	0.809	0.803	0.374	1215
RL_DQN	82.5	0.815	0.826	0.820	0.317	2095
Double_RL_DQN	83.6	0.829	0.840	0.834	0.288	2270
RL_PPO	84.7	0.843	0.850	0.846	0.263	2701
RL_A3C	84.0	0.834	0.842	0.838	0.275	2492

Hạn chế của Quá trình Thực nghiệm

Tối ưu hóa Siêu tham số (Hyperparameter Tuning): Việc tinh chỉnh siêu tham số cho các thuật toán học tăng cường vốn là một thách thức, đòi hỏi nhiều thử nghiệm và tài nguyên tính toán đáng kể. Mặc dù đã nỗ lực tối ưu hóa trong phạm vi cho phép, vẫn có khả năng rằng hiệu suất của một số mô hình (đặc biệt là A3C) có thể được cải thiện hơn nữa nếu có thêm thời gian và tài nguyên để thực hiện một quá trình tìm kiếm siêu tham số toàn diện hơn (ví dụ: sử dụng Grid Search, Random Search hoặc các phương pháp tối ưu hóa Bayesian).

Giới hạn về Kích thước và Đa dạng của Bộ dữ liệu: Mặc dù UIT-VSFC và UIT-ViFSD là những bộ dữ liệu có giá trị cho cộng đồng nghiên cứu tiếng Việt, việc huấn luyện và đánh giá trên các bộ dữ liệu lớn hơn, đa dạng hơn về lĩnh vực và phong cách ngôn ngữ sẽ giúp kiểm chứng sâu hơn khả năng tổng quát hóa của các mô hình được đề xuất.

Kiến trúc Mạng Nơ-ron trong Tác nhân RL: Các kiến trúc mạng nơ-ron được sử dụng để biểu diễn tác nhân (agent) trong các mô hình RL (ví dụ: số lớp ẩn, số lượng neuron) được lựa chọn dựa trên kinh nghiệm và các thử nghiệm sơ bộ. Việc khám phá các kiến trúc mạng phức tạp hơn, hoặc các kiến trúc được thiết kế chuyên biệt cho xử lý ngôn ngữ tự nhiên (ví dụ, tích hợp sâu hơn các cơ chế attention), có thể mang lại những cải thiện về hiệu suất.

Kết luận Rút ra từ Kết quả Thực nghiệm

Từ những phân tích và đánh giá chi tiết đã trình bày, nghiên cứu khẳng định một cách thuyết phục rằng các phương pháp học tăng cường, đặc biệt là RL_DQN và RL_PPO, sở hữu tiềm năng và hiệu quả vượt trội so với các mô hình học sâu truyền thống trong việc giải quyết bài toán phân tích cảm xúc theo chủ đề trên dữ liệu tiếng Việt.

- RL_DQN đã chứng minh sự phù hợp và hiệu quả cao trên dữ liệu có cấu trúc tương đối đơn giản và rõ ràng (như UIT-VSFC), mang lại hiệu suất ấn tượng với chi phí thời gian huấn luyện hợp lý.
- RL_PPO đã thể hiện khả năng vượt trội trên dữ liệu phức tạp và đa dạng hơn (như UIT-ViFSD), cho thấy tiềm năng lớn trong việc xử lý các tình huống thực tế với nhiều sắc thái ngôn ngữ. Mặc dù đòi hỏi thời gian huấn luyện dài hơn, sự cải thiện về độ chính xác và tính ổn định của PPO là rất đáng giá theo kết quả thu được.

- Các biến thể khác như Double_RL_DQN và RL_A3C cũng cho thấy những kết quả hứa hẹn trên dữ liệu phức tạp, tuy nhiên, có thể cần thêm các nghiên cứu chuyên sâu để có thể tối ưu hóa hoàn toàn và khai thác hết tiềm năng của chúng.

KẾT LUẬN VÀ ĐỀ XUẤT

Đề án tốt nghiệp thạc sĩ với tiêu đề "Ứng dụng học tăng cường trong phân tích cảm xúc theo chủ đề" đã được thực hiện với mục tiêu khám phá và chứng minh hiệu quả của việc áp dụng các thuật toán học tăng cường (Reinforcement Learning - RL) vào một trong những bài toán quan trọng và thách thức của Xử lý Ngôn ngữ Tự nhiên (NLP) là Phân tích Cảm xúc theo Chủ đề (Aspect-Based Sentiment Analysis - ABSA), đặc biệt trên dữ liệu tiếng Việt. Qua quá trình nghiên cứu lý thuyết và triển khai thực nghiệm, đề án đã đạt được những kết quả đáng kể và rút ra được những nhận định quan trọng, đồng thời xác định các hướng phát triển tiềm năng trong tương lai.

Kết quả đạt được

- Nghiên cứu và làm chủ cơ sở lý thuyết: Đề án đã tiến hành nghiên cứu sâu rộng về các khái niệm, mô hình toán học và các thuật toán cốt lõi của học tăng cường (bao gồm Q-Learning, DQN, PPO, A3C) cũng như các phương pháp truyền thống và hiện đại trong ABSA. Điều này đã tạo nền tảng kiến thức vững chắc cho việc thiết kế và triển khai các mô hình thực nghiệm.
- Xây dựng và triển khai thành công các mô hình học tăng cường cho ABSA: Các mô hình RL_DQN, RL_PPO, và RL_A3C đã được xây dựng và huấn luyện thành công trên hai bộ dữ liệu tiếng Việt đặc thù là UIT-VSFC và UIT-ViFSD. Quá trình này bao gồm các bước từ tiền xử lý dữ liệu, trích xuất đặc trưng ngữ nghĩa bằng PhoBERT, đến thiết kế kiến trúc agent và quy trình huấn luyện tối ưu.
- Chứng minh hiệu quả vượt trội của học tăng cường: Kết quả thực nghiệm đã cho thấy các mô hình học tăng cường, đặc biệt là RL_DQN trên dữ liệu có cấu trúc rõ ràng (UIT-VSFC) và RL_PPO trên dữ liệu phức tạp hơn (UIT-ViFSD), đạt được hiệu suất vượt trội so với các mô hình học sâu truyền thống như LSTM và Multi-BERT trên cả hai nhiệm vụ phân loại cảm xúc và xác định chủ đề. Cụ thể, mô hình RL_DQN đạt F1-score cao nhất (0.9050 cho Sentiment và 0.8536 cho Topic) trên UIT-VSFC, trong khi RL_PPO dẫn đầu trên UIT-ViFSD (F1-score 0.870 cho Sentiment và 0.846 cho Topic).
- Phân tích và đánh giá sâu sắc kết quả: Đề án đã tiến hành phân tích chi tiết các yếu tố ảnh hưởng đến hiệu suất của từng mô hình, bao gồm đặc điểm của thuật toán RL, tính

chất của bộ dữ liệu, và sự cân bằng giữa hiệu quả và chi phí tính toán. Những phân tích này cung cấp cái nhìn sâu sắc về ưu nhược điểm của từng phương pháp và điều kiện ứng dụng phù hợp.

- Đáp ứng mục tiêu đề tài: Đề án đã hoàn thành mục tiêu chính là phát triển và ứng dụng thành công phương pháp học tăng cường để xây dựng một hệ thống phân tích cảm xúc theo chủ đề, có khả năng tự động phân loại cảm xúc và nhận diện các khía cạnh liên quan với độ chính xác cao, đặc biệt là trong việc xử lý dữ liệu tiếng Việt với những đặc thù riêng.

Hạn chế của đề tài

Bên cạnh những kết quả tích cực, trong quá trình thực hiện, đề án cũng nhận thấy một số hạn chế cần được cải thiện trong các nghiên cứu tiếp theo:

- Tối ưu hóa siêu tham số: Quá trình tinh chỉnh siêu tham số cho các thuật toán RL, đặc biệt là A3C, vẫn còn không gian để cải thiện nhằm khai thác tối đa tiềm năng của thuật toán.
- Mở rộng bộ dữ liệu: Việc thử nghiệm trên các bộ dữ liệu lớn hơn, đa dạng hơn về lĩnh vực và phong cách ngôn ngữ sẽ giúp đánh giá khả năng tổng quát hóa của mô hình một cách toàn diện hơn.
- Khám phá kiến trúc mạng phức tạp hơn: Kiến trúc mạng nơ-ron cho các agent RL có thể được khám phá sâu hơn, ví dụ như tích hợp các cơ chế attention phức tạp hơn hoặc các kỹ thuật mới trong biểu diễn ngôn ngữ.
- Xử lý các trường hợp đặc biệt: Khả năng xử lý các trường hợp ngôn ngữ phức tạp như sarcasm, ẩn dụ hoặc các biểu hiện cảm xúc tinh vi vẫn cần được đầu tư nghiên cứu thêm.

Hướng nghiên cứu và phát triển trong tương lai

Từ những kết quả và hạn chế của đề tài, đề xuất một số hướng nghiên cứu và phát triển tiềm năng trong tương lai:

- Nâng cao độ chính xác và khả năng tổng quát hóa: Áp dụng các kỹ thuật tinh chỉnh siêu tham số tự động (Automated Hyperparameter Optimization). Nghiên cứu các phương pháp transfer learning và domain adaptation tiên tiến hơn để mô hình có thể dễ dàng thích ứng với các lĩnh vực và bộ dữ liệu mới mà không cần huấn luyện lại từ

đầu. Tích hợp các mô hình ngôn ngữ lớn (Large Language Models - LLMs) vào kiến trúc của agent RL để tận dụng kiến thức nền phong phú của chúng.

- Phát triển hệ thống học liên tục (Continual Learning / Lifelong Learning): Xây dựng khả năng cho mô hình tự động cập nhật và thích ứng với các xu hướng ngôn ngữ mới, các khía cạnh sản phẩm mới hoặc sự thay đổi trong cách biểu đạt cảm xúc của người dùng theo thời gian mà không bị "quên" kiến thức cũ (catastrophic forgetting).
- Xử lý đa ngôn ngữ và các ngôn ngữ ít tài nguyên: Mở rộng mô hình để có thể phân tích cảm xúc theo chủ đề cho nhiều ngôn ngữ khác nhau, đặc biệt là các ngôn ngữ ít có tài nguyên huấn luyện.
- Tăng cường khả năng giải thích (Explainability / Interpretability): Phát triển các cơ chế giúp hiểu rõ hơn tại sao mô hình đưa ra một quyết định phân loại cụ thể, tăng cường sự tin cậy và khả năng ứng dụng trong các hệ thống quan trọng.
- Ứng dụng vào các bài toán thực tế và tích hợp hệ thống: Xây dựng các giao diện ứng dụng (API) để dễ dàng tích hợp mô hình vào các hệ thống quản lý quan hệ khách hàng (CRM), chatbot, nền tảng thương mại điện tử, hoặc các công cụ phân tích thị trường. Phát triển các ứng dụng cụ thể như hệ thống tự động tóm tắt đánh giá sản phẩm theo khía cạnh và cảm xúc, hoặc hệ thống cảnh báo sớm các vấn đề tiêu cực từ phản hồi của khách hàng. Nghiên cứu việc áp dụng RL trong việc cá nhân hóa phản hồi hoặc đề xuất dựa trên phân tích cảm xúc chi tiết.