

ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHOA TOÁN - TIN HỌC

Hoàng Hải Đăng

Trần Tiến Đạt

PHÂN TÍCH CẢM XÚC KHÁCH HÀNG
TRONG NGÀNH HÀNG SÁCH

KHOÁ LUẬN TỐT NGHIỆP CỬ NHÂN

Tp. Hồ Chí Minh - 2024

ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHOA TOÁN - TIN HỌC

Hoàng Hải Đăng

Trần Tiến Đạt

PHÂN TÍCH CẢM XÚC KHÁCH HÀNG
TRONG NGÀNH HÀNG SÁCH

KHOÁ LUẬN TỐT NGHIỆP CỬ NHÂN
CHƯƠNG TRÌNH CHÍNH QUY

NGÀNH: KHOA HỌC DỮ LIỆU

Mã số sinh viên: 20280011 - 20280016

NGƯỜI HƯỚNG DẪN:

ThS. Đoàn Thị Trâm

Tp. Hồ Chí Minh - 2024

Lời cảm ơn

Lời đầu tiên, chúng tôi xin chân thành cảm ơn trường **Đại học Khoa học Tự nhiên, Đại học Quốc Gia TP. Hồ Chí Minh** và **Ban chủ nhiệm Khoa Toán - Tin học** đã tạo điều kiện cho chúng tôi hoàn thành chương trình học nói chung và khoá luận tốt nghiệp nói riêng. Nền tảng kiến thức được tích lũy qua 4 năm học cùng với kinh nghiệm tích lũy qua khoảng thời gian thực hiện khoá luận đã tạo tiền đề rất tốt cho con đường nghiên cứu khoa học về sau của chúng tôi.

Lời thứ hai, chúng tôi bày tỏ lòng biết ơn đặc biệt đến **ThS. Đoàn Thị Trâm**, người đã trực tiếp định hướng, hướng dẫn cho chúng tôi trong suốt quá trình thực hiện khoá luận. Từ giai đoạn chọn đề tài, nghiên cứu tài liệu, tìm kiếm phương pháp, thực nghiệm, kết luận đều có thầy hỗ trợ về mặt kiến thức và kinh nghiệm. Sự động viên của thầy cũng là nguồn cảm hứng rất lớn cho chúng tôi trong khi thực hiện đề tài và trên con đường nghiên cứu khoa học.

Chúng tôi cảm thấy vô cùng may mắn khi nhận được sự dạy dỗ, giúp đỡ tận tình của các thầy, cô giáo ở khoa Toán - Tin học. Các thầy cô không chỉ trao cho chúng tôi kiến thức, kỹ năng, kinh nghiệm mà còn là phong cách sống và làm việc chuyên nghiệp, trách nhiệm. Xin dành lời cảm ơn thứ ba cho các quý thầy cô, kính chúc quý thầy cô luôn khoẻ mạnh, thành công trong sự nghiệp giảng dạy.

Lời cuối cùng, chúng tôi xin gửi lời cảm ơn đến gia đình, người thân, bạn bè đã khuyến khích và tạo điều kiện cho chúng tôi hoàn thành khoá luận tốt nghiệp này.

Trân trọng.

Tp. HCM, ngày 27 tháng 07 năm 2023

Tác giả

Hoàng Hải Đăng
Trần Tiến Đạt

Mục lục

Lời nói đầu	3
Danh sách các hình	5
Danh sách các bảng	6
Danh sách các từ viết tắt	7
Danh sách thuật ngữ	8
Chương 1. Giới thiệu	9
1.1 Lý do chọn đề tài	9
1.2 Trình bày bài toán	10
Chương 2. Tổng quan lý thuyết	12
2.1 Sentiment Analysis (SA)	12
2.2 Aspect-based Sentiment Analysis (ABSA)	13
2.3 Các nghiên cứu có liên quan	15
2.4 Word embedding	16
2.5 TF-IDF	16
2.6 Recurrent Neural Network (RNN)	17
2.6.1 Cấu trúc	17
2.6.2 Hạn chế	18
2.6.3 Long Short-Term Memory (LSTM)	18
2.6.4 Gated Recurrent Unit (GRU)	22
2.7 PhoBERT	25
2.8 ELMo	26
Chương 3. Phương pháp sử dụng	27

3.1 Thu thập và tiền xử lý dữ liệu	27
3.1.1 Thu thập dữ liệu	27
3.1.2 Tiền xử lý dữ liệu	27
3.1.3 Đánh nhãn dữ liệu	29
3.2 Mô hình ELMo	31
3.3 Mô hình PhoBERT	32
3.3.1 Lớp đầu vào và các lớp ẩn	33
3.3.2 Lớp đầu ra	36
Chương 4. Thực nghiệm và đánh giá mô hình	38
4.1 Bộ dữ liệu thực nghiệm	38
4.1.1 Chuẩn bị dữ liệu	38
4.1.2 Thống kê tổng quan	38
4.2 Cài đặt thông số thực nghiệm	40
4.3 Thước đo đánh giá	41
4.4 Kết quả thực nghiệm	41
Chương 5. Kết luận và mở rộng	46
Tài liệu tham khảo	47

Lời nói đầu

Trong bối cảnh công nghệ phát triển vượt bậc, việc phân tích và hiểu rõ cảm xúc của khách hàng đã trở thành một yếu tố then chốt giúp các doanh nghiệp nâng cao chất lượng dịch vụ và tạo dựng lòng tin từ khách hàng. Ngành hàng sách, với sự đa dạng về sản phẩm và sự cạnh tranh ngày càng gay gắt, cũng không nằm ngoài xu hướng này. Hiểu được cảm xúc của khách hàng không chỉ giúp các nhà kinh doanh cải thiện dịch vụ mà còn định hướng chiến lược kinh doanh một cách hiệu quả hơn.

Khoá luận tốt nghiệp với đề tài **‘Phân tích cảm xúc khách hàng trong ngành hàng Sách’** nhằm mục đích áp dụng các kỹ thuật xử lý ngôn ngữ tự nhiên và máy học để thu thập, phân tích và đánh giá cảm xúc của khách hàng qua các đánh giá và bình luận trên các nền tảng trực tuyến. Qua đó, nghiên cứu không chỉ đưa ra cái nhìn tổng quan về cảm xúc của khách hàng mà còn đề xuất các biện pháp cải thiện dựa trên kết quả phân tích.

Nội dung báo cáo khoá luận bao gồm 5 chương:

- 1. Giới thiệu:** Tổng quan, lý do chọn đề tài và mục đích của đề tài.
- 2. Tổng quan lý thuyết:** Xác định bài toán và các kiến thức liên quan cần biết.
- 3. Phương pháp sử dụng:** Trình bày các phương pháp tiếp cận được nghiên cứu và áp dụng.
- 4. Thực nghiệm và đánh giá mô hình:** Mô tả quá trình thực hiện nghiên cứu và đánh giá hiệu quả mô hình.
- 5. Kết luận và mở rộng:** Đưa ra kết luận và trình bày hướng đi trong tương lai của bài toán.

Chúng tôi hy vọng rằng luận văn này sẽ đóng góp một phần nhỏ vào việc nâng cao nhận thức về tầm quan trọng của phân tích cảm xúc trong kinh doanh, đặc biệt là trong ngành hàng sách. Đồng thời, chúng tôi mong rằng những kiến thức

và kỹ năng thu được từ nghiên cứu này sẽ là nền tảng vững chắc cho sự nghiệp sau này của bản thân.

Danh sách các hình

Hình 2.1:	Phân loại các kỹ thuật phân tích cảm xúc	13
Hình 2.2:	RNN với vòng lặp	17
Hình 2.3:	Cấu trúc trái phằng của RNN	18
Hình 2.4:	Sự lặp lại kiến trúc module trong mạng RNN chứa một tầng ẩn	19
Hình 2.5:	Sự lặp lại kiến trúc module trong mạng LSTM chứa 4 tầng ẩn (3 sigmoid và 1 tanh)	19
Hình 2.6:	Đường đi của ô trạng thái (cell state) trong mạng LSTM . . .	19
Hình 2.7:	Một cổng của hàm sigmoid trong LSTM	20
Hình 2.8:	Forget gate layer	20
Hình 2.9:	Cập nhật giá trị cho ô trạng thái	21
Hình 2.10:	Ô trạng thái mới	21
Hình 2.11:	Điều chỉnh thông tin ở đầu ra thông qua hàm tanh	22
Hình 2.12:	Cổng xóa và cổng cập nhật trong GRU	23
Hình 2.13:	Tính toán trạng thái ẩn tiềm năng trong một GRU	24
Hình 2.14:	Tính toán trạng thái ẩn trong GRU	25
Hình 3.1:	Quy trình tiền xử lý dữ liệu	28
Hình 3.2:	Kiến trúc mô hình sử dụng word embedding ELMo	31
Hình 3.3:	Kiến trúc mô hình sử dụng PhoBERT	33
Hình 3.4:	Lớp đầu vào và các lớp ẩn	36
Hình 3.5:	Lớp đầu ra sử dụng phương pháp multi-task learning	37
Hình 4.1:	Sự phân bố cảm xúc của từng khía cạnh trong tập dữ liệu . .	39

Danh sách các bảng

Bảng 3.1: Định nghĩa các khía cạnh	30
Bảng 3.2: Định nghĩa các cảm xúc	30
Bảng 4.1: Một số ví dụ về tập dữ liệu	40
Bảng 4.2: Thống kê tổng quan về tập dữ liệu	40
Bảng 4.3: Hiệu suất của các mô hình áp dụng trên tập dữ liệu sách . . .	42
Bảng 4.4: F1-score cho tác vụ phát hiện khía cạnh (Bi-GRU + Conv1D)	43
Bảng 4.5: F1-score cho tác vụ phân loại cảm xúc (Bi-GRU + Conv1D) .	45

Danh sách các từ viết tắt

- NLP: Natural Language Processing
- ML: Machine Learning
- NN: Neuron Network
- ABSA: Aspect-based Sentiment Analysis
- DL: Deep Learning
- BERT: Bidirectional Encoder Representations from Transformers
- DNN: Deep Neuron Network
- k-NN: k-Nearest Neighbors
- SVM: Support Vector Machines
- CNN: Convolutional Neural Network
- RNN: Recurrent Neural Network
- TF-IDF: Term Frequency-Inverse Document Frequency
- CBOW: Continious Bag-Of-Words
- LSTM: Long Short-Term Memory
- GRU: Gated Recurrent Unit
- NER: Named Entity Recognition
- HTML: HyperText Markup Language
- Bi-LSTM: Bidirectional Long Short-Term Memory
- Bi-GRU: Bidirectional Gated Recurrent Unit

Danh sách thuật ngữ

Các chú thích cho thuật ngữ chuyên ngành được dùng trong khoá luận:

Natural Language Processing: Xử lý ngôn ngữ tự nhiên

Neuron Network: Mạng neuron

Aspect-based Sentiment Analysis: Một kỹ thuật phân tích văn bản giúp phân loại dữ liệu theo khía cạnh và xác định quan điểm được gán cho từng khía cạnh đó.

Span Detection: Một kỹ thuật trong NLP dùng để xác định các đoạn văn bản cụ thể trong một tài liệu mà đáp ứng một số tiêu chí nhất định.

Sentiment Score: Kết quả phân tích chỉ số cảm xúc của các đánh giá để xác định tương quan giữa các đánh giá tích cực và tiêu cực.

Positive: Tích cực.

Negative: Tiêu cực.

Neutral: Trung lập.

Segmentation: Quá trình phân chia một văn bản dài thành các đoạn nhỏ hơn có ý nghĩa, chẳng hạn như câu, đoạn văn hoặc các đơn vị nhỏ hơn như từ.

Word Embedding: Một kỹ thuật trong NLP được sử dụng để ánh xạ các từ thành các vector số trong một không gian nhiều chiều.

TF-IDF: Một kỹ thuật trong NLP được sử dụng để đánh giá tầm quan trọng của một từ trong một văn bản cụ thể so với toàn bộ tập hợp văn bản.

SpatialDropout1D: Một kỹ thuật regularization nhằm giảm overfitting bằng cách ngẫu nhiên bỏ qua một số đơn vị trong quá trình huấn luyện.

Conv1D: Một convolutional layer được sử dụng cho dữ liệu một chiều như chuỗi thời gian và dữ liệu văn bản.

Chương 1

Giới thiệu

1.1 Lý do chọn đề tài

Trong bối cảnh thị trường ngày càng cạnh tranh và khách hàng ngày càng trở nên khó tính hơn, việc hiểu rõ cảm xúc và đánh giá của khách hàng đối với sản phẩm và dịch vụ là một yếu tố quyết định đến sự thành công của doanh nghiệp. Ngành hàng sách cũng không nằm ngoài xu hướng này. Đặc biệt, với sự bùng nổ của thương mại điện tử và các nền tảng đánh giá trực tuyến, việc phân tích cảm xúc của khách hàng trở thành một công cụ mạnh mẽ giúp các nhà kinh doanh hiểu sâu hơn về nhu cầu và mong muốn của khách hàng, từ đó cải thiện chất lượng dịch vụ và gia tăng sự hài lòng của khách hàng.

Một trong những lý do cụ thể khiến chúng tôi chọn đề tài **‘Phân tích cảm xúc khách hàng trong ngành hàng Sách’** là nhờ vào một nghiên cứu của công ty nghiên cứu thị trường Nielsen¹ năm 2021. Nghiên cứu này cho thấy rằng 41.6% khách hàng Mỹ tìm kiếm lời khuyên từ người khác và 70% có xu hướng đọc các bình luận trực tuyến trước khi đưa ra quyết định mua hàng. Điều này chứng tỏ rằng cảm xúc và đánh giá của khách hàng không chỉ ảnh hưởng đến quyết định mua hàng mà còn có thể tạo ra sự khác biệt lớn về doanh số bán hàng. Điều này đã thôi thúc chúng tôi nghiên cứu sâu hơn về cách áp dụng các kỹ thuật NLP và ML để phân tích cảm xúc của khách hàng một cách hiệu quả.

Mục tiêu của đề tài là áp dụng các kỹ thuật tiên tiến để thu thập, phân tích và đánh giá cảm xúc của khách hàng từ các bình luận và đánh giá trực tuyến. Kết quả phân tích không chỉ giúp hiểu rõ hơn về cảm xúc của khách hàng mà còn cung cấp những thông tin quý giá giúp nhà sách cải thiện chất lượng dịch vụ và phát triển chiến lược kinh doanh một cách hiệu quả hơn.

¹<https://www.nielsen.com/insights/2021/making-personal-connections-how-brands-are-leveraging-social-influencers/>

Chúng tôi tin rằng việc phân tích cảm xúc khách hàng không chỉ mang lại lợi ích thiết thực cho doanh nghiệp mà còn góp phần nâng cao trải nghiệm của khách hàng, tạo ra một môi trường kinh doanh lành mạnh và bền vững trong ngành hàng sách.

1.2 Trình bày bài toán

Phân tích cảm xúc theo khía cạnh (ABSA) là một nhiệm vụ trong NLP. Thay vì chỉ phân tích cảm xúc tổng thể của một câu hoặc đoạn văn, bài toán ABSA giúp phân tích cảm xúc theo từng khía cạnh được thể hiện trong một câu hoặc đoạn văn. Ví dụ, trong các đánh giá về sách - truyện, khách hàng không phải lúc nào cũng đưa ra đánh giá chung về sản phẩm mà họ có xu hướng đánh giá dựa trên các khía cạnh của sách như nội dung, chất lượng,... Khách hàng có thể thích bìa sách nhưng không thích nội dung của sách vì nó quá tệ. Bài toán ABSA giúp nhận biết được những khía cạnh nào mà khách hàng yêu thích và những khía cạnh nào mà khách hàng không yêu thích, từ đó hỗ trợ nhà kinh doanh nâng cao chất lượng sản phẩm và dịch vụ nhằm đáp ứng nhu cầu của khách hàng. Đây là lý do tại sao ABSA thực sự cần thiết để hiểu sâu hơn về những gì khách hàng nghĩ về sản phẩm.

Với sự bùng nổ của thương mại điện tử, hàng triệu đánh giá được tạo ra mỗi ngày từ khách hàng. Họ để lại rất nhiều thông tin hữu ích không chỉ cho người bán để hiểu được khách hàng thích và không thích gì về sản phẩm của họ mà còn cho các khách hàng khác đọc và cân nhắc trải nghiệm của những người mua trước đó trước khi quyết định mua hàng. Có nhiều loại sản phẩm trong thương mại điện tử, nhưng ngày nay sản phẩm liên quan tới sách - truyện là một trong những sản phẩm được ưa chuộng nhất để mua trực tuyến. Thêm vào đó, các đánh giá về sản phẩm sách - truyện chứa rất nhiều khía cạnh, từ giá cả, chất lượng bên ngoài đến nội dung bên trong, nhờ đó mà nhiều phân tích có thể được thực hiện để hiểu sâu hơn thái độ của khách hàng đối với 1 quyển sách hay truyện. Đây là lý do tại sao chúng tôi chọn các đánh giá về sách - truyện để xây dựng một tập dữ liệu nhằm giải quyết bài toán ABSA trong các đánh giá thương mại điện tử.

Khoá luận này có ba đóng góp chính:

- Thứ nhất, cung cấp một tập dữ liệu tiếng việt về bình luận đánh giá của khách hàng về sách. Góp phần tạo nên sự đa dạng trong việc nghiên cứu lĩnh vực ABSA.
- Thứ hai, đề xuất một số kiến trúc mô hình học sâu cho bài toán ABSA, góp phần tạo nên sự đa dạng trong phương pháp nghiên cứu cho bài toán ABSA.
- Cuối cùng, một so sánh toàn diện giữa các mô hình được thực hiện để tìm ra phương pháp tốt nhất cho bài toán phân tích cảm xúc theo khía cạnh.

Chương 2

Tổng quan lý thuyết

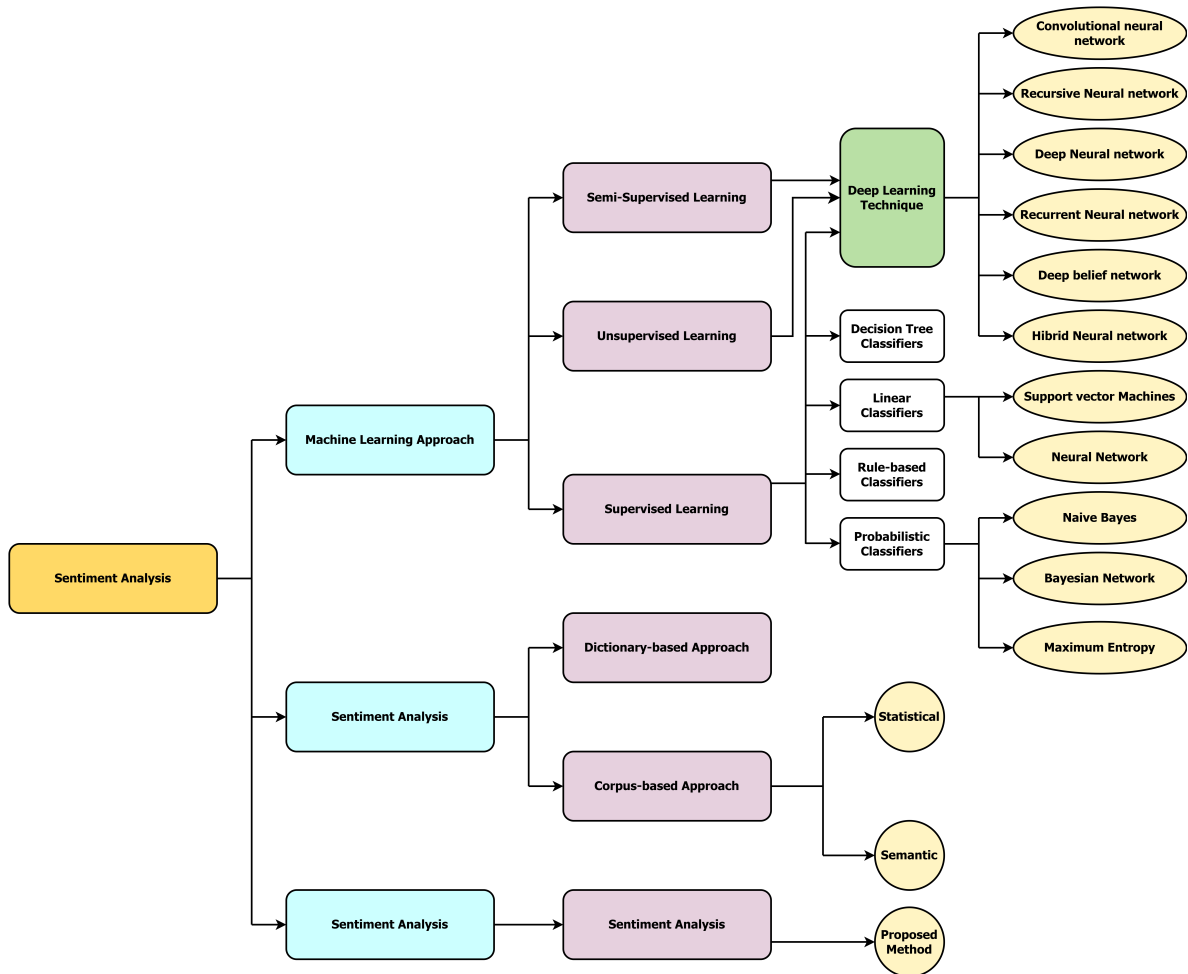
2.1 Sentiment Analysis (SA)

Phân tích cảm xúc [1] là một quá trình trích xuất thông tin về một thực thể và tự động xác định bất kỳ tính chủ quan nào của thực thể đó. Mục tiêu là xác định liệu văn bản do người dùng tạo ra thể hiện ý kiến Positive, Negative hay Neutral của họ hay không. Phân loại cảm xúc có thể được thực hiện trên ba mức độ trích xuất: mức độ khía cạnh hoặc đặc điểm, mức độ câu và mức độ tài liệu. Hiện tại, có ba phương pháp để giải quyết vấn đề phân tích cảm xúc: (1) các kỹ thuật dựa trên từ điển, (2) các kỹ thuật dựa trên ML và (3) các phương pháp kết hợp.

- **Các kỹ thuật dựa trên từ điển** là những kỹ thuật đầu tiên được sử dụng cho phân tích cảm xúc. Chúng được chia thành hai cách tiếp cận: dựa trên từ điển và dựa trên tập hợp văn bản. Trong cách tiếp cận đầu tiên, phân loại cảm xúc được thực hiện bằng cách sử dụng một từ điển các thuật ngữ, chẳng hạn như những từ được tìm thấy trong SentiWordNet và WordNet. Tuy nhiên, phân tích cảm xúc dựa trên tập hợp văn bản không dựa vào một từ điển được xác định trước mà dựa vào phân tích thống kê nội dung của một tập hợp các tài liệu, sử dụng các kỹ thuật dựa trên k-NN, Conditional Random Field và Hidden Markov Models cùng các kỹ thuật khác.
- **Các kỹ thuật dựa trên ML** được đề xuất cho các vấn đề phân tích cảm xúc có thể được chia thành hai nhóm: (1) các mô hình truyền thống và (2) các mô hình DL. Các mô hình truyền thống đề cập đến các kỹ thuật ML cổ điển, chẳng hạn như Naïve Bayes classifier, maximum entropy classifier hoặc SVM. Đầu vào cho các thuật toán này bao gồm các đặc trưng từ vựng, các đặc trưng dựa trên từ điển cảm xúc, các thành phần từ loại hoặc các tính từ và trạng từ. Độ chính xác của các hệ thống này phụ thuộc vào các đặc trưng

được chọn. Các mô hình DL có thể cung cấp kết quả tốt hơn so với các mô hình truyền thống. Nhiều loại mô hình DL có thể được sử dụng cho phân tích cảm xúc, bao gồm CNN, DNN và RNN. Các phương pháp này giải quyết các vấn đề phân loại ở mức tài liệu, mức câu hoặc mức khía cạnh.

- **Các phương pháp kết hợp** kết hợp các kỹ thuật dựa trên từ điển và ML. Các từ điển cảm xúc thường đóng vai trò quan trọng trong hầu hết các phương pháp này. Hình 2.1 minh họa phân loại các kỹ thuật phân tích cảm xúc.



Hình 2.1: Phân loại các kỹ thuật phân tích cảm xúc

2.2 Aspect-based Sentiment Analysis (ABSA)

Phân tích cảm xúc dựa trên khía cạnh [2] là một phân nhánh của phân tích cảm xúc chi tiết. ABSA tập trung vào việc xác định cảm xúc đối với các thực thể cụ thể hoặc các thuộc tính/đặc điểm của chúng gọi là các khía cạnh. Một khía cạnh

có thể được thể hiện rõ ràng trong văn bản (khía cạnh rõ ràng) hoặc không có trong văn bản nhưng được ngụ ý từ ngữ cảnh (khía cạnh ngầm). Hơn nữa, cảm xúc ở cấp độ khía cạnh có thể khác nhau giữa các khía cạnh và khác với cảm xúc tổng thể của câu hoặc tài liệu. Một số nghiên cứu còn phân biệt khía cạnh thành thuật ngữ khía cạnh và danh mục khía cạnh, trong đó thuật ngữ khía cạnh đề cập đến biểu hiện khía cạnh trong văn bản đầu vào (ví dụ: ‘pizza’) và danh mục khía cạnh là một cấu trúc ẩn, thường là một danh mục cao cấp hơn của các thuật ngữ khía cạnh (ví dụ: ‘thực phẩm’) được xác định hoặc cho sẵn. Những ví dụ sau đây minh họa các thuật ngữ của ABSA:

Ví dụ 1 (từ một đánh giá nhà hàng): **‘Nhà hàng đắt đỏ, nhưng thực đơn rất phong phú.’**

Câu này có một khía cạnh rõ ràng ‘thực đơn’ (thuật ngữ cảm xúc: ‘phong phú’, tính phân cực của cảm xúc: Positive), một khía cạnh ngầm ‘giá cả’ (thuật ngữ cảm xúc: ‘đắt đỏ’, tính phân cực của cảm xúc: Negative). Tùy thuộc vào các danh mục mục tiêu, các khía cạnh có thể được phân loại vào các danh mục, chẳng hạn như ‘thực đơn’ vào ‘tổng quát’ và ‘giá cả’ vào ‘giá cả’.

Ví dụ 2 (từ một đánh giá laptop): **‘Mang theo cực kì tiện lợi, dễ dàng kết nối với WIFI tại thư viện và nhiều nơi khác.’**

Câu này có hai khía cạnh ngầm: ‘tính di động’ (thuật ngữ cảm xúc: ‘tiện lợi’, tính phân cực của cảm xúc: Positive), ‘khả năng kết nối’ (thuật ngữ cảm xúc: ‘dễ dàng’, tính phân cực của cảm xúc: Positive). Các khía cạnh này có thể được phân loại vào các danh mục, chẳng hạn như cả hai dưới danh mục ‘laptop’ (khác với ‘phần mềm’ hoặc ‘hỗ trợ’).

Ví dụ 3 (từ một đánh giá khóa học): **‘Khoá học rất khó và nó có một khối lượng công việc điên rồ, tôi không khuyến khích sinh viên mới vào học mặc dù buổi hướng dẫn và giảng viên rất hữu ích.’**

Hai khía cạnh rõ ràng trong ví dụ 3 là ‘buổi hướng dẫn’ và ‘giảng viên’ (thuật ngữ cảm xúc: ‘hữu ích’, tính phân cực: Positive). Các khía cạnh ngầm là ‘nội dung’ (thuật ngữ cảm xúc: ‘rất khó’, tính phân cực của cảm xúc: Negative), ‘khối lượng công việc’ (thuật ngữ cảm xúc: ‘khối lượng công việc điên rồ’, tính phân cực của

cảm xúc: Negative) và ‘khóa học’ (thuật ngữ cảm xúc: ‘không khuyến khích’, tính phân cực của cảm xúc: Negative). Một minh họa cho các danh mục khía cạnh là việc gán khía cạnh ‘giảng viên’ vào danh mục tổng quát hơn như là ‘nhân viên’ và ‘buổi hướng dẫn’ vào danh mục ‘thành phần khóa học’.

Như được minh họa ở trên, tính chi tiết cao giúp ABSA có mục tiêu cụ thể và cung cấp nhiều thông tin hơn so với phân tích cảm xúc ở mức tài liệu hoặc câu. Do đó, ABSA có thể được ứng dụng để đo lường trọng số thuộc tính trong đánh giá tổng thể, tóm tắt ý kiến dựa trên khía cạnh, hệ thống đề xuất tự động và cá nhân hoá.

2.3 Các nghiên cứu có liên quan

Trong tiếng Việt, có một số bộ dữ liệu về phân tích cảm xúc trong nhiều lĩnh vực. Năm 2018, Minh Huyen và các đồng nghiệp đã công bố bộ dữ liệu SA-VLSP2018 [3] cho bài toán phân tích cảm xúc dựa trên khía cạnh về các lĩnh vực khách sạn và nhà hàng trong hội thảo VLSP. Năm 2021, Kim Thanh và các đồng nghiệp đã áp dụng Span Detection [4] cho bài toán phân tích cảm xúc dựa trên khía cạnh và đạt hiệu suất 62,76% F1-score cho bộ dữ liệu UIT-ViSD4SA. Năm 2018, Van Kiet và các đồng nghiệp đã công bố bộ dữ liệu UIT-VSFC [5], bao gồm hơn 16.000 câu phản hồi từ sinh viên. Bộ dữ liệu này được sử dụng cho phân loại cảm xúc và phân loại chủ đề.

Năm 2010, một nhóm tác giả đã triển khai một phương pháp phân tích cảm xúc tự động cho các đánh giá phim đã được đề xuất và thực hiện [6]. Phương pháp này sử dụng phân tích chi tiết để xác định cả hướng và mức độ cảm xúc của người đánh giá đối với các khía cạnh khác nhau của bộ phim. Nó áp dụng một phương pháp ngôn ngữ học để tính toán cảm xúc của một thuật ngữ từ Sentiment Score trước đó.

Năm 2018, Hai Ha và các đồng nghiệp đã chỉ ra rằng khối lượng nội dung do người dùng tạo ra trên web ngày càng tăng đã biến phân tích cảm xúc thành một công cụ thiết yếu trong việc trích xuất thông tin về trạng thái cảm xúc của con người [7]. Thông qua bài báo, tác giả đã cung cấp một cái nhìn so sánh tổng quan về DL

cho bài toán phân tích cảm xúc dựa trên khía cạnh để áp dụng các phương pháp khác nhau tùy vào ngữ cảnh.

Năm 2020, Oanh và các đồng nghiệp đã mô tả ABSA đại diện cho nhiệm vụ xác định các cực cảm xúc cho các thực thể cụ thể và các khía cạnh được đề cập trong đánh giá của khách hàng [8]. Bài báo này trình bày một mô hình phân cấp mới và hiệu quả sử dụng mô hình ngôn ngữ tiền huấn luyện đại diện cho BERT.

Năm 2020, một nghiên cứu được thực hiện bởi Van Thin và các đồng nghiệp nhằm xây dựng một mô hình DNN để giải quyết hai nhiệm vụ trong phân tích cảm xúc dựa trên khía cạnh của bộ dữ liệu tiếng Việt [9]. Họ nhận thấy mô hình của họ hoạt động tốt hơn so với các phương pháp tiếp cận thông thường trên cả hai nhiệm vụ cho hai lĩnh vực và đạt được F1-score là 64,78% cho lĩnh vực nhà hàng và 70,90% cho lĩnh vực khách sạn.

2.4 Word embedding

Word embedding [10] là một kỹ thuật cho mô hình ngôn ngữ và học đặc trưng, trong đó mỗi từ được ánh xạ tới một vector các giá trị thực sao cho các từ mang nghĩa tương tự có đại diện tương tự nhau. Việc học giá trị có thể được thực hiện bằng các mạng neuron. Một hệ thống word embedding thường được sử dụng là Word2vec (GloVe hoặc Gensim) [11], chứa các mô hình như skip-gram và CBOW. Cả hai mô hình đều dựa trên xác suất xuất hiện gần nhau của các từ. Skip-gram cho phép bắt đầu với một từ và dự đoán các từ có khả năng xuất hiện xung quanh nó. CBOW thì ngược lại, dự đoán một từ có khả năng xuất hiện dựa trên các ngữ cảnh cụ thể.

2.5 TF-IDF

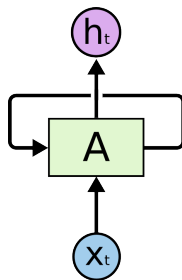
TF-IDF [12] là một thước đo thống kê phản ánh tầm quan trọng của một từ trong một văn bản cụ thể so với toàn bộ tập hợp văn bản. Thước đo này xem xét tần suất của từ trong tài liệu mục tiêu cũng như tần suất của từ trong các tài liệu khác. Tần suất của một từ trong tài liệu mục tiêu càng cao và tần suất của nó trong các tài liệu khác càng thấp thì tầm quan trọng của nó càng lớn. Lớp vectorizer trong thư viện scikit-learn thường được sử dụng để tính TF-IDF.

Cả word embedding và TF-IDF đều được sử dụng làm các đặc trưng đầu vào của các thuật toán DL trong NLP. Trong nhiệm vụ phân tích cảm xúc, các tập hợp dữ liệu thô được chuyển thành các vector liên tục mang giá trị thực.

2.6 Recurrent Neural Network (RNN)

2.6.1 Cấu trúc

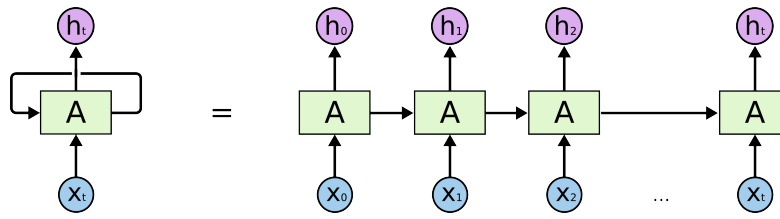
Trong lý thuyết về ngôn ngữ, ngữ nghĩa của một câu được tạo thành từ mối liên kết của những từ trong câu theo một cấu trúc ngữ pháp. Nếu xét từng từ một đứng riêng lẻ ta không thể hiểu được nội dung của toàn bộ câu, nhưng dựa trên những từ xung quanh ta có thể hiểu được trọn vẹn một câu nói. Như vậy cần phải có một kiến trúc đặc biệt hơn cho các mạng neuron biểu diễn ngôn ngữ nhằm mục đích liên kết các từ liên trước với các từ ở hiện tại để tạo ra mối liên hệ xâu chuỗi. RNN [13] đã được thiết kế đặc biệt để giải quyết yêu cầu này.



Hình 2.2: RNN với vòng lặp [14]

Hình 2.2 biểu diễn kiến trúc của một RNN. Trong kiến trúc này mạng neuron sử dụng một đầu vào là một vector x_t và trả ra đầu ra là một giá trị ẩn h_t . Đầu vào được đầu vào với một thân mạng neural A có tính chất truy hồi và thân này được đầu vào tới đầu ra h_t

Vòng lặp A ở thân mạng neuron là điểm mấu chốt trong nguyên lý hoạt động của mạng neural truy hồi. Đây là chuỗi sao chép nhiều lần của cùng một kiến trúc nhằm cho phép các thành phần có thể kết nối liên mạch với nhau theo mô hình chuỗi. Đầu ra của vòng lặp trước chính là đầu vào của vòng lặp sau. Nếu trải phẳng thân mạng neural A ta sẽ thu được một mô hình như Hình 2.3:



Hình 2.3: Cấu trúc trải phẳng của RNN [14]

2.6.2 Hạn chế

Một trong những đặc điểm nổi bật của RNN là khả năng kết nối thông tin từ quá khứ với nhiệm vụ hiện tại. Ví dụ, trong câu ‘học sinh đang tới trường học’, từ ‘trường học’ có thể được dự đoán ngay lập tức mà không cần thông tin bổ sung từ các câu xung quanh. Tuy nhiên, có những tình huống yêu cầu nhiều thông tin hơn, chẳng hạn: ‘Hôm qua, A đi học nhưng không mang áo mưa. Trên đường đi học, trời mưa. Cặp sách của A bị ướt’. Trong trường hợp này, để suy ra từ ‘ướt’, cần phải biết các sự kiện trước đó như trời mưa và không mang áo mưa. Những liên kết ngữ nghĩa dài như vậy được gọi là phụ thuộc dài hạn (long-term dependencies).

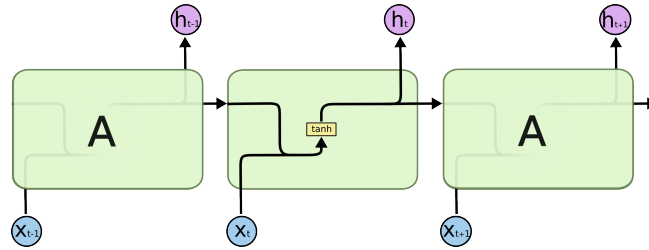
Về lý thuyết, mạng RNN có khả năng giải quyết các phụ thuộc dài hạn. Tuy nhiên, trong thực tế, RNN thường gặp khó khăn khi học các phụ thuộc này. Nguyên nhân chính là do hiện tượng triệt tiêu đạo hàm (vanishing gradient) của hàm cost function khi trải qua chuỗi dài các tính toán truy hồi. Để khắc phục hiện tượng này, một phiên bản cải tiến của mạng RNN là mạng LSTM (Long Short-Term Memory) đã được ra đời, nhờ vào cơ chế đặc biệt của nó.

2.6.3 Long Short-Term Memory (LSTM)

LSTM [13], được giới thiệu bởi Hochreiter và Schmidhuber vào năm 1997 [15], là một kiến trúc đặc biệt của mạng neuron hồi quy (RNN), có khả năng học và ghi nhớ các phụ thuộc dài hạn. Kiến trúc này đã trở nên phổ biến và được sử dụng rộng rãi cho đến ngày nay. LSTM khắc phục nhiều hạn chế của RNN truyền thống, đặc biệt là vấn đề triệt tiêu đạo hàm. Mặc dù cấu trúc của LSTM phức tạp hơn, nó vẫn giữ được ý tưởng chính của RNN là sự sao chép các kiến trúc theo dạng chuỗi.

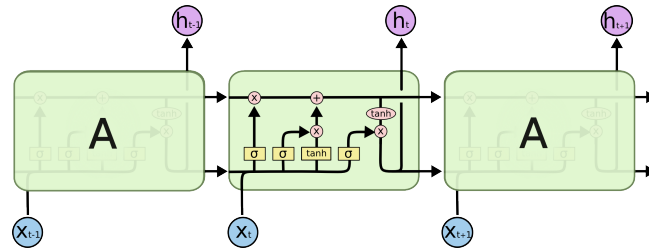
Một mạng RNN tiêu chuẩn có kiến trúc đơn giản, chẳng hạn với một tầng ẩn sử

dùng hàm tanh như Hình 2.4.



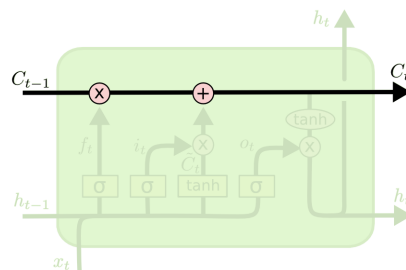
Hình 2.4: Sự lặp lại kiến trúc module trong mạng RNN chứa một tầng ẩn [14]

LSTM cũng có cấu trúc chuỗi tương tự, nhưng phần kiến trúc lặp lại phức tạp hơn, bao gồm bốn tầng ẩn (ba tầng sigmoid và một tầng tanh) tương tác với nhau theo một cách đặc biệt.



Hình 2.5: Sự lặp lại kiến trúc module trong mạng LSTM chứa 4 tầng ẩn (3 sigmoid và 1 tanh) [14]

Ý tưởng chính của LSTM là sử dụng ô trạng thái (cell state), được thể hiện qua đường chạy ngang trên đỉnh đồ thị như Hình 2.6.

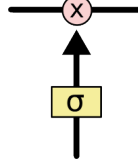


Hình 2.6: Đường đi của ô trạng thái (cell state) trong mạng LSTM [14]

Ô trạng thái hoạt động như một băng chuyền, truyền thông tin dọc theo chuỗi với ít sự can thiệp, giúp cho thông tin được truyền ổn định. LSTM có khả năng xóa và thêm thông tin vào ô trạng thái thông qua các cấu trúc gọi là cổng (gates).

Cổng là cơ chế điều chỉnh luồng thông tin, bao gồm một tầng ẩn sử dụng hàm

sigmoid và một toán tử nhân. Hàm sigmoid cho đầu ra là một giá trị xác suất từ 0 đến 1, quyết định lượng thông tin được phép đi qua. Một mạng LSTM có ba cổng để bảo vệ và kiểm soát các ô trạng thái.

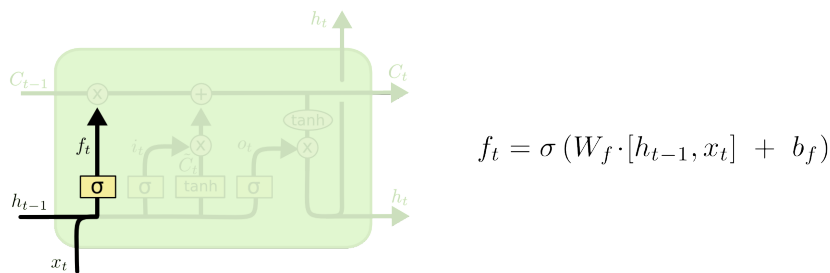


Hình 2.7: Một cổng của hàm sigmoid trong LSTM [14]

Các bước hoạt động của LSTM

Bước đầu tiên là quyết định thông tin nào được phép đi qua ô trạng thái, được kiểm soát bởi hàm sigmoid trong tầng quên (forget gate layer). Đầu tiên nó nhận đầu vào là 2 giá trị h_{t-1} và x_t và trả về một giá trị nằm trong khoảng 0 và 1 cho mỗi giá trị của ô trạng thái C_{t-1} . Nếu giá trị bằng 1 thể hiện ‘giữ toàn bộ thông tin’ và bằng 0 thể hiện ‘bỏ qua toàn bộ chúng’.

Trở lại ví dụ về ngôn ngữ, chúng ta đang cố gắng dự báo từ tiếp theo dựa trên toàn bộ những từ trước đó. Trong những bài toán như vậy, ô trạng thái có thể bao gồm loại của chủ ngữ hiện tại, để cho đại từ ở câu tiếp theo được sử dụng chính xác. Tuy nhiên chủ ngữ không phải khi nào cũng cố định. Khi chúng ta nhìn thấy một chủ ngữ mới, chúng ta muốn quên đi loại của một chủ ngữ cũ. Do đó tầng quên cho phép cập nhật thông tin mới và lưu giữ giá trị của nó khi có thay đổi theo thời gian.

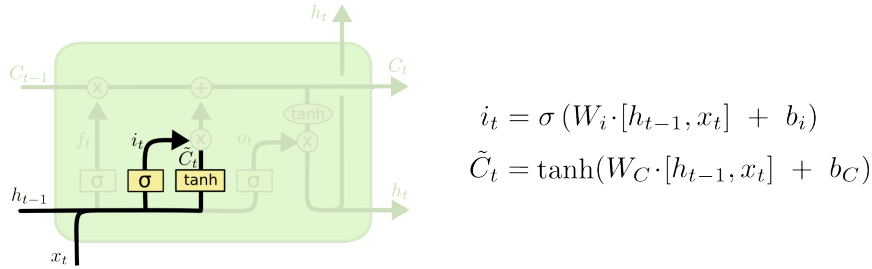


Hình 2.8: Forget gate layer [14]

Bước tiếp theo quyết định loại thông tin nào sẽ được lưu trữ trong ô trạng thái. Bước này bao gồm 2 phần. Phần đầu tiên là một tầng ẩn của hàm sigmoid được

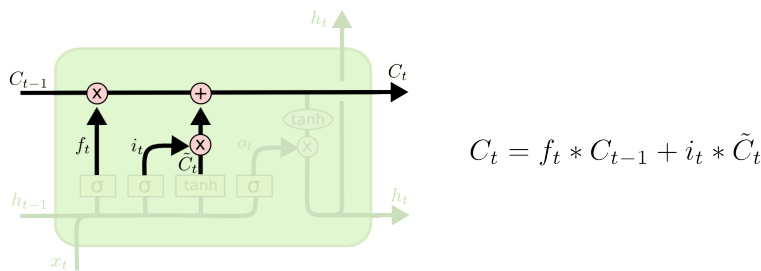
gọi là tầng cổng vào (input gate layer) quyết định giá trị bao nhiêu sẽ được cập nhật. Tiếp theo, tầng ẩn hàm tanh sẽ tạo ra một vector của một giá trị trạng thái mới \tilde{C}_t . Tiếp theo kết hợp kết quả của 2 tầng này để tạo thành một cặp nhật cho trạng thái.

Trong ví dụ của mô hình ngôn ngữ, chúng ta muốn thêm một chủ ngữ mới vào ô trạng thái để thay thế phần trạng thái cũ muốn quên đi.



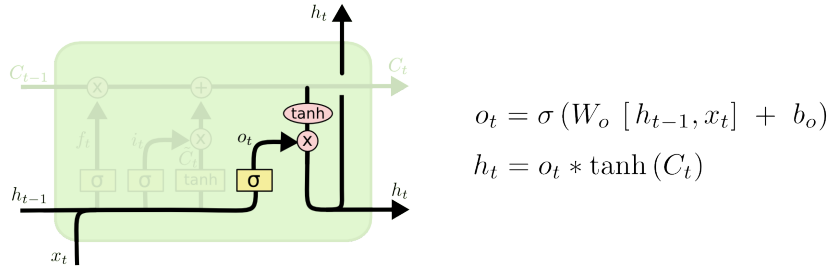
Hình 2.9: Cập nhật giá trị cho ô trạng thái [14]

Ô trạng thái cũ được cập nhật, C_{t-1} sang một trạng thái mới C_t . Trạng thái cũ nhân với giá trị từ tầng quên để loại bỏ thông tin không cần thiết. Phần tử $i_t * \tilde{C}_t$ là một giá trị mới được tính toán tương ứng với giá trị được cập nhật vào mỗi trạng thái.



Hình 2.10: Ô trạng thái mới [14]

Cuối cùng, đầu ra của LSTM dựa trên ô trạng thái đã qua tầng sigmoid và hàm tanh, quyết định phần nào của ô trạng thái sẽ ở đầu ra.



Hình 2.11: Điều chỉnh thông tin ở đầu ra thông qua hàm tanh [14]

2.6.4 Gated Recurrent Unit (GRU)

GRU [13] là một loại mạng neuron hồi quy (RNN) tiên tiến trong lĩnh vực học DL, được thiết kế để cải thiện hiệu suất so với các RNN truyền thống bằng cách giảm thiểu vấn đề vanishing gradient. Cụ thể, GRU được trang bị các cơ chế học tập để xác định thời điểm nên cập nhật và khi nào nên xóa trạng thái ẩn. Điều này mang lại cho GRU tính linh hoạt và hiệu quả cao hơn trong việc xử lý chuỗi dữ liệu dài và phức tạp.

Ví dụ, nếu ký tự đầu tiên trong chuỗi dữ liệu có mức độ quan trọng cao, GRU có thể học cách không cập nhật trạng thái ẩn sau lần quan sát đầu tiên. Ngược lại, GRU cũng có thể bỏ qua những quan sát tạm thời không quan trọng và chỉ cập nhật trạng thái ẩn khi cần thiết.

Chính nhờ những cơ chế này mà GRU có khả năng duy trì và sử dụng thông tin quan trọng trong suốt quá trình huấn luyện mô hình, đồng thời loại bỏ những thông tin không cần thiết một cách hiệu quả. Điều này giúp cải thiện đáng kể hiệu suất và độ chính xác của các mô hình học sâu khi làm việc với các dữ liệu chuỗi. Dưới đây là cơ chế hoạt động và lợi ích của GRU trong các ứng dụng thực tế.

Cổng xóa và cổng cập nhật

Trong GRU, hai thành phần quan trọng là cổng xóa (reset gate) và cổng cập nhật (update gate). Chúng được thiết kế dưới dạng các vector có các phần tử nằm trong khoảng từ 0 đến 1, nhằm mục đích biểu diễn các tổ hợp logic. Các cổng này cho phép mô hình kiểm soát chính xác trạng thái ẩn tại mỗi thời điểm.

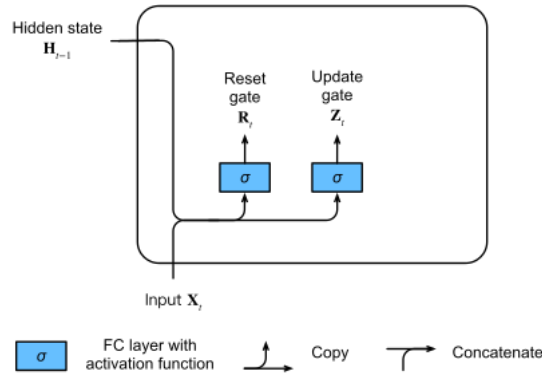
Cổng xóa: Cổng này quyết định bao nhiêu phần của trạng thái trước đây sẽ được

giữ lại và bao nhiêu phần sẽ bị xóa. Khi giá trị của cổng xóa gần bằng 1, phần lớn thông tin của trạng thái trước đó được giữ lại. Ngược lại, khi giá trị gần bằng 0, phần lớn thông tin bị loại bỏ.

Cổng cập nhật: Cổng này quyết định bao nhiêu phần của trạng thái mới sẽ giống với trạng thái cũ. Nếu giá trị của cổng cập nhật gần bằng 1, trạng thái ẩn mới sẽ gần giống với trạng thái trước đó. Nếu giá trị gần bằng 0, trạng thái ẩn mới sẽ chủ yếu dựa vào thông tin mới từ đầu vào hiện tại.

Cách tiếp cận này cho phép GRU điều chỉnh linh hoạt mức độ ảnh hưởng của các trạng thái trước đó và các thông tin mới, giúp mô hình nắm bắt tốt hơn các mối quan hệ dài hạn và loại bỏ các thông tin không cần thiết.

Hình 2.12 minh họa các đầu vào cho cả cổng xóa và cổng cập nhật trong GRU, với đầu vào ở bước thời gian hiện tại \mathbf{X}_t và trạng thái ẩn ở bước thời gian trước đó \mathbf{H}_{t-1} . Đầu ra được tạo bởi một tầng kết nối đầy đủ với hàm kích hoạt sigmoid.



Hình 2.12: Cổng xóa và cổng cập nhật trong GRU [16]

Tại bước thời gian t , với đầu vào minibatch là $\mathbf{X}_t \in \mathbb{R}^{n \times d}$ (số lượng mẫu: n , số lượng đầu vào: d) và trạng thái ẩn ở bước thời gian gần nhất là $\mathbf{H}_{t-1} \in \mathbb{R}^{n \times h}$ (số lượng trạng thái ẩn: h), cổng xóa $\mathbf{R}_t \in \mathbb{R}^{n \times h}$ và cổng cập nhật $\mathbf{Z}_t \in \mathbb{R}^{n \times h}$ được tính như sau:

$$\mathbf{R}_t = \sigma(\mathbf{X}_t \mathbf{W}_{xr} + \mathbf{H}_{t-1} \mathbf{W}_{hr} + \mathbf{b}_r),$$

$$\mathbf{Z}_t = \sigma(\mathbf{X}_t \mathbf{W}_{xz} + \mathbf{H}_{t-1} \mathbf{W}_{hz} + \mathbf{b}_z).$$

Ở đây, $\mathbf{W}_{xr}, \mathbf{W}_{xz} \in \mathbb{R}^{d \times h}$ và $\mathbf{W}_{hr}, \mathbf{W}_{hz} \in \mathbb{R}^{h \times h}$ là các tham số trọng số và $\mathbf{b}_r, \mathbf{b}_z \in \mathbb{R}^{1 \times h}$ là các hệ số điều chỉnh. Hàm sigmoid được sử dụng để biến đổi các giá trị

đầu vào nằm trong khoảng $(0, 1)$.

Hoạt động của cổng xóa

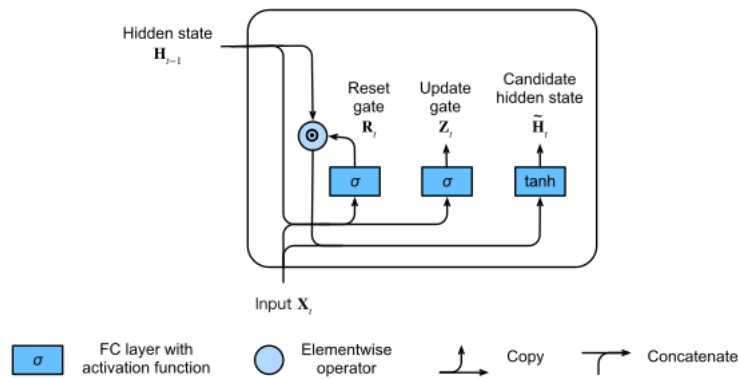
Để hiểu rõ cách GRU hoạt động, ta bắt đầu bằng việc tích hợp cổng xóa vào cơ chế cập nhật trạng thái tiềm ẩn. Trong RNN thông thường, trạng thái ẩn được cập nhật theo công thức:

$$\mathbf{H}_t = \tanh(\mathbf{X}_t \mathbf{W}_{xh} + \mathbf{H}_{t-1} \mathbf{W}_{hh} + \mathbf{b}_h).$$

Điều này về cơ bản giống với những gì đã thảo luận ở phần trước, mặc dù có thêm tính phi tuyến dưới dạng hàm tanh để đảm bảo rằng các giá trị trạng thái ẩn nằm trong khoảng $(-1, 1)$. Nếu muốn giảm ảnh hưởng của các trạng thái trước đó, ta có thể nhân \mathbf{H}_{t-1} với \mathbf{R}_t theo từng phần tử. Nếu các phần tử trong cổng xóa \mathbf{R}_t có giá trị gần với 1, kết quả sẽ giống RNN thông thường. Nếu tất cả các phần tử của cổng xóa \mathbf{R}_t gần với 0, trạng thái ẩn sẽ là đầu ra của một perceptron đa tầng với đầu vào là \mathbf{X}_t . Bất kỳ trạng thái ẩn nào tồn tại trước đó đều được đặt lại về giá trị mặc định. Tại đây nó được gọi là trạng thái ẩn tiềm năng và chỉ là tiềm năng vì ta vẫn cần kết hợp thêm đầu ra của cổng cập nhật.

$$\tilde{\mathbf{H}}_t = \tanh(\mathbf{X}_t \mathbf{W}_{xh} + (\mathbf{R}_t \odot \mathbf{H}_{t-1}) \mathbf{W}_{hh} + \mathbf{b}_h).$$

Hình 2.13 minh họa luồng tính toán sau khi áp dụng cổng xóa. Ký hiệu \odot biểu thị phép nhân theo từng phần tử giữa các tensor.



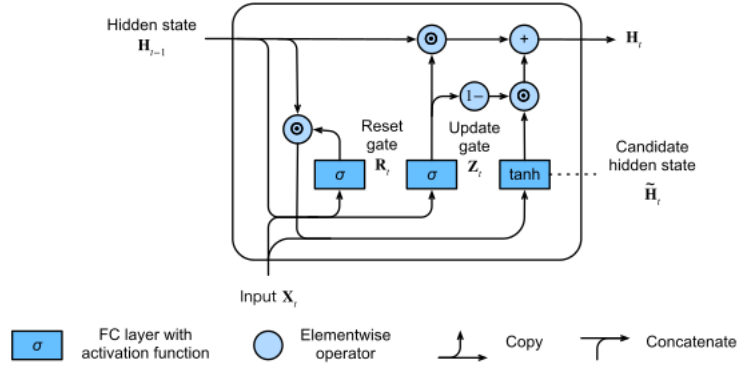
Hình 2.13: Tính toán trạng thái ẩn tiềm năng trong một GRU [16]

Hoạt động của cổng cập nhật

Tiếp theo ta sẽ kết hợp hiệu ứng của cổng cập nhật \mathbf{Z}_t như trong Hình 2.14. Cổng

này xác định mức độ giống nhau giữa trạng thái mới \mathbf{H}_t và trạng thái cũ \mathbf{H}_{t-1} , cũng như mức độ trạng thái ẩn tiềm năng $\tilde{\mathbf{H}}_t$ được sử dụng. Biến cổng (gating variable) \mathbf{Z}_t được sử dụng cho mục đích này, bằng cách áp dụng tổ hợp lỗi giữa trạng thái cũ và trạng thái tiềm năng. Ta có phương trình cập nhật cuối cùng cho GRU:

$$\mathbf{H}_t = \mathbf{Z}_t \odot \mathbf{H}_{t-1} + (1 - \mathbf{Z}_t) \odot \tilde{\mathbf{H}}_t.$$



Hình 2.14: Tính toán trạng thái ẩn trong GRU [16]

Nếu các giá trị trong cổng cập nhật \mathbf{Z}_t bằng 1, nó chỉ đơn giản giữ lại trạng thái cũ. Trong trường hợp này, thông tin từ \mathbf{X}_t về cơ bản được bỏ qua, tương đương với việc bỏ qua bước thời gian t trong chuỗi phụ thuộc. Ngược lại, nếu \mathbf{Z}_t gần giá trị 0, trạng thái ẩn \mathbf{H}_t sẽ gần với trạng thái ẩn tiềm năng $\tilde{\mathbf{H}}_t$. Những thiết kế trên có thể giúp giải quyết vấn đề vanishing gradient trong các mạng RNN và nắm bắt tốt hơn sự phụ thuộc xa trong chuỗi thời gian.

2.7 PhoBERT

PhoBERT [17] là một mô hình ngôn ngữ tiên tiến được thiết kế đặc biệt cho tiếng Việt, dựa trên kiến trúc BERT [18] nổi tiếng. Mô hình này được phát triển bởi VinAI Research, với mục tiêu cải thiện hiệu suất của các hệ thống xử lý ngôn ngữ tự nhiên (NLP) cho tiếng Việt.

Giống như BERT, PhoBERT sử dụng một kiến trúc Transformer hai chiều. Điều này cho phép mô hình nắm bắt ngữ cảnh của từ dựa trên cả hai hướng (trái và phải) trong một câu. PhoBERT được huấn luyện trên một lượng lớn dữ liệu tiếng Việt, bao gồm các văn bản từ Wikipedia, các trang web tin tức và các tài liệu văn

học.

2.8 ELMo

ELMo (Embeddings from Language Models) là một trong những mô hình tiên tiến trong lĩnh vực word embedding, được phát triển bởi AllenNLP [19]. ELMo cung cấp các biểu diễn ngữ nghĩa động cho các từ, nghĩa là các vector từ của ELMo có thể thay đổi dựa trên ngữ cảnh mà từ đó xuất hiện trong câu.

Khác với các mô hình word embedding tĩnh như Word2Vec [11] hay GloVe [20], ELMo sử dụng một mô hình ngôn ngữ hai chiều (BiLM) dựa trên các mạng LSTM. Mô hình này được huấn luyện trên một khối lượng văn bản lớn để dự đoán từ tiếp theo trong một chuỗi từ, cả từ trái sang phải và từ phải sang trái. Điều này cho phép ELMo nắm bắt các mối quan hệ ngữ nghĩa phức tạp và phụ thuộc xa trong văn bản.

Chương 3

Phương pháp sử dụng

3.1 Thu thập và tiền xử lý dữ liệu

3.1.1 Thu thập dữ liệu

Với mục tiêu phân tích cảm xúc theo khía cạnh cho đánh giá về sản phẩm sách - truyện trên sàn thương mại điện tử, chúng tôi đã tiến hành thu thập dữ liệu về bình luận, đánh giá của khách hàng trên sàn thương mại điện tử Tiki¹. Chúng tôi đã dùng hai thư viện Python là requests và BeautifulSoup4 [21] kết hợp với Open Product API² do Tiki cung cấp để tiến hành thu thập dữ liệu.

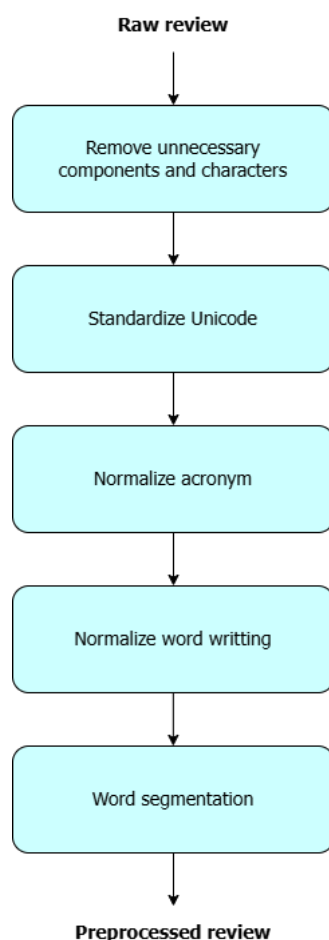
3.1.2 Tiền xử lý dữ liệu

Xử lý Ngôn ngữ Tự nhiên nói chung và xử lý tiếng Việt nói riêng, bước đầu tiên chính là tiền xử lý dữ liệu. Tiền xử lý dữ liệu là quá trình chuẩn hoá dữ liệu và loại bỏ những thành phần không cần thiết để trích xuất thông tin tốt hơn từ dữ liệu thô. Nếu bước này được thực hiện tốt, hiệu suất của mô hình sẽ được cải thiện.

Trong quá trình thu thập dữ liệu từ Tiki, có không ít các đánh giá bị rỗng hoặc trùng lặp, vì vậy chúng tôi đã tiến hành loại bỏ chúng ra khỏi tập dữ liệu. Sau đó, một quy trình tiền xử lý dữ liệu được áp dụng cho từng đánh giá trong tập dữ liệu. Các bước xử lý được mô tả như Hình 3.1.

¹<https://tiki.vn/>

²<https://open.tiki.vn/docs/docs/current/api-references/product-api/>



Hình 3.1: Quy trình tiền xử lý dữ liệu

Đầu tiên, chúng tôi loại bỏ các thành phần không cần thiết trong các đánh giá. Các thành phần này bao gồm: kí tự newline (`'\n\r'`), HTML code, biểu tượng cảm xúc, link, email, số điện thoại và các kí tự không cần thiết khác. Việc loại bỏ này giúp giảm độ phức tạp, hạn chế các thông tin gây nhiễu và không liên quan của dữ liệu.

Tiếp theo, chúng tôi tiến hành chuẩn hoá bộ kí tự thành UTF-8. Việc chuẩn hóa về UTF-8 nhằm đảm bảo tính nhất quán trong cách biểu diễn ký tự. Ngoài ra, các ký tự đặc biệt hoặc dấu trong tiếng Việt có thể bị mất hoặc hiển thị không đúng nếu không được mã hóa đúng cách. Chuẩn hóa về UTF-8 giúp bảo vệ và duy trì toàn bộ thông tin từ dữ liệu ban đầu.

Bước tiếp theo là xử lý các từ viết tắt và các thuật ngữ trong lĩnh vực sách. Các từ viết tắt thông dụng có thể kể đến như: 'ndung' (nội dung), 'sp' (sản phẩm), 'df'

(dead format, tương ứng với gãy gập, trầy xước hoặc dính nước, dính bẩn),... hoặc các thuật ngữ như ‘font’ (kiểu chữ), ‘seal’ (niêm phong),... Các mô hình NLP cần dữ liệu nhất quán để học và dự đoán chính xác. Chuẩn hóa từ viết tắt và thuật ngữ giúp mô hình hiểu rõ hơn về ngữ cảnh và tăng độ chính xác trong việc phân tích cảm xúc.

Tiếp theo, chúng tôi xử lý các từ có âm cuối được viết kéo dài (chẳng hạn như ‘xinnn’). Các từ có âm cuối kéo dài thường không tuân theo chuẩn chính tả, dẫn đến sự không nhất quán trong dữ liệu. Việc chuẩn hóa chúng giúp dữ liệu trở nên đồng nhất, từ đó cải thiện chất lượng của quá trình xử lý và phân tích dữ liệu.

Cuối cùng là bước word segmentation. Trong tiếng Việt, một từ có thể được cấu thành bởi 2 từ khác nhau (từ ghép), chẳng hạn như ‘nhà sách’ (mỗi từ có nghĩa khác nhau khi đứng một mình và khi kết hợp lại). Vì vậy, việc phân đoạn cho các đánh giá tiếng Việt là cần thiết và chúng tôi dùng VnCoreNLP toolkit [22] để thực hiện bước này.

3.1.3 Đánh nhãn dữ liệu

Đối với bài toán phân tích cảm xúc theo khía cạnh, để chuẩn bị dữ liệu, chúng tôi đã tiến hành nghiên cứu và phân loại các nhóm khía cạnh phù hợp với bài toán. Quá trình này được thực hiện trên công cụ mã nguồn mở Label Studio [23]. Sau quá trình tiền xử lý, dữ liệu được nhập vào Label Studio để đánh nhãn. Bảng 3.1 định nghĩa các khía cạnh mà một đánh giá về sản phẩm sách - truyện có thể có. Có tổng cộng 8 khía cạnh: *Nội Dung*, *Tác Giả*, *Chất Lượng và Hình Thức*, *Sản Xuất và Phân Phối*, *Đóng Gói và Vận Chuyển*, *Giá Cả và Ưu Đãi*, *Dịch Vụ Khách Hàng*, *Trải Nghiệm Tổng Thể*. Mỗi khía cạnh có thể nhận một trong ba giá trị cảm xúc: Positive, Negative hoặc Neutral. Bảng 3.2 định nghĩa các mức độ cảm xúc tương ứng với từng khía cạnh.

Khía cạnh	Định nghĩa
Nội Dung	Đánh giá đề cập đến cốt truyện và ý nghĩa được truyền tải qua câu chuyện.
Tác Giả	Đánh giá đề cập đến lối hành văn của tác giả, chất lượng của bản dịch.
Chất Lượng và Hình Thức	Đánh giá đề cập đến vẻ ngoài của sách như bìa, màu sắc, chất lượng in của bìa hoặc chất lượng của giấy sách.
Sản Xuất và Phân Phối	Đánh giá đề cập đến nhà xuất bản cũng như nền tảng bán hàng.
Đóng Gói và Vận Chuyển	Đánh giá đề cập đến thái độ của nhân viên giao hàng, tình trạng hộp, có bọc bảo vệ hay không, thời gian giao hàng.
Giá Cả và Ưu Đãi	Đánh giá đề cập đến giá thành của sản phẩm có phù hợp với chất lượng sản phẩm hay không, các khuyến mãi hoặc quà tặng được áp dụng lên sản phẩm.
Dịch Vụ Khách Hàng	Đánh giá đề cập đến thời gian phản hồi và hoàn trả hàng lỗi.
Trải Nghiệm Tổng Thể	Đánh giá đề cập đến trải nghiệm của khách hàng sau khi sử dụng sản phẩm.

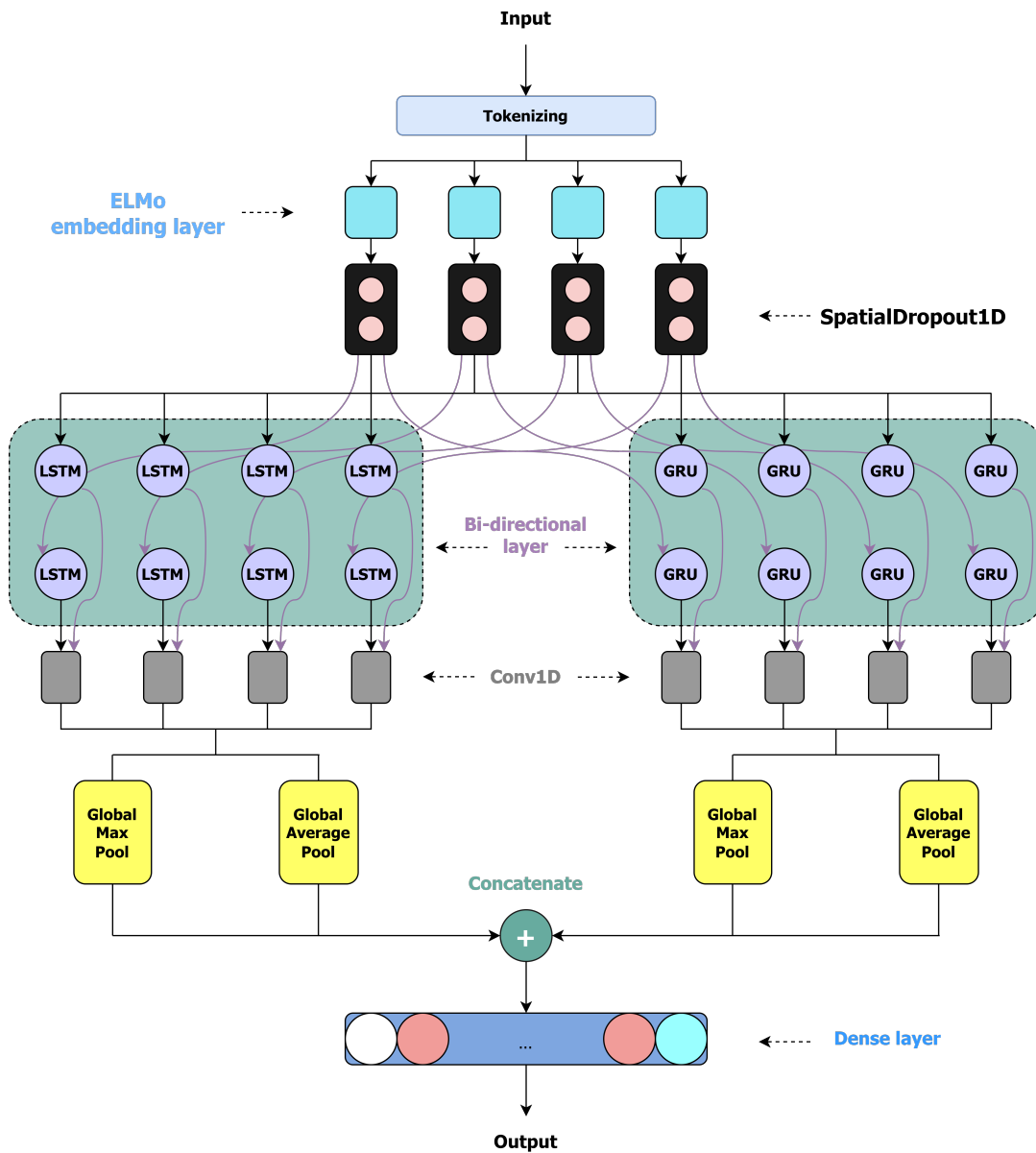
Bảng 3.1: Định nghĩa các khía cạnh

Cảm xúc	Định nghĩa
Positive	Trạng thái tâm lý khi khách hàng cảm thấy hạnh phúc, vui vẻ hoặc hài lòng về một tình huống, sự kiện hoặc trải nghiệm nào đó, thường đi kèm với cảm giác lạc quan và sự thích thú.
Negative	Trạng thái tâm lý khi khách hàng cảm thấy buồn chán, lo lắng hoặc không hài lòng về một tình huống, sự kiện hoặc trải nghiệm nào đó, thường đi kèm với cảm giác căng thẳng và bất mãn.
Neutral	Trạng thái tâm lý khi khách hàng không cảm thấy đặc biệt tích cực hay tiêu cực về một tình huống, sự kiện hoặc trải nghiệm nào đó, thường đi kèm với trạng thái không quan tâm hoặc không có cảm xúc đặc biệt.

Bảng 3.2: Định nghĩa các cảm xúc

3.2 Mô hình ELMo

Năm 2022, Quang Linh và các đồng nghiệp đã đề xuất các mô hình phân tích cảm xúc theo khía cạnh cho bộ dữ liệu đánh giá sản phẩm son môi [24] trên sàn thương mại điện tử Shopee³. Có hai phương pháp học chính được giới thiệu: (1) single-task learning và (2) multi-task learning. Trong bài luận này, chúng tôi chỉ đề cập đến phương pháp multi-task learning. Kiến trúc mô hình được mô tả như Hình 3.2.



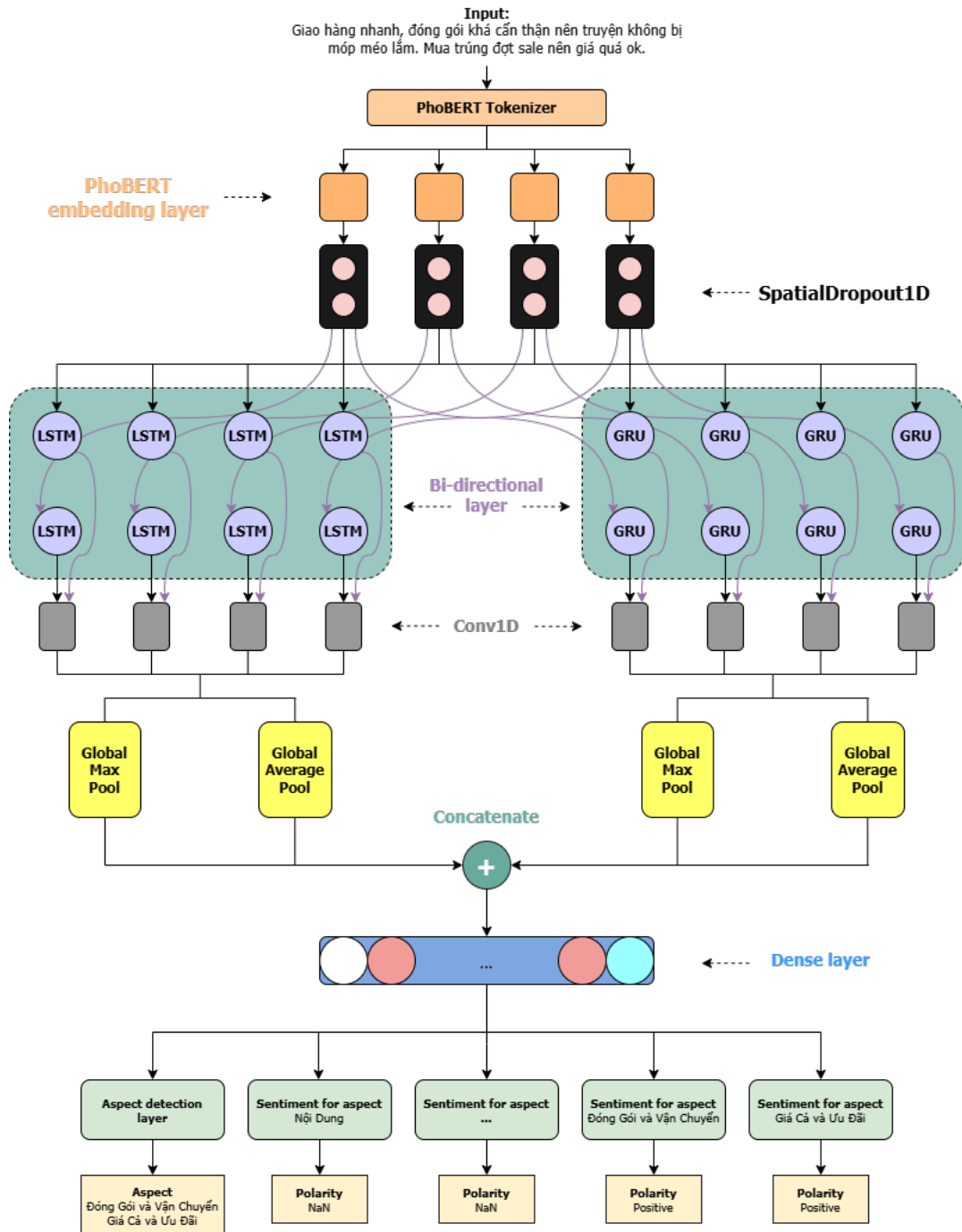
Hình 3.2: Kiến trúc mô hình sử dụng word embedding ELMo

³<https://shopee.vn/>

Sau khi tiền xử lý các đánh giá, một lớp tokenizer được sử dụng để chuyển từ thành các chỉ số dựa trên bộ từ vựng. Word embedding tiền huấn luyện ELMo [25] được sử dụng để khởi tạo cho lớp embedding và lớp này tạo ra một vector đại diện cho mỗi từ. Lớp SpatialDropout1D giúp giảm vấn đề overfitting. Một lớp Bi-LSTM [15] và Bi-GRU [26] được sử dụng song song để thu thập nhiều thông tin nhất có thể. Lớp Bi-LSTM có thể lưu trữ thông tin có giá trị trong các đánh giá và sử dụng nó để dự đoán nhãn. Sau các lớp Bi-LSTM hoặc Bi-GRU, các lớp Conv1D chuyển đổi ma trận đa chiều thành ma trận 1 chiều, GlobalMaxPooling và GlobalAveragePooling sẽ trích xuất phần tử lớn nhất của ma trận cũng như phần tử trung bình. Tất cả các lớp pooling được nối tiếp nhau và đi qua một lớp dense trước khi truyền đến các lớp đầu ra.

3.3 Mô hình PhoBERT

Phần này cung cấp thông tin về mô hình được sử dụng cho bài toán phân tích cảm xúc theo khía cạnh của chúng tôi. Lấy cảm hứng từ mô hình được đề cập trong Mục 3.2, chúng tôi đã thiết kế một mô hình phân tích cảm xúc theo khía cạnh cho tập dữ liệu về đánh giá sản phẩm sách - truyện. Điểm khác biệt ở mô hình này là lớp tokenizer và embedding được xây dựng từ mô hình tiền huấn luyện PhoBERT, sử dụng phiên bản PhoBERT_{LARGE} [17]. Kiến trúc mô hình được mô tả như Hình 3.3. Mô hình có hai thành phần chính: (1) lớp đầu vào và các lớp ẩn, (2) lớp đầu ra. Mục 3.3.1 mô tả kiến trúc mô hình đối với lớp đầu vào và các lớp ẩn. Mục 3.3.2 cung cấp thêm chi tiết về các lớp đầu ra, sử dụng phương pháp multi-task learning.



Hình 3.3: Kiến trúc mô hình sử dụng PhoBERT

3.3.1 Lớp đầu vào và các lớp ẩn

Sau khi các đánh giá được tiền xử lý, bước tiếp theo là sử dụng một lớp tokenizer để chuyển đổi các từ thành các chỉ số dựa trên bộ từ vựng đã được xây dựng trước đó. Lớp tokenizer này có nhiệm vụ chia tách văn bản thành các token và ánh xạ chúng tới các chỉ số tương ứng, dựa trên từ điển từ vựng được khởi tạo từ mô hình

tiền huấn luyện PhoBERT. Mô hình PhoBERT, một biến thể của BERT được điều chỉnh đặc biệt cho tiếng Việt, cung cấp một bộ từ vựng và các biểu diễn từ được huấn luyện trước trên một tập dữ liệu tiếng Việt phong phú.

Sau khi các từ đã được chuyển đổi thành các chỉ số, lớp embedding sẽ tạo ra các vector đại diện cho mỗi từ. Các vector này chứa thông tin ngữ nghĩa của các từ và được khởi tạo từ các trọng số đã học trước của mô hình PhoBERT. Lớp embedding giúp chuyển đổi không gian chỉ số sang không gian liên tục, nơi mà các từ có nghĩa tương tự nhau sẽ có các vector gần nhau. Điều này rất quan trọng để mô hình có thể nắm bắt được mối quan hệ ngữ nghĩa giữa các từ trong câu.

Để giảm thiểu vấn đề overfitting, một lớp SpatialDropout1D được sử dụng. Lớp này hoạt động bằng cách ngẫu nhiên bỏ qua một số phần tử trong các vector đầu vào trong suốt quá trình huấn luyện, giúp mô hình tổng quát hóa tốt hơn trên dữ liệu chưa thấy trước đó. SpatialDropout1D đặc biệt hiệu quả trong việc xử lý dữ liệu tuần tự như văn bản, nơi mà các đơn vị gần nhau có thể có mối liên hệ chặt chẽ.

Tiếp theo, một lớp Bi-LSTM [15] và một lớp Bi-GRU [26] được sử dụng song song để thu thập thông tin từ cả hai chiều của chuỗi dữ liệu. Lớp Bi-LSTM và Bi-GRU đều có khả năng nắm bắt được các phụ thuộc dài hạn trong chuỗi dữ liệu, nhưng mỗi lớp có những ưu điểm riêng. Bi-LSTM có khả năng ghi nhớ thông tin trong một khoảng thời gian dài hơn và có khả năng xử lý các vấn đề phức tạp về chuỗi thời gian, trong khi Bi-GRU nhẹ hơn và yêu cầu ít tài nguyên tính toán hơn. Việc sử dụng cả hai lớp này song song giúp mô hình khai thác tối đa thông tin từ các đánh giá.

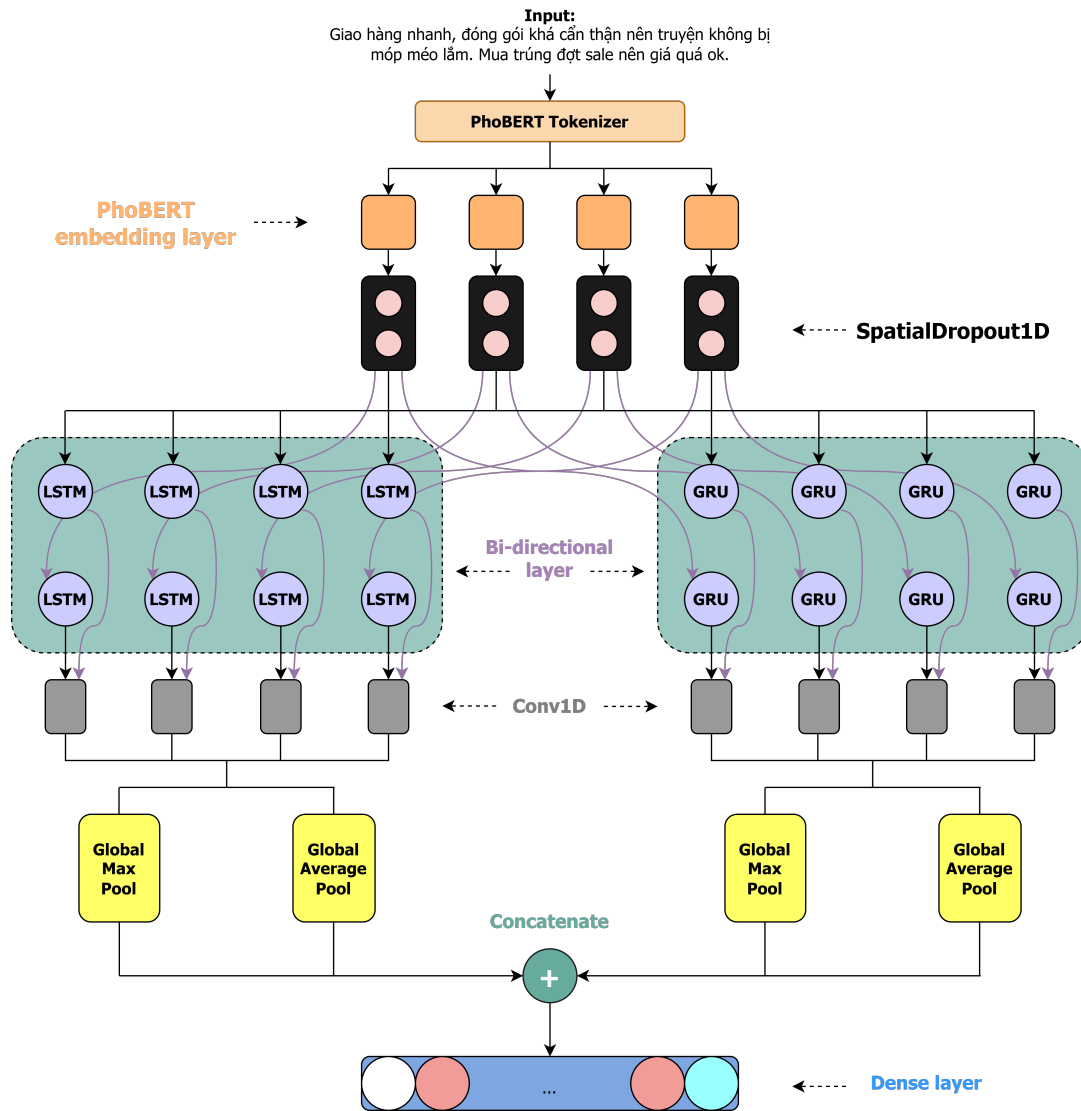
Sau khi qua các lớp Bi-LSTM và Bi-GRU, các lớp Conv1D được áp dụng để chuyển đổi ma trận đa chiều thành ma trận một chiều. Conv1D giúp trích xuất các đặc trưng cục bộ trong chuỗi dữ liệu, giúp mô hình nhận dạng được các mẫu quan trọng trong văn bản.

Tiếp theo, các lớp GlobalMaxPooling và GlobalAveragePooling được sử dụng để trích xuất thông tin quan trọng từ ma trận đầu ra của Conv1D. Lớp GlobalMaxPooling chọn ra phần tử có giá trị lớn nhất từ mỗi kênh đặc trưng, trong khi lớp

GlobalAveragePooling tính giá trị trung bình của các phần tử trong mỗi kênh đặc trưng. Kết quả của hai lớp pooling này được nối tiếp nhau, giúp mô hình nắm bắt được cả các đặc trưng cục bộ mạnh nhất và trung bình.

Cuối cùng, tất cả các đặc trưng đã trích xuất được kết hợp lại và đi qua một lớp dense trước khi truyền đến các lớp đầu ra. Lớp dense này thực hiện việc tích hợp thông tin từ các đặc trưng đã trích xuất và chuẩn bị đầu ra cho bước cuối cùng là dự đoán nhãn. Nhờ vào các bước xử lý và các lớp mô hình này, hệ thống có thể lưu trữ và sử dụng các thông tin có giá trị trong các đánh giá để đưa ra các dự đoán chính xác.

Để tìm kiến trúc mô hình tốt nhất, chúng tôi xếp các lớp lần lượt từ lớp Bi-LSTM hoặc Bi-GRU đến Bi-LSTM + Conv1D hoặc Bi-GRU + Conv1D đến phiên bản đầy đủ các lớp như Hình 3.4.

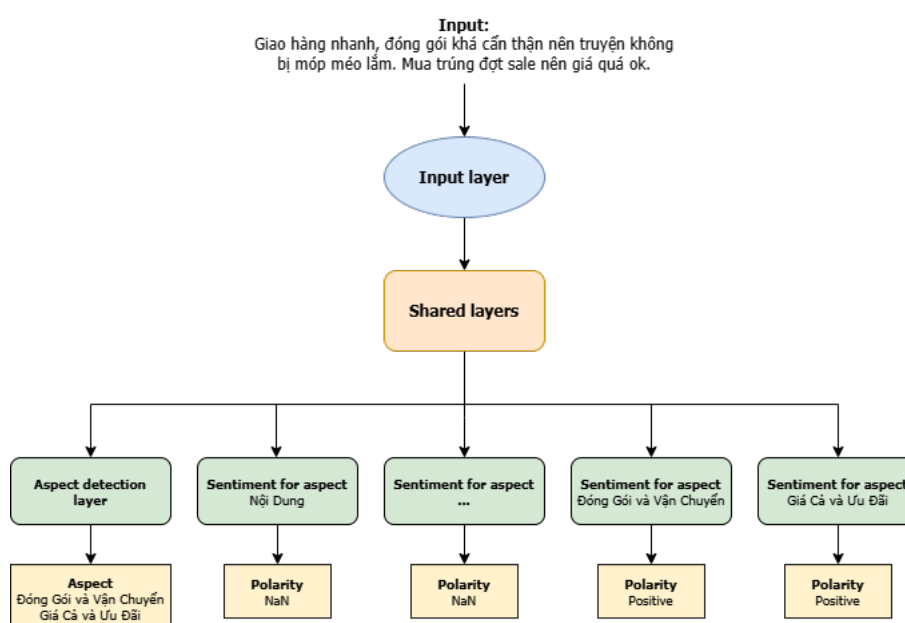


Hình 3.4: Lớp đầu vào và các lớp ẩn

3.3.2 Lớp đầu ra

Trong phương pháp học đa nhiệm (multi-task learning), các khía cạnh và cảm xúc tương ứng được dự đoán đồng thời. Phương pháp này đã được sử dụng trong nhiều nghiên cứu trước đây [24][27][28] và đã chứng minh hiệu quả của nó trong nhiệm vụ phân tích cảm xúc theo khía cạnh. Lấy cảm hứng từ các nghiên cứu này, chúng tôi thiết kế một mạng neuron hồi quy đa nhiệm để giải quyết bài toán. Lớp đầu vào và các lớp ẩn trong mô hình học đa nhiệm đã được thể hiện trong Mục 3.3.1 và chúng đóng vai trò là các lớp chia sẻ. Tuy nhiên, đối với các lớp đặc thù thực hiện các nhiệm vụ, có một tác vụ phát hiện khía cạnh và tám tác vụ phân loại cảm xúc.

Lớp đầu ra cho mô hình học đa nhiệm được minh họa như Hình 3.5. Đối với tác vụ phát hiện khía cạnh, lớp đầu ra là một lớp dense với 8 node, tương ứng với 8 khía cạnh. Sigmoid được sử dụng làm hàm kích hoạt vì tác vụ này là multi-label classification và mỗi đánh giá có thể có hơn một khía cạnh. Đối với các tác vụ phân loại cảm xúc, có 8 tác vụ tương ứng với 8 khía cạnh cần phân loại cảm xúc. Trong mỗi tác vụ, đầu ra có 4 node thể hiện Positive, Neutral, Negative hoặc NaN (không có cảm xúc cho khía cạnh này vì nó không được đề cập trong đánh giá). Softmax được sử dụng làm hàm kích hoạt cho các tác vụ này.



Hình 3.5: Lớp đầu ra sử dụng phương pháp multi-task learning

Chương 4

Thực nghiệm và đánh giá mô hình

4.1 Bộ dữ liệu thực nghiệm

4.1.1 Chuẩn bị dữ liệu

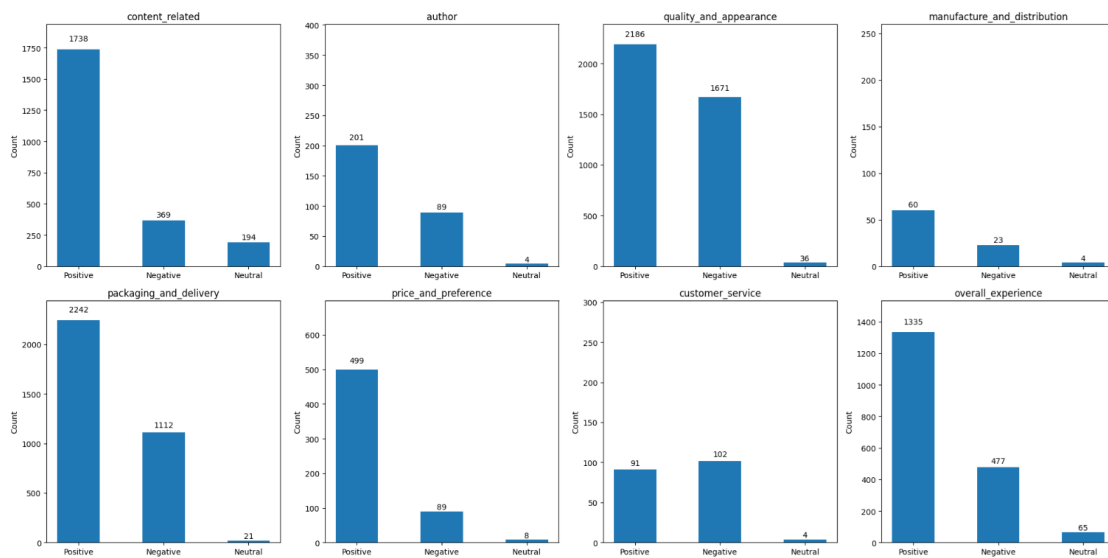
Chúng tôi đã xây dựng một tập dữ liệu tiếng Việt cho nhiệm vụ phân tích cảm xúc theo khía cạnh. Tập dữ liệu này chứa các đánh giá bằng tiếng Việt về nhiều loại sách - truyện như: sách tâm lý, sách self-help, tiểu thuyết nước ngoài, truyện tranh... được crawl từ sàn thương mại điện tử Tiki¹. Các đánh giá được thực hiện trong khoảng thời gian từ tháng 2/2024 đến tháng 4/2024. Có hai tác vụ phụ trong tập dữ liệu này: phát hiện khía cạnh và phân loại cảm xúc. Trong tác vụ phát hiện khía cạnh, chúng tôi tập trung vào việc tìm các khía cạnh được đề cập trong đánh giá. Có tổng cộng 8 khía cạnh đã được định nghĩa trong Bảng 3.1, bao gồm: Nội Dung, Tác Giả, Chất Lượng và Hình Thức, Sản Xuất và Phân Phối, Đóng Gói và Vận Chuyển, Giá Cả và Ưu Đãi, Dịch Vụ Khách Hàng, Trải Nghiệm Tổng Thể. Tác vụ phụ khác là phân loại cảm xúc Positive, Neutral hoặc Negative cho các khía cạnh này. Chúng tôi chia tập dữ liệu thành ba phần: 80% bộ dữ liệu cho tập train, 10% bộ dữ liệu cho tập validation và 10% bộ dữ liệu cho tập test. Tập train và validation sẽ giúp chúng tôi xây dựng các mô hình phát hiện và phân loại, tập test được sử dụng để đánh giá hiệu suất của các mô hình.

4.1.2 Thống kê tổng quan

Tập dữ liệu của chúng tôi chứa 6799 mẫu, bao gồm nội dung đánh giá, 8 khía cạnh và cảm xúc tương ứng với từng khía cạnh. Số lượng mẫu trong tập train, validation và test lần lượt là 5415, 692, 692. Bảng 4.1 thể hiện một số ví dụ về các đánh giá trong tập dữ liệu của chúng tôi và phân loại các khía cạnh, cảm xúc tương ứng. Hình 4.1 minh họa sự phân bố cảm xúc của mỗi khía cạnh trong tập

¹<https://tiki.vn/>

dữ liệu. Tổng thể, cảm xúc tích cực chiếm phần lớn trong tất cả các khía cạnh. Ngoài ra, số lượng đánh giá có khía cạnh Chất Lượng và Hình Thức là 3893 đánh giá, chiếm hơn 57% số đánh giá, cho thấy sự quan tâm cao của khách hàng về khía cạnh này khi mua sách. Sự mất cân bằng cao giữa các cảm xúc có thể thấy trong khía cạnh Giá Cả và Ưu Đãi. Nguyên nhân là giá trên các nền tảng thương mại điện tử thực sự cạnh tranh và rẻ hơn so với các cửa hàng vật lý, nên thường nhận được các đánh giá tốt. Trong khi đó, khía cạnh Đóng Gói và Vận Chuyển nhận được các đánh giá tốt vì đây là một khía cạnh mà người bán có thể kiểm soát và quản lý để tạo ấn tượng ban đầu tốt từ người mua. Bảng 4.2 hiển thị các thống kê tổng quan về tập dữ liệu của chúng tôi.



Hình 4.1: Sự phân bố cảm xúc của từng khía cạnh trong tập dữ liệu

Đánh giá	Phân loại
Đóng gói kỹ, sách đẹp dù bị cắn gáy và trầy 1 tẹo ở mặt trc, khổ sách mình ko ưng lắm nhưng bù lại chất lượng giấy, màu cùng các font chữ hài hòa.	Đóng Gói và Vận Chuyển: Positive, Chất Lượng và Hình Thức: Positive
Mua ngay lúc sale giá ưu đãi rất thích lại còn dc tặng thêm cuốn sách nữa. 10đ ko có nhưng. Đóng gói kỹ, mới đặt lúc sáng chiều đã giao hàng, shipper thân thiện.	Giá Cả và Ưu Đãi: Positive, Đóng Gói và Vận Chuyển: Positive
Sách mình mua chung 1 lúc 4 cuốn: Lâu đài bay của pháp sư Howl, Trà Hoa Nữ, Sự im lặng của bầy cừu và Tàn ngày để lại. 4 cuốn đều ở trong tình trạng rất tệ, cũ, gáy bị long tróc, bìa sách xuất đến thua cả sách ở tiệm sách cũ. Lại còn ko có bao ni lông bọc bên ngoài sách (cả 4 cuốn đều ở chung 1 tình trạng). Đây là lần trải nghiệm sách tệ nhất của mình ở tiki.	Chất Lượng và Hình Thức: Negative, Đóng Gói và Vận Chuyển: Negative, Trải Nghiệm Tổng Thể: Negative

Bảng 4.1: Một số ví dụ về tập dữ liệu

Set	Review	Positive	Neutral	Negative	Total sentiment
Train	5145	6649	273	3174	10096
Validation	692	856	29	379	1264
Test	692	847	34	379	1260

Bảng 4.2: Thống kê tổng quan về tập dữ liệu

4.2 Cài đặt thông số thực nghiệm

Các mô hình được huấn luyện trên Kaggle Notebook², sử dụng GPU P100. Lớp embedding có số chiều là 1024 nếu sử dụng mô hình tiền huấn luyện PhoBERT hoặc word embedding tiền huấn luyện ELMo. Số lượng đơn vị trong Bi-LSTM hoặc Bi-GRU là 400 và hàm tanh được sử dụng làm hàm kích hoạt trong các lớp này. Trong lớp Conv1D, kernel size là 2 và filter là 128, nghĩa là giảm số chiều đầu vào từ 400 xuống 128.

Chúng tôi sử dụng binary crossentropy làm hàm mất mát cho tác vụ phát hiện khía cạnh và categorical crossentropy cho các tác vụ phân loại cảm xúc. Adam [29] được sử dụng làm optimizer với learning rate là 0.0001. Batch size là 32 và

²<https://tiki.vn/>

đầu vào được padding 256 token. Số vòng lặp huấn luyện là 100. Early Stopping được sử dụng để hạn chế vấn đề overfitting.

4.3 Thước đo đánh giá

Do sự mất cân bằng về khía cạnh và cảm xúc trong tập dữ liệu thử nghiệm, F1-score được sử dụng để đánh giá hiệu suất của các mô hình nhằm xem xét sự mất cân bằng và cung cấp một góc nhìn tổng quát hơn về hiệu quả của các mô hình dự đoán. Công thức của F1-score [30] như sau với n là số lượng mẫu và TP, FP và FN lần lượt là TruePositive, FalsePositive và FalseNegative:

$$\text{Precision} = \frac{\sum_{i=1}^n TP_i}{\sum_{i=1}^n TP_i + FP_i}$$

$$\text{Recall} = \frac{\sum_{i=1}^n TP_i}{\sum_{i=1}^n TP_i + FN_i}$$

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

4.4 Kết quả thực nghiệm

Weighted average F1-score được sử dụng để đánh giá hiệu suất mô hình. $F1_{ad}$ và $F1_{sc}$ lần lượt là weighted average F1-score trong tác vụ phát hiện khía cạnh và tác vụ phân loại cảm xúc. Bảng 4.3 minh họa hiệu suất của các mô hình trên tập dữ liệu.

Nhìn chung, hầu hết mô hình sử dụng ELMo embedding cho kết quả tốt hơn mô hình sử dụng PhoBERT embedding cho cả hai tác vụ là phát hiện khía cạnh và phân loại cảm xúc. Tuy nhiên kết quả không có sự khác biệt quá lớn.

Đối với các mô hình sử dụng PhoBERT làm lớp embedding thì mô hình dùng lớp Bi-GRU + Conv1D cho kết quả tốt nhất với $F1_{ad}$ đạt 89.59% và $F1_{sc}$ đạt 87.73%. Tiếp đến là mô hình dùng lớp Bi-LSTM + Bi-GRU + Conv1D với $F1_{ad}$ đạt 88.75% và $F1_{sc}$ đạt 87.41%. Mô hình dùng lớp Bi-LSTM có $F1_{ad}$ nhỉnh hơn mô hình dùng lớp Bi-LSTM + Conv1D (88.69% và 88.42%), nhưng $F1_{sc}$ thấp hơn

(86.7% và 86.97%). Mô hình dùng lớp Bi-GRU cho kết quả thấp nhất với $F1_{ad}$ đạt 86.47% và $F1_{sc}$ đạt 85.12%.

Đối với các mô hình sử dụng ELMo làm lớp embedding thì mô hình dùng lớp Bi-GRU + Conv1D cho kết quả tốt nhất với $F1_{ad}$ đạt 90.45% và $F1_{sc}$ đạt 89.5%. Mô hình dùng lớp Bi-LSTM + Bi-GRU + Conv1D có $F1_{ad}$ nhỉnh hơn mô hình dùng lớp Bi-LSTM + Conv1D (89.97% và 89.54%), nhưng $F1_{sc}$ thấp hơn (88.82% và 88.83%). Tiếp đến là mô hình dùng lớp Bi-LSTM với $F1_{ad}$ đạt 88.29% và $F1_{sc}$ đạt 87.37%. Mô hình dùng lớp Bi-GRU cho kết quả thấp nhất với $F1_{ad}$ đạt 88,24% và $F1_{sc}$ đạt 87.19%.

Các mô hình sử dụng lớp embedding ELMo có hiệu suất nhỉnh hơn so với PhoBERT trong hầu hết các trường hợp, cho thấy ELMo embedding có thể nắm bắt các đặc điểm ngôn ngữ hiệu quả hơn trong tác vụ này. Đặc biệt là mô hình với sự kết hợp giữa ELMo embedding, Bi-GRU và Conv1D cho hiệu suất cao nhất trong tất cả các cấu hình, với $F1_{ad}$ đạt 90.45% và $F1_{sc}$ đạt 89.50%.

Việc thêm Conv1D vào Bi-LSTM hoặc Bi-GRU có xu hướng cải thiện các F1-score, cho thấy việc kết hợp các lớp recurrent với các lớp convolutional giúp mô hình nắm bắt thông tin ngữ cảnh hiệu quả hơn.

Mô hình Bi-GRU + Conv1D đặc biệt hoạt động tốt trên cả hai loại embedding, cho thấy rằng GRU kết hợp với các lớp convolutional là một phương pháp mạnh mẽ cho bài toán phân loại cảm xúc theo khía cạnh.

Embedding layer	PhoBERT		ELMo	
Metric	$F1_{ad}$	$F1_{sc}$	$F1_{ad}$	$F1_{sc}$
Bi-LSTM	88.69%	86.70%	88.29%	87.37%
Bi-LSTM + Conv1D	88.42%	86.97%	89.54%	88.83%
Bi-GRU	86.47%	85.12%	88.24%	87.19%
Bi-GRU + Conv1D	89.59%	87.73%	90.45%	89.50%
Bi-LSTM + Bi-GRU + Conv1D	88.75%	87.41%	89.97%	88.82%

Bảng 4.3: Hiệu suất của các mô hình áp dụng trên tập dữ liệu sách

Bảng 4.4 minh họa F1-score cho tác vụ phát hiện khía cạnh của cấu hình có hiệu suất tốt nhất (Bi-GRU + Conv1D) trên cả hai lớp embedding. Nhìn chung, mô

hình sử dụng ELMo embedding cho kết quả tốt hơn mô hình sử dụng PhoBERT embedding. Mô hình sử dụng PhoBERT embedding tuy cho kết quả kém hơn nhưng chênh lệch không nhiều, trong đó F1-score cho tác vụ phát hiện khía cạnh ‘*Sản Xuất và Phân Phối*’ và ‘*Dịch Vụ Khách Hàng*’ giữa hai lớp embedding là như nhau. Có thể thấy mô hình của chúng tôi có lợi thế trong việc phát hiện khía cạnh, với hầu hết các khía cạnh đều đạt F1-score hơn 84%.

Khía cạnh	PhoBERT	ELMo
Nội Dung	84.00%	86.20%
Tác Giả	94.40%	94.67%
Chất Lượng và Hình Thức	84.46%	86.82%
Sản Xuất và Phân Phối	96.98%	96.98%
Đóng Gói và Vận Chuyển	90.62%	92.35%
Giá Cả và Ưu Đãi	93.26%	89.71%
Dịch Vụ Khách Hàng	95.69%	95.69%
Trải Nghiệm Tổng thể	71.76%	74.57%

Bảng 4.4: F1-score cho tác vụ phát hiện khía cạnh (Bi-GRU + Conv1D)

Bảng 4.5 minh họa F1-score cho tác vụ phân loại cảm xúc của cấu hình có hiệu suất tốt nhất (Bi-GRU + Conv1D) trên cả hai lớp embedding. Nhìn chung, mô hình sử dụng ELMo embedding cho kết quả tốt hơn mô hình sử dụng PhoBERT embedding trong hầu hết các khía cạnh và cảm xúc tương ứng.

Đối với mô hình sử dụng PhoBERT embedding, việc phân loại cảm xúc Positive cho kết quả tốt nhất ở khía cạnh ‘*Đóng Gói và Vận Chuyển*’, với F1-score đạt 87.32%. Tiếp theo sau đó lần lượt là các khía cạnh ‘*Chất Lượng và Hình Thức*’ (77.82%), ‘*Nội Dung*’ (71.28%), ‘*Giá Cả và Ưu Đãi*’ (65.96%) và ‘*Trải Nghiệm Tổng thể*’ (46.77%). Mô hình gặp khó khăn trong việc phân loại cảm xúc Positive cho các khía cạnh ‘*Tác Giả*’, ‘*Sản Xuất và Phân Phối*’ và ‘*Dịch Vụ Khách Hàng*’, với các F1-score đều là 0%. Việc phân loại cảm xúc Negative cho kết quả tốt nhất ở khía cạnh ‘*Chất Lượng và Hình Thức*’, với F1-score đạt 75.41%. Tiếp theo sau đó lần lượt là các khía cạnh ‘*Đóng Gói và Vận Chuyển*’ (71.82%), ‘*Nội Dung*’ (45.95%) và ‘*Trải Nghiệm Tổng thể*’ (20.83%). Mô hình gặp khó khăn trong việc phân loại cảm xúc Negative cho các khía cạnh ‘*Tác Giả*’, ‘*Sản Xuất và Phân*

Phối’, *Giá Cả và Ưu Đãi*’ và *Dịch Vụ Khách Hàng*’, với các F1-score đều là 0%.

Đối với mô hình sử dụng ELMo embedding, việc phân loại cảm xúc Positive cho kết quả tốt nhất ở khía cạnh *Đóng Gói và Vận Chuyển*’, với F1-score đạt 90.23%.

Tiếp theo sau đó lần lượt là các khía cạnh *Chất Lượng và Hình Thức*’ (82.6%), *Nội Dung*’ (76.49%), *Giá Cả và Ưu Đãi*’ (71.91%), *Trải Nghiệm Tổng thể*’ (60.56%) và *Trải Nghiệm Tổng thể*’ (40%). Mô hình gặp khó khăn trong việc phân loại cảm xúc Positive cho các khía cạnh *Tác Giả*’ và *Sản Xuất và Phân Phối*’, với các F1-score đều là 0%. Việc phân loại cảm xúc Negative cho kết quả tốt nhất ở khía cạnh *Chất Lượng và Hình Thức*’, với F1-score đạt 78.66%. Tiếp theo sau đó lần lượt là các khía cạnh *Đóng Gói và Vận Chuyển*’ (73.83%), *Nội Dung*’ (58.82%), *Trải Nghiệm Tổng thể*’ (40.54%), *Dịch Vụ Khách Hàng*’ (28.57%), *Tác Giả*’ (22.22%) và *Giá Cả và Ưu Đãi*’ (13.33%). Mô hình gặp khó khăn trong việc phân loại cảm xúc Negative cho khía cạnh *Sản Xuất và Phân Phối*’, với F1-score là 0%.

Cả hai mô hình đều gặp khó khăn lớn trong việc phân loại cảm xúc Neutral, với F1-score bằng 0% ở tất cả các khía cạnh. Lý do là vì số lượng đánh giá tích cực và tiêu cực cho sản phẩm sách - truyện trên sàn thương mại điện tử Tiki nhiều hơn hẳn so với số lượng đánh giá trung lập. Đối với tác vụ phân loại cảm xúc, kết quả ở Bảng 4.5 cho thấy rằng mô hình phân loại cảm xúc Positive tốt hơn các cảm xúc khác. Các khía cạnh như *Nội Dung*’, *Chất Lượng và Hình Thức*’ và *Đóng Gói và Vận Chuyển*’ có sự biến động trong các F1-score, đặc biệt là với cảm xúc Positive và Negative bởi vì các khía cạnh này được đề cập rất nhiều trong các đánh giá. Tuy nhiên, mô hình gặp khó khăn trong việc phân loại cảm xúc cho các khía cạnh *Tác Giả*’ và *Sản Xuất và Phân Phối*’ vì số lượng đánh giá đề cập đến các khía cạnh này là không nhiều so với các khía cạnh khác. Sự mất cân bằng của bộ dữ liệu là một trong những thách thức lớn và sẽ được giải quyết trong tương lai.

Embedding layer	PhoBERT			ELMo		
Cảm xúc	Positive	Negative	Neutral	Positive	Negative	Neutral
Nội Dung	71.28%	45.95%	0.00%	76.49%	58.82%	0.00%
Tác Giả	0.00%	0.00%	0.00%	0.00%	22.22%	0.00%
Chất Lượng và Hình Thức	77.82%	75.41%	0.00%	82.60%	78.66%	0.00%
Sản Xuất và Phân Phối	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Đóng Gói và Vận Chuyển	87.32%	71.82%	0.00%	90.23%	73.83%	0.00%
Giá Cả và Ưu Đãi	65.96%	0.00%	0.00%	71.91%	13.33%	0.00%
Dịch Vụ Khách Hàng	0.00%	0.00%	0.00%	40.00%	28.57%	0.00%
Trải Nghiệm Tổng Thể	46.77%	20.83%	0.00%	60.56%	40.54%	0.00%

Bảng 4.5: F1-score cho tác vụ phân loại cảm xúc (Bi-GRU + Conv1D)

Chương 5

Kết luận và mở rộng

Bài luận này đã xử lý bài toán phân tích cảm xúc theo khía cạnh của các đánh giá về sản phẩm sách - truyện trên các trang web thương mại điện tử. Trong bài luận này, chúng tôi đã xây dựng một tập dữ liệu mới chứa 5799 đánh giá về sách trên các nền tảng thương mại điện tử cho nhiệm vụ phân tích cảm xúc dựa trên khía cạnh. Có 12620 cặp khía cạnh và cảm xúc trong tập dữ liệu này. Về kiến trúc mô hình, sự kết hợp giữa ELMo embedding layer, Bi-GRU và Conv1D vượt trội hơn so với các cấu hình khác, với F1-score tốt nhất đạt 90.45% cho tác vụ phát hiện khía cạnh và 89.50% cho tác vụ phân loại cảm xúc.

Trong tương lai, chúng tôi đang xem xét xây dựng một hệ thống tự động thu thập đánh giá trên các nền tảng thương mại điện tử, xử lý các đánh giá, dự đoán khía cạnh, cảm xúc của các đánh giá và hiển thị kết quả trên một dashboard. Hệ thống này sẽ mang lại cái nhìn tổng quan cho người bán, giúp họ hiểu rõ hơn về sản phẩm và giúp người mua dễ dàng đưa ra quyết định trước khi mua hàng. Ngoài ra, chúng tôi sẽ áp dụng một số mô hình hiện đại như Large Language Model để cải thiện độ chính xác của bài toán phân loại cảm xúc theo khía cạnh.

Tài liệu tham khảo

- [1] BK Bhavitha, Anisha P Rodrigues, and Niranjana N Chiplunkar. “Comparative study of machine learning techniques in sentimental analysis”. In: *2017 International conference on inventive communication and computational technologies (ICICCT)*. IEEE. 2017, pp. 216–221.
- [2] Paul Denny, Katerina Taskova, Jörg Wicker, et al. “A Systematic Review of Aspect-based Sentiment Analysis (ABSA): Domains, Methods, and Trends”. In: *arXiv preprint arXiv:2311.10777* (2023).
- [3] Huyen TM Nguyen, Hung V Nguyen, Quyen T Ngo, Luong X Vu, Vu Mai Tran, Bach X Ngo, and Cuong A Le. “VLSP shared task: sentiment analysis”. In: *Journal of Computer Science and Cybernetics* 34.4 (2018), pp. 295–310.
- [4] Kim Thi-Thanh Nguyen, Sieu Khai Huynh, Luong Luc Phan, Phuc Huynh Pham, Duc-Vu Nguyen, and Kiet Van Nguyen. “Span detection for aspect-based sentiment analysis in Vietnamese”. In: *arXiv preprint arXiv:2110.07833* (2021).
- [5] Kiet Van Nguyen, Vu Duc Nguyen, Phu XV Nguyen, Tham TH Truong, and Ngan Luu-Thuy Nguyen. “UIT-VSFC: Vietnamese students’ feedback corpus for sentiment analysis”. In: *2018 10th international conference on knowledge and systems engineering (KSE)*. IEEE. 2018, pp. 19–24.
- [6] Tun Thura Thet, Jin-Cheon Na, and Christopher SG Khoo. “Aspect-based sentiment analysis of movie reviews on discussion boards”. In: *Journal of information science* 36.6 (2010), pp. 823–848.
- [7] Hai Ha Do, Penatiyana WC Prasad, Angelika Maag, and Abeer Alsadoon. “Deep learning for aspect-based sentiment analysis: a comparative review”. In: *Expert systems with applications* 118 (2019), pp. 272–299.
- [8] Oanh Thi Tran and Viet The Bui. “A bert-based hierarchical model for vietnamese aspect based sentiment analysis”. In: *2020 12th International Conference on Knowledge and Systems Engineering (KSE)*. IEEE. 2020, pp. 269–274.
- [9] Dang Van Thin, Duc-Vu Nguyen, Kiet Van Nguyen, Ngan Luu-Thuy Nguyen, and Anh Hoang-Tu Nguyen. “Multi-task learning for aspect and polarity recognition on vietnamese datasets”. In: *International Conference of the Pacific Association for Computational Linguistics*. Springer. 2019, pp. 169–180.
- [10] Yang Li and Tao Yang. “Word embedding for understanding natural language: a survey”. In: *Guide to big data applications* (2018), pp. 83–104.
- [11] Kenneth Ward Church. “Word2Vec”. In: *Natural Language Engineering* 23.1 (2017), pp. 155–162.
- [12] Prafulla Bafna, Dhanya Pramod, and Anagha Vaidya. “Document clustering: TF-IDF approach”. In: *2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*. IEEE. 2016, pp. 61–66.

- [13] Farhad Morteza pour Shiri, Thinagaran Perumal, Norwati Mustapha, and Raihani Mohamed. “A comprehensive overview and comparative analysis on deep learning models: CNN, RNN, LSTM, GRU”. In: *arXiv preprint arXiv:2305.17473* (2023).
- [14] Christopher Olah et al. “Understanding lstm networks”. In: (2015).
- [15] Sepp Hochreiter and Jürgen Schmidhuber. “Long short-term memory”. In: *Neural computation* 9.8 (1997), pp. 1735–1780.
- [16] Xiaolin Hu and P Balasubramaniam. *Recurrent neural networks*. Vol. 400. InTech, 2008.
- [17] Dat Quoc Nguyen and Anh Tuan Nguyen. “PhoBERT: Pre-trained language models for Vietnamese”. In: *arXiv preprint arXiv:2003.00744* (2020).
- [18] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. “Bert: Pre-training of deep bidirectional transformers for language understanding”. In: *arXiv preprint arXiv:1810.04805* (2018).
- [19] Matt Gardner, Joel Grus, Mark Neumann, Oyvind Tafjord, Pradeep Dasigi, Nelson Liu, Matthew Peters, Michael Schmitz, and Luke Zettlemoyer. “Allennlp: A deep semantic natural language processing platform”. In: *arXiv preprint arXiv:1803.07640* (2018).
- [20] Jeffrey Pennington, Richard Socher, and Christopher D Manning. “Glove: Global vectors for word representation”. In: *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*. 2014, pp. 1532–1543.
- [21] Ayat Abodayeh, Reem Hejazi, Ward Najjar, Leena Shihadeh, and Rabia Latif. “Web Scraping for Data Analytics: A BeautifulSoup Implementation”. In: *2023 Sixth International Conference of Women in Data Science at Prince Sultan University (WiDS PSU)*. IEEE. 2023, pp. 65–69.
- [22] Thanh Vu, Dat Quoc Nguyen, Dai Quoc Nguyen, Mark Dras, and Mark Johnson. “VnCoreNLP: A Vietnamese natural language processing toolkit”. In: *arXiv preprint arXiv:1801.01331* (2018).
- [23] Maxim Tkachenko, Mikhail Malyuk, Andrey Holmanyuk, and Nikolai Liubimov. “Label studio: Data labeling software”. In: *Open source software available from <https://github.com/heartexlabs/label-studio>* 2022 (2020).
- [24] Quang-Linh Tran, Phan Thanh Dat Le, and Trong-Hop Do. “Aspect-based sentiment analysis for Vietnamese reviews about beauty product on E-commerce websites”. In: *Proceedings of the 36th Pacific Asia Conference on Language, Information and Computation*. 2022, pp. 767–776.
- [25] Xuan-Son Vu, Thanh Vu, Son N Tran, and Lili Jiang. “Etnlp: A visual-aided systematic approach to select pre-trained embeddings for a downstream task”. In: *arXiv preprint arXiv:1903.04433* (2019).
- [26] Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. “Empirical evaluation of gated recurrent neural networks on sequence modeling”. In: *arXiv preprint arXiv:1412.3555* (2014).
- [27] Ruidan He, Wee Sun Lee, Hwee Tou Ng, and Daniel Dahlmeier. “An interactive multi-task learning network for end-to-end aspect-based sentiment analysis”. In: *arXiv preprint arXiv:1906.06906* (2019).
- [28] Heng Yang, Biqing Zeng, Jianhao Yang, Youwei Song, and Ruyang Xu. “A multi-task learning model for chinese-oriented aspect polarity classification and aspect term extraction”. In: *Neuro-computing* 419 (2021), pp. 344–356.

- [29] Diederik P Kingma and Jimmy Ba. “Adam: A method for stochastic optimization”. In: *arXiv preprint arXiv:1412.6980* (2014).
- [30] Marina Sokolova and Guy Lapalme. “A systematic analysis of performance measures for classification tasks”. In: *Information processing & management* 45.4 (2009), pp. 427–437.