

# Nhận diện chứng cứ trong phân tích cảm xúc dựa trên các khía cạnh tiếng Việt

Trần Triệu Vũ<sup>1,2</sup>, Phạm Đức Thế<sup>1,2</sup>, Nguyễn Văn Hữu Nghĩa<sup>1,2</sup>,  
Đỗ Trọng Hợp<sup>1,3</sup>, Lưu Thanh Sơn<sup>1,3</sup>, Nguyễn Thành Luân<sup>1,3</sup>

<sup>1</sup>Đại Học Công Nghệ Thông Tin – Đại Học Quốc Gia TP.HCM, Việt Nam

<sup>2</sup>{19522539,19522253,19521900}@gm.uit.edu.vn

<sup>3</sup>{hopdt,sonlt,luannt}@uit.edu.vn

## Tóm Tắt Nội Dung

Phân tích cảm xúc dựa trên khía cạnh là một bài toán có vai trò quan trọng trong xử lý ngôn ngữ tự nhiên và trí tuệ nhân tạo. Gần đây, các nhà nghiên cứu chỉ tập trung vào phát hiện khía cạnh và phân loại cảm xúc mà bỏ qua tác vụ phụ là phát hiện những chứng cứ về quan điểm của người dùng, có tiềm năng to lớn trong các ứng dụng thực tế. Trong báo cáo này, chúng em sẽ tìm hiểu và trình bày cách cài đặt một mô hình BiLSTM-CRF cho bài toán “Nhận diện chứng cứ trong phân tích cảm xúc dựa trên các khía cạnh tiếng Việt” trên bộ dữ liệu UIT-ViSD4SA bao gồm 35,396 chứng cứ đã được gán nhãn trên 11,122 bình luận phản hồi của người dùng về các sản phẩm điện thoại thông minh trên các trang thương mại điện tử. Mô hình Bidirectional Long Short-Term Memory (BiLSTM) với Conditional Random Field (CRF) (BiLSTM-CRF) cho tác vụ phát hiện chứng cứ trong phân tích cảm xúc theo khía cạnh trên tiếng Việt. Kết quả tốt nhất theo độ đo  $F1_{macro}$  là 54.07% sử dụng BiLSTM-CRF kết hợp với syllable embedding và character embedding (CharLSTM). Bộ dữ liệu được nhóm tác giả chia sẻ công khai tại: <https://github.com/kimkim00/UIT-ViSD4SA> để phục vụ cho mục đích nghiên cứu.

## 1 Giới Thiệu

Thông thường, trước khi mua một món hàng hoặc quyết định sử dụng một dịch vụ, mọi người thường có xu hướng tìm kiếm lời khuyên từ những người đi trước đã mua hàng hoặc sử dụng dịch vụ đó. Với sự phát triển nhanh chóng của internet, ngày càng có nhiều người tìm đến những lời khuyên từ các website, trang thương mại điện tử, diễn đàn hay các kênh đánh giá sản phẩm. Các đánh giá của người dùng có sẵn trên các nền tảng thương mại điện tử đang trở thành nguồn tài nguyên quý giá cho cả khách hàng và nhà sản xuất. Đối với khách hàng,

nguồn dữ liệu này cung cấp thông tin về sản phẩm và lời khuyên hữu ích giúp họ tránh mua phải sản phẩm hoặc đăng ký dịch vụ không phù hợp với nhu cầu cá nhân. Mặt khác, đánh giá của người dùng cũng là thông tin quý giá đối với doanh nghiệp và nếu được sử dụng đúng cách và hiệu quả, dữ liệu này có thể giúp cho doanh nghiệp định hướng nâng cao chất lượng sản phẩm, xác định chính xác đối tượng khách hàng cho từng phân khúc.

*Phân tích cảm xúc dựa trên khía cạnh (Aspect-based sentiment analysis – ABSA)* (Hu and Liu, 2004) về phản hồi của người dùng là một tác vụ đầy thách thức thu hút nhiều sự quan tâm của cả giới học thuật lẫn các doanh nghiệp (Jo and Oh, 2011; Kiritchenko et al., 2014; Chen et al., 2017). Đầu vào là một phản hồi cụ thể về một sản phẩm hoặc dịch vụ, nhiệm vụ chính của ABSA là phát hiện những gì đang được thảo luận, sau đó đưa ra phân tích cảm xúc cho khía cạnh đã phát hiện đó. Vấn đề của ABSA có thể được chia thành ba tác vụ cơ bản như sau: *phát hiện khía cạnh (aspect detection)*, *diễn tả mục đích của ý kiến (opinion target expression – OTE)*, *trạng thái cảm xúc (sentiment polarity)*. Trong báo cáo này, chúng em tập trung vào việc phát hiện ý kiến của người dùng dựa trên các khía cạnh và cảm xúc của họ, mà chúng em gọi là nhận diện chứng cứ cho ABSA. Cụ thể, khi một đánh giá được đưa ra “*Mặc dù điện thoại rất tốt nhưng giá quá đắt!*”, việc phát hiện chứng cứ cho tác vụ ABSA nhằm trích xuất ra hai ý kiến “*điện thoại rất tốt*” và “*giá quá đắt*”, sau đó phân loại chúng dựa trên các khía cạnh phù hợp và các trạng thái cảm xúc tương ứng. Tác vụ được mô tả như sau:

- **Input:** Phản hồi của khách hàng  $C$  về điện thoại thông minh có  $n$  ký tự.
- **Output:** Một hoặc nhiều *chứng cứ (span)* được trích xuất trực tiếp từ phản hồi của khách hàng  $C$  với khía cạnh của nó. Mỗi

chứng cứ được trích từ vị trí  $i$  đến vị trí  $j$  sao  
 $0 \leq i, j \leq n$  và  $i \leq j$ .

*Giao diện người dùng (User interface)* đóng góp một phần quan trọng trong trải nghiệm mua sắm trên các nền tảng thương mại điện tử. Giao diện người dùng của các trang thương mại điện tử trở nên thuận tiện hơn bao giờ hết với sự trợ giúp của các kỹ thuật ABSA. Nếu một trang thương mại điện tử áp dụng ABSA cho nền tảng của họ, khách hàng có thể tập trung vào các bài đánh giá về khía cạnh mà họ quan tâm một cách hiệu quả nhất. Bên cạnh đó, chủ sở hữu trang thương mại điện tử có thể theo dõi chất lượng sản phẩm và dịch vụ của họ với sự trợ giúp của ABSA. Một số nền tảng thương mại điện tử của Trung Quốc như Taobao, Dianping triển khai giao diện người dùng dựa trên ABSA để cải thiện trải nghiệm người dùng. Do đó, tiềm năng và tầm quan trọng của các kỹ thuật ABSA đối với lĩnh vực này là vô cùng lớn. Mặt khác, các trang thương mại điện tử có mặt tại Việt Nam vẫn còn non kém trong việc cung cấp thông tin phản hồi cho người dùng. Hầu hết các nền tảng thương mại điện tử tại Việt Nam đều cung cấp hệ thống phản hồi đơn giản: người dùng để lại bình luận trên hệ thống cùng với hệ thống đánh giá 5 sao như trong Hình 1. Các nền tảng có hệ thống này bao gồm thegioididong<sup>1</sup>, fptshop<sup>2</sup>, shopee<sup>3</sup>, tiki<sup>4</sup> và lazada<sup>5</sup>. Khác với phần còn lại, foody<sup>6</sup> (một nền tảng đánh giá nhà hàng) cho phép người dùng phản hồi trên thang điểm 10 và cung cấp số điểm đó trên một số khía cạnh cụ thể (vị trí, giá cả, chất lượng, dịch vụ và không gian). Do đó, chúng em tập trung vào nhận diện chứng cứ cho bài toán ABSA, không chỉ nhận diện các *khía cạnh (aspect)* và *trạng thái cảm xúc (polarity)* của chúng mà còn nhận diện các chứng cứ của quan điểm đó.

Phần còn lại của báo cáo được tổ chức như sau. Trong Phần 2, chúng em trình bày về các công trình liên quan. Trong Phần 3, chúng em trình bày bộ dữ liệu sẽ sử dụng để giải quyết bài toán. Hướng tiếp cận bài toán được mô tả chi tiết trong Phần 4. Trong Phần 5, chúng em triển khai mô hình BiLSTM-CRF để giải quyết vấn đề và phân tích để tìm ra điểm yếu của phương pháp trên tập dữ liệu của các tác giả. Cuối cùng, Phần 6 là rút ra kết luận.

<sup>1</sup><https://www.thegioididong.com/>

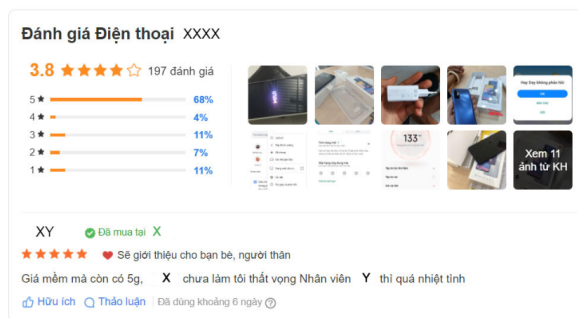
<sup>2</sup><https://fptshop.com.vn/>

<sup>3</sup><https://shopee.vn/>

<sup>4</sup><https://tiki.vn/>

<sup>5</sup><https://www.lazada.vn/>

<sup>6</sup><https://www.foody.vn>



Hình 1: Phản hồi từ một trang thương mại điện tử tại Việt Nam

## 2 Công Trình Liên Quan

Các bộ dữ liệu SemEval bao gồm các đánh giá của người dùng từ các trang thương mại điện tử được thúc đẩy cho nghiên cứu có liên quan nhiều đến ABSA (Li et al., 2019; Luo et al., 2020; Chen and Qian, 2020). Bộ dữ liệu SemEval-2014 task 4 (SE-ABSA14) (Pontiki et al., 2014) bao gồm các đánh giá về nhà hàng và máy tính xách tay. Bộ dữ liệu về nhà hàng bao gồm năm danh mục khía cạnh (Food, Service, Price, Ambience và Anecdotes/Miscellaneous) và bốn nhãn trạng thái cảm xúc (Positive, Negative, Conflict và Neutral). Bộ dữ liệu về máy tính xách tay chỉ được gán nhãn để thực hiện tác vụ nhận diện các khía cạnh và phân loại trạng thái cảm xúc. Bộ dữ liệu SemEval-2015 Task 12 (SE-ABSA15) (Pontiki et al., 2015) được xây dựng dựa trên SE-ABSA14. SE-ABSA15 mô tả danh mục khía cạnh của nó như một loại thực thể được kết hợp với một loại thuộc tính (ví dụ: Food#Style Options) và loại bỏ nhãn có trạng thái cảm xúc là Conflict. Bộ dữ liệu SemEval-2016 task-5 (SE-ABSA16) (Pontiki et al., 2016) đã mở rộng SE-ABSA15 cho các lĩnh vực mới như Khách sạn (Hotels), Điện tử gia dụng (Consumer Electronics), Viễn thông (Telecom), Bảo tàng (Museums) và các ngôn ngữ khác (Hà Lan, Pháp, Nga, Tây Ban Nha, Tiếng Thổ Nhĩ Kỳ và tiếng Ả Rập).

So với sự thịnh vượng của các ngôn ngữ có tài nguyên phong phú như tiếng Anh, tiếng Trung hoặc tiếng Tây Ban Nha, số lượng bộ dữ liệu chất lượng cao của Việt Nam là rất thấp. Năm 2018, tác vụ ABSA đầu tiên bằng tiếng Việt được thêm vào cuộc thi xử lý ngôn ngữ tự nhiên nổi tiếng được tổ chức bởi cộng đồng Xử lý Ngôn ngữ và Giọng nói tiếng Việt (Vietnamese Language and Speech Processing – VLSP) (Nguyen et al., 2018). VLSP đã cung cấp một bộ dữ liệu ABSA bao gồm các đánh giá về khách sạn và nhà hàng. Thật không may,

Input	Output
<p>Máy đẹp, sang <small>DESIGN#POSITIVE</small>, sd thì rất là ok <small>GENERAL#POSITIVE</small> máy <small>BATTERY#POSITIVE</small></p> <p>mượt <small>PERFORMANCE#POSITIVE</small>. Pin sd cũng rất lâu mới hết, nhiều khi cả ngày và qua luôn ngày hôm sau mới sạc, sạc rất nhanh khoảng chừng 1 tiếng 5 phút là đầy rồi, ko lâu như iPhones mất gần 3 đến 4 tiếng đầy <small>BATTERY#POSITIVE</small>. Chỉ sd để lướt web, facebook, youtube. Nghe nhạc rất hay đặc biệt là nghe bằng tai nghe AKG. <small>GENERAL#POSITIVE</small>. Rất xứng đáng với số tiền bỏ ra <small>GENERAL#POSITIVE</small>.</p>	<p>0, 13, "DESIGN#POSITIVE" 15, 31, "GENERAL#POSITIVE" 32, 40, "PERFORMANCE#POSITIVE" 42, 175, "BATTERY#POSITIVE" 315, 346, "GENERAL#POSITIVE"</p>

Bảng 1: Ví dụ minh họa các chứng cứ để phân tích cảm xúc dựa trên các khía cạnh bằng tiếng Việt

bộ dữ liệu VLSP lấy cảm hứng từ SE-ABSA15 chỉ được gán nhãn entity#attribute cho danh mục khía cạnh và cảm xúc của nó mà bỏ qua phần *trích xuất mục đích của ý kiến (Opinion Target Extraction)*. Nguyen và cộng sự (Nguyen et al., 2019) đã đề xuất bộ dữ liệu trên các lĩnh vực tương tự như nhà hàng và khách sạn, chỉ bao gồm 7,828 đánh giá ở cấp tài liệu (document-level) với bảy khía cạnh kết hợp với năm trạng thái cảm xúc cho hai tác vụ. Dang và cộng sự (Van Thin et al., 2021) cũng đã xây dựng một bộ dữ liệu cho cùng lĩnh vực như hai công trình trước được gán nhãn với độ đồng thuận cao giữa các người gán nhãn (annotator) ở cấp độ câu (sentence level). Mai và cộng sự (Mai and Le, 2018) đã thu thập và gán nhãn bộ dữ liệu ABSA tiếng Việt chỉ gồm 2,098 câu cho hai tác vụ: trích xuất mục đích của ý kiến và nhận diện trạng thái cảm xúc cho lĩnh vực điện thoại thông minh. Họ đã trình bày một mô hình đa tác vụ cho hai tác vụ bằng cách sử dụng cơ chế gán nhãn chuỗi được liên kết với mạng neuron hồi quy hai chiều (bidirectional recurrent neural networks – BRNN) và trường điều kiện ngẫu nhiên (conditional random field – CRF). Để đánh giá phân tích cảm xúc dựa trên khía cạnh cho thương mại điện tử di động (mobile e-commerce), Phan và cộng sự (Phan et al., 2021) đã tạo ra bộ dữ liệu điểm chuẩn (benchmark dataset) (UIT-ViSFD) với 11,122 nhận xét dựa trên cơ chế gán nhãn chặt chẽ. Hơn nữa, họ đã phát triển một hệ thống nghiên cứu dư luận xã hội (social listening) bằng tiếng Việt dựa trên ABSA. Để nhận diện chứng cứ trong ABSA, Hu và cộng sự (Hu et al., 2019) đã đề xuất một framework trích xuất chứng cứ sau đó mới phân loại, trong đó nhiều chứng cứ được trích xuất trực tiếp từ câu dựa trên các khoảng ký tự có trạng thái cảm xúc tương ứng, cuối cùng được phân loại bằng các biểu diễn chứng cứ của chúng. Công việc này được lấy cảm hứng từ những tiến bộ trong khả năng máy

hiểu và trả lời câu hỏi (Seo et al., 2016; Xu et al., 2018), tác vụ ở đây là trích xuất một khoảng văn bản liên tục từ tài liệu làm câu trả lời cho câu hỏi (Rajpurkar et al., 2016; Nguyen et al., 2020). Nguyen và cộng sự (Nguyen et al., 2021) đã xây dựng bộ dữ liệu UIT-ViSD4SA với 11,122 bình luận của người dùng về các sản phẩm điện thoại thông minh, các bình luận được gán khoảng ký tự cùng nhãn khía cạnh, trạng thái cảm xúc tương ứng phục vụ giải quyết bài toán nhận diện chứng cứ. Nhóm tác giả sử dụng mô hình BiLSTM-CRF với sự kết hợp của các embedding về âm tiết, ký tự và ngữ cảnh (từ XLM-RoBERTa), kết quả tốt nhất mà họ đạt được là 62.76%  $F1_{macro}$  cho tác vụ nhận diện aspect, 49.77%  $F1_{macro}$  cho tác vụ nhận diện polarity, 45.70%  $F1_{macro}$  cho tác vụ nhận diện aspect#polarity.

### 3 Bộ Dữ Liệu

Bộ dữ liệu được sử dụng trong báo cáo của chúng em có tên là UIT-ViSD4SA, bộ dữ liệu là các phản hồi về điện thoại thông minh bằng tiếng Việt chuẩn cho ABSA và nhận diện chứng cứ. Tất cả các phản hồi trong UIT-ViSD4SA được thu thập từ một nền tảng thương mại điện tử. Có 11,122 bình luận của người dùng và mỗi bình luận được gán nhãn thủ công các chứng cứ theo các khoảng ký tự của nó với nhãn là một trong mười khía cạnh và một trong ba trạng thái cảm xúc tương ứng. Bảng 1 cho thấy một điểm dữ liệu minh họa đã được chú thích.

Theo các hướng dẫn gán nhãn do Phan và cộng sự đề xuất (Phan et al., 2021), các tác giả thêm một số định nghĩa và quy tắc làm cơ sở cho việc xây dựng dữ liệu. Các tác giả sử dụng lại danh sách mười khía cạnh được định nghĩa trước như trong Bảng 2, với mỗi khía cạnh được đề cập trong bài đánh giá sẽ được gán một trạng thái cảm xúc tương ứng là Positive, Neutral hoặc Negative. Chứng cứ được định nghĩa là khoảng ngắn nhất chứa ý kiến



Aspect	Định nghĩa
SCREEN	Nhận xét của người dùng về chất lượng màn hình, kích thước, màu sắc và công nghệ hiển thị.
CAMERA	Các nhận xét đề cập đến chất lượng của máy ảnh, độ rung, độ trễ, lấy nét và màu sắc hình ảnh.
FEATURES	Trải nghiệm của người dùng liên quan tới các tính năng như cảm biến vân tay, kết nối wifi, cảm ứng và nhận diện khuôn mặt của điện thoại.
BATTERY	Các nhận xét mô tả dung lượng pin và chất lượng pin.
PERFORMANCE	Những đánh giá mô tả chip xử lý, hiệu suất sử dụng và độ mượt mà của điện thoại.
STORAGE	Các ý kiến đề cập đến khả năng lưu trữ, khả năng mở rộng dung lượng qua thẻ nhớ.
DESIGN	Các bài đánh giá đề cập đến kiểu dáng, thiết kế và vỏ.
PRICE	Các bình luận trình bày giá cụ thể của điện thoại.
GENERAL	Các đánh giá của khách hàng nhận xét chung về điện thoại.
SER&ACC	Các ý kiến đề cập đến dịch vụ bán hàng, bảo hành và đánh giá các phụ kiện của điện thoại.

Bảng 2: Danh sách mười khía cạnh và định nghĩa của chúng

của người dùng về một khía cạnh nào đó đã được định nghĩa trước. Với mỗi bình luận, người gán nhãn được yêu cầu đưa ra các khoảng ký tự của các chứng cứ có trong bình luận đó và gán nhãn khía cạnh, trạng thái cảm xúc tương ứng cho chứng cứ đó. Giả sử ta có một bình luận, khi một chứng cứ được nhận diện trong bình luận đó một cách đủ rõ ràng hay ẩn ý, khía cạnh có trạng thái cảm xúc của chứng cứ đó được gán nhãn là *khía cạnh#trạng thái cảm xúc (aspect#polarity)* như trong Bảng 1.

Hình 2 trình bày sự phân bố của mười danh mục khía cạnh trong tập dữ liệu UIT-ViSD4SA. Mọi người có xu hướng đánh giá tổng quan một chiếc điện thoại thông minh, với 22,76% bài đánh giá đề cập đến GENERAL. Người dùng thường quan tâm đến các khía cạnh liên quan đến nhu cầu của họ hơn, chẳng hạn như PERFORMANCE, BATTERY, FEATURES và CAMERA.

Thông kê về tập dữ liệu UIT-ViSD4SA được trình bày trong Bảng 3. Tập dữ liệu UIT-ViSD4SA bao gồm 35,396 chứng cứ trên 11,122 bình luận. Qua phân tích của các tác giả, tập dữ liệu có sự phân bố không đồng đều các nhãn cảm xúc. Nhãn Positive chiếm số lượng nhiều nhất, tiếp theo là Negative và Neutral. Trung bình, mỗi bình luận có ba chứng cứ, mỗi chứng cứ dài khoảng 32 ký tự.

## 4 Hướng Tiếp Cận

Với bài toán “*Nhận diện chứng cứ trong phân tích cảm xúc dựa trên các khía cạnh tiếng Việt*”, chúng em chọn hướng tiếp cận giống tác giả là xử lý bài toán theo hướng *nhận diện thực thể có tên (named entity recognition)*. Chúng em sử dụng mô hình BiLSTM-CRF (Huang et al., 2015) kết hợp với các embedding để giải quyết bài toán. Mô hình BiLSTM-CRF trong báo cáo sử dụng 2 lớp embedding: *embedding âm tiết (syllable embedding)* kết hợp với *embedding ký tự (character embedding)* được mô tả trong Hình 3. Ngoài ra chúng em còn

chọn thêm một số pre-trained embedding<sup>7</sup> nổi bậc trong tiếng Việt để tiến hành thực nghiệm và so sánh kết quả.

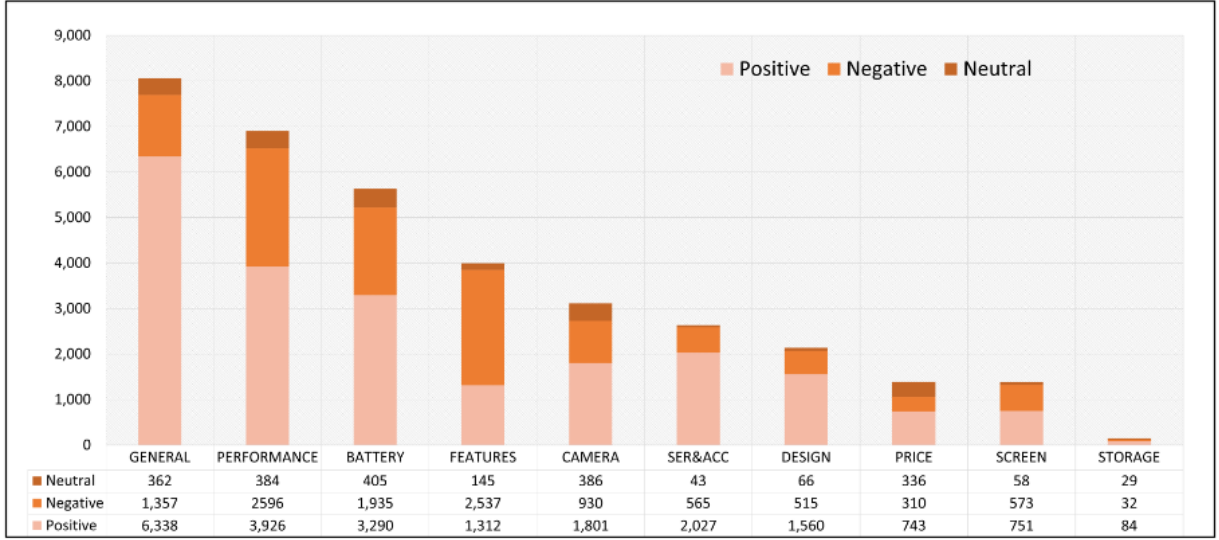
### 4.1 Kết Hợp Lớp Embedding

Lớp embedding lấy đầu vào là một chuỗi có  $N$  tokens  $(x_1, x_2, \dots, x_N)$ , và đầu ra là các vector với số chiều cố định biểu diễn cho từng token  $(e_1, e_2, \dots, e_N)$ . Chúng em sử dụng phương pháp embedding kết hợp giữa syllable embedding (Nguyen et al., 2017) và character embedding (CharLSTM). Các pre-trained embedding sử dụng trong bài đều sử dụng mức độ syllable và có số chiều là 300.

### 4.2 Bidirectional Long Short-Term Memory (BiLSTM)

Một mạng bộ nhớ dài-ngắn hạn (*long-short term memory – LSTM*) là một dạng đặc biệt của *mạng neuron hồi quy (recurrent neural network – RNN)* được giới thiệu bởi Hochreiter và cộng sự, (Hochreiter and Schmidhuber, 1997), có thể nắm bắt mối quan hệ ngữ nghĩa dài hạn bằng cách duy trì một ô nhớ (*memory cell*) lưu trữ thông tin ngữ cảnh. LSTM không bị các vấn đề “*vanishing gradient*” và “*exploding gradient*”. LSTM được trang bị ô nhớ với cơ chế điều chỉnh thích ứng giúp điều chỉnh thông tin được thêm vào hoặc xóa khỏi ô. Ô nhớ được cập nhật liên tục trong quá trình mã hóa và tốc độ thông tin được xác định bởi ba cổng hạt nhân, bao gồm cổng đầu vào (*input gate*), cổng quên (*forget gate*) và cổng đầu ra (*output gate*). Về mặt hình thức, quá trình mã hóa tại thời điểm  $t$  được thực hiện như sau:

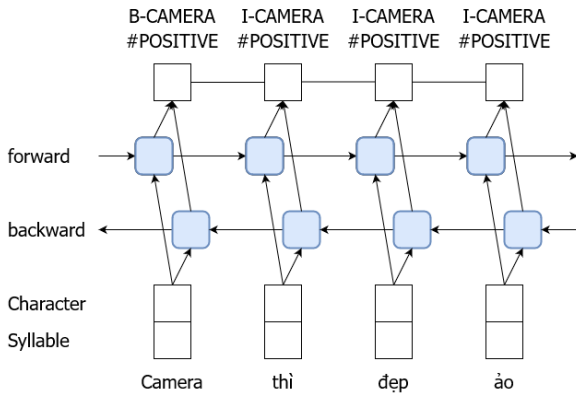
<sup>7</sup>Embedding đã được huấn luyện trước trên các bộ dữ liệu lớn



Hình 2: Chi tiết sự phân bố của 10 loại khía cạnh

Set	Comment	Average aspect per comment	Average span length	Positive	Negative	Neutral	Total span
Train	7,784	3.2	32.6	15,356	7,793	1,560	35,396
Dev	1,113	3.1	32.4	2,110	1,144	241	
Test	2,225	3.2	32.5	4,266	2,269	413	

Bảng 3: Thống kê tổng quan về bộ dữ liệu UIT-ViSD4SA



Hình 3: Mạng BiLSTM-CRF với các lớp embedding

vector word embedding và vector trạng thái ẩn tại thời điểm  $t$ . Cả  $\sigma$  và  $\tanh$  đều là các hàm kích hoạt (activation function) và  $\odot$  đại diện cho phép nhân từng phần tử (element-wise product) của 2 ma trận.  $W^*$  và  $b^*$  là các tham số trong mạng cung cấp ma trận trọng số và vector bias. Mặc dù LSTM có thể giải quyết vấn đề phụ thuộc tồn tại lâu, nhưng nó vẫn mất một số thông tin ngữ nghĩa do cách mã hóa tuần tự của LSTM. Ví dụ,  $h_t$  chỉ chứa thông tin ngữ nghĩa trước thời điểm  $t$ . Do đó, một LSTM hai chiều (BiLSTM) là cần thiết để mô hình hóa cả thông tin ngữ cảnh forward và backward như trong phương trình (7,8) và hai trạng thái ẩn được ghép nối để thu được kết quả cuối cùng dưới dạng phương trình (9):

$$i_t = \sigma(W_{hi}h_{t-1} + W_{ei}e_t^w + b_i) \quad (1)$$

$$f_t = \sigma(W_{hf}h_{t-1} + W_{ef}e_t^w + b_f) \quad (2)$$

$$\tilde{c}_t = \tanh(W_{hc}h_{t-1} + W_{ec}e_t^w + b_c) \quad (3)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \quad (4)$$

$$o_t = \sigma(W_{ho}h_{t-1} + W_{eo}e_t^w + b_o) \quad (5)$$

$$h_t = o_t \odot \tanh(c_t) \quad (6)$$

$$\vec{h}_t = F(e_t^w, \vec{h}_{t-1}) \quad (7)$$

$$\overleftarrow{h}_t = F(e_t^w, \overleftarrow{h}_{t-1}) \quad (8)$$

$$h_t = [\vec{h}_t, \overleftarrow{h}_t] \quad (9)$$

### 4.3 Conditional Random Fields (CRF)

Trường điều kiện ngẫu nhiên (Conditional Random Fields – CRF) (Lafferty et al., 2001) là một framework mô hình chuỗi thừa kế tất cả các ưu

điểm của mô hình MEMM (McCallum et al., 2000; Ratnaparkhi, 1996) đồng thời giải quyết vấn đề “label bias”. Với CRF, các đầu vào và đầu ra được kết nối trực tiếp, không giống như các mạng LSTM và BiLSTM nơi sử dụng các ô nhớ/thành phần hồi quy. Cho một tập dữ liệu huấn luyện  $D = (x^1, y^1), \dots, (x^N, y^N)$  của  $N$  chuỗi dữ liệu được gán nhãn  $x^i$  và chuỗi nhãn tương ứng của chúng là  $y^i$ , CRF cực đại log-likelihood có điều kiện của chuỗi nhãn dựa trên chuỗi dữ liệu được thể hiện như sau:

$$L = \sum_{i=1}^N \log(P(y^i|x^i)) - \sum_{k=1}^K \frac{\lambda_k^2}{2\sigma^2} \quad (10)$$

## 5 Thực Nghiệm và Kết Quả

### 5.1 Cài Đặt Thực Nghiệm

Theo định dạng IOB<sup>8</sup>, bộ dữ liệu được chuyển thành 3 bộ dữ liệu gồm: bộ dữ liệu chỉ chứa các nhãn khía cạnh (SCREEN, BATTERY, CAMERA,...), bộ dữ liệu chỉ chứa nhãn cảm xúc (POSITIVE, NEUTRAL và NEGATIVE) và bộ dữ liệu chứa cả nhãn khía cạnh và cảm xúc (SCREEN#POSITIVE, BATTERY#NEGATIVE, CAMERA#NEUTRAL...) để đánh giá toàn diện cách tiếp cận của chúng em. Word embeddings của chúng em có hai phần: syllable (1) và character (2), với kích thước embedding là 300. Chúng em đặt các hidden layer của LSTM là 100, tỷ lệ dropout là 0,33 và batch size là 300 với 30 epochs để huấn luyện. Tất cả các thực nghiệm đều được thực hiện trên Google Colab Pro (Tesla P100 GPU).

### 5.2 Độ Đo Đánh Giá

Trong báo cáo này, chúng em sử dụng ba độ đo để đánh giá: Precision, Recall và F1-score. Để có được cái nhìn toàn diện, chúng em tính toán các độ đo đánh giá này trên cả micro và macro. Chúng em sử dụng framework sequeval (Nakayama, 2018) để tính toán các độ đo trên. sequeval là một framework của Python để đánh giá nhãn chuỗi. sequeval có thể đánh giá hiệu suất của các tác vụ phân loại như named-entity recognition, part-of-speech tagging, semantic role labeling và một số tác vụ khác.

### 5.3 Kết Quả Thực Nghiệm

Bảng 4 trình bày kết quả thực nghiệm sử dụng mô hình BiLSTM-CRF với lần lượt các syllable embedding (không sử dụng pre-trained, sử dụng pre-trained PhoW2V, cc.vi.300.vec, wiki.vi.vec) và sự

<sup>8</sup>Viết tắt của inside, outside, beginning

kết hợp giữa các syllable embedding trên với character embedding charLSTM trên lần lượt các tác vụ nhận diện aspect, polarity, aspect#polarity. Đối với tác vụ nhận diện aspect, mô hình cho kết quả tốt nhất khi kết hợp pre-trained wiki.vi.vec với charLSTM cho phần embedding với  $F1_{macro} = 54.07\%$ . Đối với tác vụ nhận diện polarity, mô hình cho kết quả tốt nhất khi chỉ sử dụng pre-trained wiki.vi.vec cho phần embedding với  $F1_{macro} = 38.50\%$ . Đối với tác vụ nhận diện aspect#polarity, mô hình cho kết quả tốt nhất khi chỉ sử dụng pre-trained wiki.vi.vec cho phần embedding với  $F1_{macro} = 38.31\%$ . Nhìn chung, khi kết hợp embedding charLSTM cho các tác vụ thì chỉ có tác vụ nhận diện aspect là cho kết quả cao hơn đôi chút còn tác vụ nhận diện polarity và aspect#polarity kết quả không những không cải thiện mà còn có phần thấp hơn khi không áp dụng charLSTM.

Kết quả chi tiết cho mỗi lớp của từng tác vụ được trình bày trong Bảng 5, 6 và 7 (với nhãn aspect#polarity, chúng em chỉ hiển thị F1-score). Đối với tác vụ nhận diện aspect, chỉ có một aspect có F1-score cao trên 70% (CAMERA) trong khi phần còn lại nằm trong khoảng 50-60%, đặc biệt F1-score về PRICE và STORAGE tương đối thấp (dưới 50%). Đối với tác vụ nhận diện polarity, kết quả giảm dần theo thứ tự POSITIVE, NEGATIVE, NEUTRAL. Kết quả của tác vụ nhận diện aspect#polarity có thể được coi là kết hợp của hai tác vụ trước đó: các nhãn aspect có hiệu suất cao trước đó kết hợp với positive cho kết quả cao nhất. Kết quả này giải thích sự thiếu đồng nhất về số lượng trong các nhãn polarity (nhãn NEUTRAL chỉ chiếm 6,25% trong bộ dữ liệu UIT-ViSD4SA, chi tiết trong Hình 2). Nói chung, cách tiếp cận của chúng em đạt được hiệu suất tốt hơn khi nhận diện chứng cứ cho tác vụ aspect hơn là nhận diện chứng cứ cho tác vụ polarity và aspect#polarity. Tuy nhiên, khả năng nhận diện chứng cứ đối với tất cả các loại nhãn vẫn còn hạn chế (F1-score dưới 80%).

### 5.4 Phân Tích Kết Quả Dự Đoán

Bảng 8 cho thấy một số trường hợp được dự đoán bởi mô hình BiLSTM-CRF. Sau khi xem xét các trường hợp, chúng em nhận thấy rằng mô hình phạm phải ba loại lỗi phổ biến là: không thể nhận diện được chứng cứ, phân loại sai cảm xúc và phát hiện sai ranh giới của các chứng cứ. Như đã quan sát trong câu 1 và câu 2, mô hình nhận diện as-

Tác vụ	Pre-trained embedding	Character	$P_{micro}$	$R_{micro}$	$F1_{micro}$	$P_{macro}$	$R_{macro}$	$F1_{macro}$
Aspect	Không	Không	51.73	57.79	54.59	49.07	53.63	51.12
	PhoW2V	Không	54.53	56.80	55.64	51.67	52.78	52.09
	cc.vi.300.vec	Không	52.71	58.27	55.35	50.96	52.78	52.34
	wiki.vi.vec	Không	55.06	59.21	57.06	52.00	54.95	53.38
	Không	Có	55.27	60.36	57.70	52.24	56.33	54.05
	PhoW2V	Có	56.97	57.47	57.22	54.45	53.86	54.05
	cc.vi.300.vec	Có	53.35	56.77	55.00	51.35	53.17	52.14
	wiki.vi.vec	Có	57.32	58.41	<b>57.86</b>	53.89	54.34	<b>54.07</b>
Polarity	Không	Không	44.68	43.65	44.16	36.10	35.15	35.61
	PhoW2V	Không	43.49	49.93	46.49	36.59	39.68	37.88
	cc.vi.300.vec	Không	44.04	50.32	46.97	35.49	41.52	38.24
	wiki.vi.vec	Không	46.83	49.30	46.03	38.46	38.78	<b>38.50</b>
	Không	Có	48.42	48.38	<b>48.40</b>	38.69	37.70	38.11
	PhoW2V	Có	44.87	47.32	46.06	39.05	36.37	36.88
	cc.vi.300.vec	Có	45.24	46.25	45.74	36.32	36.38	36.32
	wiki.vi.vec	Có	48.38	47.60	47.99	37.54	38.65	37.82
Aspect-polarity	Không	Không	48.31	53.63	50.83	33.98	37.41	35.33
	PhoW2V	Không	49.53	54.38	51.84	34.66	37.44	35.52
	cc.vi.300.vec	Không	50.67	53.85	52.21	35.42	38.56	36.71
	wiki.vi.vec	Không	56.42	53.44	<b>54.89</b>	41.36	37.09	<b>38.31</b>
	Không	Có	50.39	53.06	51.69	34.87	35.88	34.77
	PhoW2V	Có	51.15	53.57	52.33	36.03	36.59	35.69
	cc.vi.300.vec	Có	48.75	54.69	51.55	35.39	37.31	35.45
	wiki.vi.vec	Có	50.63	53.72	52.13	38.19	37.23	35.28

Bảng 4: Tổng quan kết quả thử nghiệm

Aspect	Precision	Recall	F1-score
BATTERY	63.63	67.58	65.54
CAMERA	70.43	71.65	<b>71.04</b>
DESIGN	58.26	59.94	59.09
FEATURES	49.04	51.00	50.00
GENERAL	60.62	60.62	60.62
PERFORMANCE	52.90	53.89	53.43
PRICE	35.93	36.89	36.40
SCREEN	57.26	59.74	58.47
SER&ACC	53.76	50.78	52.22
STORAGE	37.04	31.25	33.90

Bảng 5: Kết quả trên mỗi lớp cho nhãn aspect

Aspect	Negative	Neutral	Positive
BATTERY	44.57	27.33	70.46
CAMERA	51.36	44.83	<b>72.26</b>
DESIGN	28.17	08.00	66.54
FEATURES	44.57	25.00	60.55
GENERAL	41.31	46.32	62.97
PERFORMANCE	38.49	15.91	65.03
PRICE	21.28	17.48	42.16
SCREEN	34.29	33.33	67.82
SER&ACC	14.06	00.00	61.45
STORAGE	00.00	00.00	43.75

Bảng 7: F1-score trên mỗi lớp cho nhãn aspect#polarity

Sentiment	Precision	Recall	F1-score
NEGATIVE	34.59	35.49	35.03
NEUTRAL	26.49	21.68	23.85
POSITIVE	54.30	59.16	<b>56.63</b>

Bảng 6: Kết quả trên mỗi lớp cho nhãn polarity

pect#polarity không phát hiện được chứng cứ “*thử chiến game cũng ok*” và “*cảm ứng mượt*” nhưng khi dự đoán trên hai mô hình nhận diện aspect và polarity riêng biệt thì kết quả dự đoán lại rất chính xác, điều này chứng tỏ với tác vụ aspect#polarity

vẫn còn nhiều thách thức đối với các mô hình học máy hiện tại. Với những trường hợp phân loại sai, chúng em nhận thấy rằng nhiều trường hợp nhầm lẫn này có chứa từ mượn tiếng Anh. Ví dụ: “*luot wep xem fim*” trong ngữ cảnh này là về thời gian sử dụng pin của điện thoại và được gán nhãn là BATTERY#POSITIVE. Tuy nhiên, mô hình đã hiểu sai và phân loại nó thành PERFORMANCE (mô hình có nhãn aspect) hoặc SCREEN#NEGATIVE (mô hình có nhãn aspect#polarity) do sự phức tạp của từ mượn và có thể do chưa bắt được đầy đủ ngữ



	Nhãn mục tiêu	Aspect prediction	Polarity prediction	Aspect#polarity prediction
1	Mua cho mẹ sài nên cũng không đòi hỏi gì nhiều, <b>máy đẹp</b> DESIGN#POSITIVE camera siêu ảo, <b>thử chiến game cũng ok</b> PERFORMANCE#POSITIVE, <b>pin sài dc 2 ngày với loạt wep xem fim</b> BATTERY#POSITIVE, <b>nhân viên tgdd an minh KG phục vụ qua nhiệt tình</b> SER&ACC#POSITIVE cho 5*	máy đẹp DESIGN camera siêu ảo CAMERA X thử chiến game cũng ok PERFORMANCE pin sài dc 2 ngày BATTERY luot wep xem fim PERFORMANCE X nhân viên tgdd an minh KG phục vụ qua nhiệt tình SER&ACC	sài nên cũng không đòi hỏi gì nhiều POSITIVE X máy đẹp POSITIVE camera siêu ảo POSITIVE X thử chiến game cũng ok POSITIVE pin sài dc 2 ngày luot wep xem fim POSITIVE nhân viên tgdd an minh KG phục vụ qua nhiệt tình POSITIVE cho 5* POSITIVE X	máy đẹp DESIGN#POSITIVE camera siêu ảo CAMERA#POSITIVE X pin sài dc 2 ngày BATTERY#POSITIVE luot wep xem fim SCREEN#NEGATIVE X nhân viên tgdd an minh KG phục vụ qua nhiệt tình SER&ACC#POSITIVE cho 5* SER&ACC#POSITIVE X
2	Mình mới mua. mình <b>thấy mẫu đẹp</b> DESIGN#POSITIVE <b>pin trâu</b> BATTERY#POSITIVE <b>cảm ứng mượt</b> FEATURES#POSITIVE <b>được em nhân viên ĐMX tư vấn rất nhiệt tình</b> SER&ACC#POSITIVE	thấy mẫu đẹp DESIGN pin trâu BATTERY cảm ứng mượt FEATURES nhân viên ĐMX tư vấn rất nhiệt tình SER&ACC	thấy mẫu đẹp POSITIVE pin trâu POSITIVE cảm ứng mượt POSITIVE nhân viên ĐMX tư vấn rất nhiệt tình POSITIVE	thấy mẫu đẹp DESIGN#POSITIVE pin trâu BATTERY#POSITIVE nhân viên ĐMX tư vấn rất nhiệt tình SER&ACC#POSITIVE

Bảng 8: Một ví dụ về kết quả dự đoán của mô hình. Các đoạn được in đậm với các aspect và các polarity của chúng được đưa ra dưới dạng các chỉ số phụ. Các dự đoán không chính xác được đánh dấu là X

cảnh trước đó. Đặc điểm này cần được quan tâm trong các nghiên cứu sau này vì đặc điểm ngôn ngữ tiếng Việt (đặc biệt là trong lĩnh vực công nghệ) thường bao gồm nhiều từ mượn với nghĩa có thể giống hoặc khác với ngôn ngữ gốc. Bên cạnh đó, việc phát hiện ranh giới của các chứng cứ trong các phản hồi của khách hàng cũng còn nhiều thách thức do sự phức tạp trong ngữ cảnh và cách các khía cạnh bổ nghĩa cho nhau. Ví dụ: “pin sài dc 2 ngày với loạt wep xem fim” là một chứng cứ liên tục nhưng mô hình nhận diện aspect và aspect#polarity đã xác định sai ranh giới của chứng cứ này và chia nó ra thành 2 chứng cứ khác nhau: BATTERY và PERFORMANCE (mô hình có nhãn aspect), BATTERY#POSITIVE và SCREEN#NEGATIVE (mô hình có nhãn aspect#polarity), sự phức tạp về ngữ cảnh cụ thể ở đây là “luot wep xem fim” cũng có thể hiểu là một khía cạnh về PERFORMANCE nhưng trong ngữ cảnh này lại được sử dụng với mục đích thực sự là bổ nghĩa cho “pin sài dc 2 ngày” vì vậy được an-

notator gán nhãn là BATTERY. Ngoài ra, có một số lỗi xuất phát từ bộ dữ liệu bị gán nhãn sai, tuy nhiên mô hình lại dự đoán rất chính xác. Ví dụ: “camera siêu ảo” là về camera của điện thoại thông minh và đáng lẽ nên được gán nhãn là CAMERA#POSITIVE nhưng lại không được gán bất kỳ một nhãn nào, đây là một sai sót trong quá trình gán nhãn dữ liệu nhưng có thể thấy mô hình đã học được khá tốt đặc điểm này (CAMERA) và dự đoán được những câu có nhãn ban đầu không chính xác có thể vì số lượng nhãn CAMERA#POSITIVE trong bộ dữ liệu khá nhiều và mô hình học được tốt trên nhãn này (F1-score trên 70%) nên có thể bắt được những lỗi sai này nhưng nếu đổi lại nhãn khác thì chưa chắc mô hình có thể bắt được mà ngược lại khiến mô hình hiểu nhầm và dẫn đến kết quả tệ đi.

## 6 Kết Luận

Báo cáo trên đây là toàn bộ những tìm hiểu của chúng em về bài toán “Nhận diện chứng cứ dựa



trên các khía cạnh cho phân tích cảm xúc tiếng Việt”. Chúng em tìm hiểu bài toán thông qua bộ dữ liệu UIT-ViSD4SA, một bộ dữ liệu mới cho bài toán này bao gồm 35,396 chứng cứ đã được gán nhãn trên 11,122 bình luận của khách hàng về các sản phẩm điện thoại thông minh. Trong đó mỗi bình luận bao gồm một số chứng cứ với nhãn là một trong 10 khía cạnh chi tiết và trạng thái cảm xúc tương ứng. Hiện tại chúng em đã cài đặt thành công mô hình BiLSTM-CRF sử dụng 2 lớp embedding về âm tiết và ký tự, kết quả cao nhất mà chúng em đạt được là 54.07%  $F1_{macro}$  trên tác vụ nhận diện chứng cứ cho khía cạnh, 38.50%  $F1_{macro}$  trên tác vụ nhận diện chứng cứ cho trạng thái cảm xúc, 38.31%  $F1_{macro}$  trên tác vụ nhận diện chứng cứ cho khía cạnh#trạng thái cảm xúc. Kết quả đạt được bên trên còn khá thấp (thấp hơn kết quả mà nhóm tác giả đã đạt được lần lượt là 62.76%, 49.77%, 45.70%), chưa thể đáp ứng được nhu cầu sử dụng, và trên cả bài báo của nhóm tác giả cũng chưa đạt được kết quả mà họ thật sự mong muốn, nó đặt ra một thử thách cho các nhóm nghiên cứu sau về việc cải thiện kết quả cho bài toán.

Hướng phát triển trước mắt của nhóm là tiếp tục cài đặt lớp embedding cho ngữ cảnh để hoàn thiện mô hình BiLSTM-CRF sử dụng 3 lớp embedding âm tiết, ký tự và ngữ cảnh. Sau đó nhóm sẽ tìm hiểu thêm về các mô hình pre-trained hoặc các cấu trúc khác dựa trên nền tảng máy đọc hiểu thay cho BiLSTM-CRF. “Nhận diện chứng cứ dựa trên các khía cạnh cho phân tích cảm xúc tiếng Việt” là một bài toán còn mới, hay và thiết thực để nghiên cứu trong tương lai gần.

## Tài Liệu Tham Khảo

- Peng Chen, Zhongqian Sun, Lidong Bing, and Wei Yang. 2017. Recurrent attention network on memory for aspect sentiment analysis. In *Proceedings of the 2017 conference on empirical methods in natural language processing*, pages 452–461.
- Zhuang Chen and Tieyun Qian. 2020. Relation-aware collaborative learning for unified aspect-based sentiment analysis. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3685–3694.
- Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation*, 9(8):1735–1780.
- Minghao Hu, Yuxing Peng, Zhen Huang, Dongsheng Li, and Yiwei Lv. 2019. Open-domain targeted sentiment analysis via span-based extraction and classification. *arXiv preprint arXiv:1906.03820*.
- Minqing Hu and Bing Liu. 2004. Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 168–177.
- Zhiheng Huang, Wei Xu, and Kai Yu. 2015. Bidirectional lstm-crf models for sequence tagging. *arXiv preprint arXiv:1508.01991*.
- Yohan Jo and Alice H Oh. 2011. Aspect and sentiment unification model for online review analysis. In *Proceedings of the fourth ACM international conference on Web search and data mining*, pages 815–824.
- Svetlana Kiritchenko, Xiaodan Zhu, Colin Cherry, and Saif Mohammad. 2014. Nrc-canada-2014: Detecting aspects and sentiment in customer reviews. In *Proceedings of the 8th international workshop on semantic evaluation (SemEval 2014)*, pages 437–442.
- John Lafferty, Andrew McCallum, and Fernando CN Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data.
- Xin Li, Lidong Bing, Piji Li, and Wai Lam. 2019. A unified model for opinion target extraction and target sentiment prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 6714–6721.
- Huaishao Luo, Lei Ji, Tianrui Li, Nan Duan, and Daxin Jiang. 2020. Grace: Gradient harmonized and cascaded labeling for aspect-based sentiment analysis. *arXiv preprint arXiv:2009.10557*.
- Long Mai and Bac Le. 2018. Aspect-based sentiment analysis of vietnamese texts with deep learning. In *Asian Conference on Intelligent Information and Database Systems*, pages 149–158. Springer.
- Andrew McCallum, Dayne Freitag, and Fernando CN Pereira. 2000. Maximum entropy markov models for information extraction and segmentation. In *Icml*, volume 17, pages 591–598.
- Hiroki Nakayama. 2018. *sequeval: A python framework for sequence labeling evaluation*. Software available from <https://github.com/chakki-works/sequeval>.
- Dat Quoc Nguyen, Thanh Vu, Dai Quoc Nguyen, Mark Dras, and Mark Johnson. 2017. From word segmentation to pos tagging for vietnamese. *arXiv preprint arXiv:1711.04951*.
- Huyen TM Nguyen, Hung V Nguyen, Quyen T Ngo, Luong X Vu, Vu Mai Tran, Bach X Ngo, and Cuong A Le. 2018. Vlsr shared task: sentiment analysis. *Journal of Computer Science and Cybernetics*, 34(4):295–310.

Kiet Nguyen, Vu Nguyen, Anh Nguyen, and Ngan Nguyen. 2020. A vietnamese dataset for evaluating machine reading comprehension. In <i>Proceedings of the 28th International Conference on Computational Linguistics</i> , pages 2595–2605.	950
Kim Thi-Thanh Nguyen, Sieu Khai Huynh, Luong Luc Phan, Phuc Huynh Pham, Duc-Vu Nguyen, and Kiet Van Nguyen. 2021. Span detection for aspect-based sentiment analysis in vietnamese. <i>arXiv preprint arXiv:2110.07833</i> .	951
Minh-Hao Nguyen, Tri Minh Nguyen, Dang Van Thin, and Ngan Luu-Thuy Nguyen. 2019. A corpus for aspect-based sentiment analysis in vietnamese. In <i>2019 11th International Conference on Knowledge and Systems Engineering (KSE)</i> , pages 1–5. IEEE.	952
Luong Luc Phan, Phuc Huynh Pham, Kim Thi-Thanh Nguyen, Tham Thi Nguyen, Sieu Khai Huynh, Luan Thanh Nguyen, Tin Van Huynh, and Kiet Van Nguyen. 2021. Sa2sl: From aspect-based sentiment analysis to social listening system for business intelligence. <i>arXiv preprint arXiv:2105.15079</i> .	953
Maria Pontiki, Dimitrios Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Mohammad Al-Smadi, Mahmoud Al-Ayyoub, Yanyan Zhao, Bing Qin, Orphée De Clercq, et al. 2016. Semeval-2016 task 5: Aspect based sentiment analysis. In <i>International workshop on semantic evaluation</i> , pages 19–30.	954
Maria Pontiki, Dimitrios Galanis, Harris Papageorgiou, Suresh Manandhar, and Ion Androutsopoulos. 2015. Semeval-2015 task 12: Aspect based sentiment analysis. In <i>Proceedings of the 9th international workshop on semantic evaluation (SemEval 2015)</i> , pages 486–495.	955
Maria Pontiki, Dimitris Galanis, John Pavlopoulos, Harris Papageorgiou, Ion Androutsopoulos, and Suresh Manandhar. 2014. <a href="#">SemEval-2014 task 4: Aspect based sentiment analysis</a> . In <i>Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)</i> , pages 27–35, Dublin, Ireland. Association for Computational Linguistics.	956
Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. 2016. Squad: 100,000+ questions for machine comprehension of text. <i>arXiv preprint arXiv:1606.05250</i> .	957
Adwait Ratnaparkhi. 1996. A maximum entropy model for part-of-speech tagging. In <i>Conference on empirical methods in natural language processing</i> .	958
Minjoon Seo, Aniruddha Kembhavi, Ali Farhadi, and Hannaneh Hajishirzi. 2016. Bidirectional attention flow for machine comprehension. <i>arXiv preprint arXiv:1611.01603</i> .	959
Dang Van Thin, Ngan Luu-Thuy Nguyen, Tri Minh Truong, Lac Si Le, and Duy Tin Vo. 2021. Two new large corpora for vietnamese aspect-based sentiment analysis at sentence level. <i>Transactions on Asian and Low-Resource Language Information Processing</i> , 20(4):1–22.	960
Hu Xu, Bing Liu, Lei Shu, and Philip S Yu. 2018. Double embeddings and cnn-based sequence labeling for aspect extraction. <i>arXiv preprint arXiv:1805.04601</i> .	961