

# Phân loại cảm xúc các bình luận Tiếng Việt theo khía cạnh

Trần Nhật Tiến<sup>1[18521493]</sup>, Kiên Tiến Đạt<sup>2[18520568]</sup>, Nguyễn Tấn Phong<sup>3[18521239]</sup>, Hồ Đình Long<sup>4[18521022]</sup>, Đỗ Hùng Dũng<sup>5[18520629]</sup>, Nguyễn Văn Kiệt, and Lưu Thanh Sơn

Trường Đại học Công nghệ Thông tin - Đại học Quốc gia Thành phố Hồ Chí Minh

**Tóm tắt nội dung** Trong bài báo này chúng tôi sẽ sử dụng các phương pháp transfer learning khác nhau để giải quyết bài toán phân loại cảm xúc cho các bình luận Tiếng Việt theo khía cạnh.

**Keywords:** Bert · Aspect-Based Sentiment Analysis · Phobert · vibert4new

## 1 Giới thiệu

Nhận diện cảm xúc văn bản là một trong những bài toán phổ biến nhất trong lĩnh vực xử lý ngôn ngữ tự nhiên của giới khoa học dữ liệu. Ngày nay, nhu cầu đó càng ngày càng lớn hơn với sự phát triển mạnh mẽ của thương mại điện tử và các dịch vụ trực tuyến.

Với một khối lượng dữ liệu lớn và công kênh được tạo ra, khi tiếp cận nó bằng phương pháp thủ công sẽ không mang lại kết quả và tốc độ mong muốn. Vì vậy các mô hình học máy đã ra đời để giải quyết bài toán này. Yêu cầu của bài toán là nhận diện cảm xúc của bình luận kèm theo khía cạnh của nó.

- Input : một bình luận tiếng Việt
- Output : khía cạnh + cảm xúc

Cấu trúc của bài báo được tổ chức như sau. Phần 2 chúng tôi giới thiệu, mô tả chi tiết cách quy trình xây dựng bộ dữ liệu sử dụng trong bài báo. Phần 3, chúng tôi phân tích chi tiết bộ dữ liệu. Phần 4 trình bày quy trình tiền xử lý dữ liệu. Phần 5, chúng tôi sẽ trình bày các mô hình và kết quả. Phần 6 là trình bày kết luận và hướng phát triển của bài toán.

## 2 Công trình liên quan

Trong thập kỷ, phân tích tình cảm dựa trên khía cạnh nhận đã được nhiều sự quan tâm bởi các ứng dụng tiềm năng của nó chẳng hạn như phản hồi của người phân tích từ nhận xét của người dùng trực tuyến. Phân tích tình cảm dựa trên khía cạnh hoặc dựa trên khía cạnh khai thác ý kiến đã được điều tra và giới thiệu bởi Hu và Liu [1], trên các khía cạnh của đánh giá sản phẩm áp dụng một bộ quy tắc dựa trên các quan sát thống kê. Những từ được đào tạo trước (Mikolov

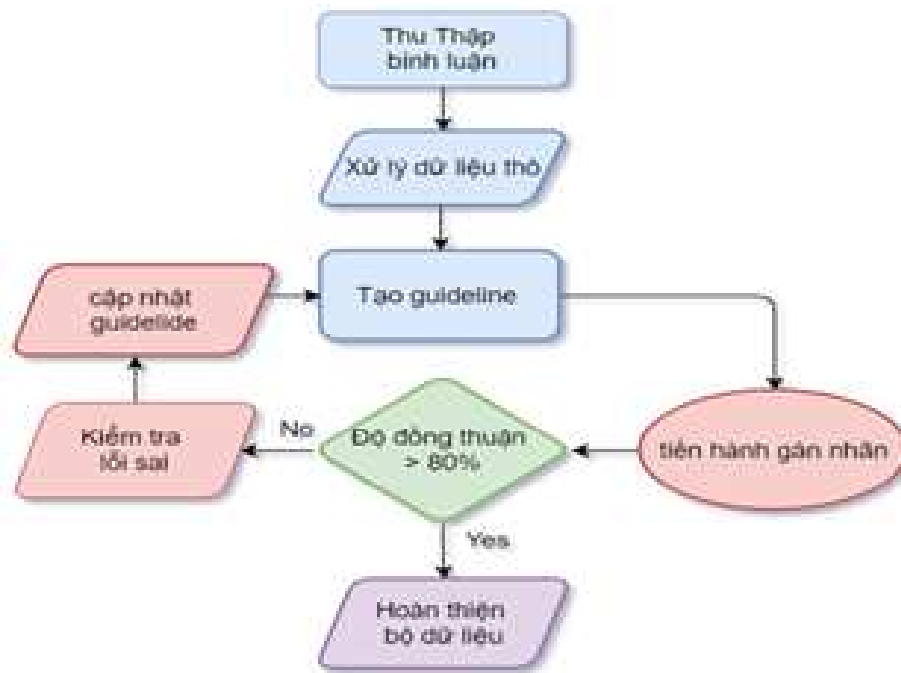
và cộng sự.,2013 [2] ; Pennington và cộng sự . ,2014[3] là một thành phần quan trọng của hệ thống NLP hiện đại có thể cung cấp những cải tiến tuyệt vời so với các phép nhúng đã học từ đầu. Máy hỗ trợ vectơ (SVM) là một thuật toán mạnh mẽ được chọn nhiều nhất cho nhiệm vụ này [4] , [5] - [6]. Các tác giả đã sử dụng SVM nhiều bộ phân loại SVM nhị phân kết hợp với các tính năng thủ công khác nhau. Đối với ngôn ngữ ít tài nguyên như tiếng Việt, đã có một số nghiên cứu về nhiệm vụ phân tích khía cạnh. Tuy nhiên, họ chỉ tập trung vào thiết kế kiến trúc dựa trên việc học có giám sát và phương pháp học sâu để giải quyết nhiệm vụ này. Kết quả của họ khá ấn tượng, tuy nhiên, những phương pháp này đã lỗi thời so với kỹ thuật hiện đại trong sự phát triển của Xử lý ngôn ngữ tự nhiên (NLP). Gần đây, kiến trúc BERT [7] đã cho thấy hiệu quả trên các nhiệm vụ khác nhau trong NLP. Bên cạnh đó, có rất nhiều các biến thể của mô hình BERT được xuất bản cho ngôn ngữ tiếng Việt như BERT đa ngôn ngữ [8]. Những mô hình này được đào tạo trước trên Wikipedia và Dữ liệu thu thập thông tin chung cho các ngôn ngữ khác nhau, bao gồm cả tiếng Việt. Vì những lý do đó ,trong bài báo này, chúng tôi sẽ áp dụng các mô hình này để giải quyết bài toán.

### 3 Bộ dữ liệu

Bộ dữ liệu chúng tôi xây dựng và sử dụng trong bài báo có tên là bộ dữ liệu bình luận tiếng Việt đánh giá ứng dụng Google photos. Thu thập từ ứng dụng Google photos. Bộ dữ liệu có 8 nhãn khía cạnh bao gồm: chung, sao lưu, chỉnh sửa, tính năng, , giao diện, cập nhật, độ ổn định, khác và 3 nhãn cảm xúc: tích cực, tiêu cực, trung tính

#### 3.1 Tạo bộ dữ liệu

Quy trình xây dựng dữ liệu được chia làm 3 phần chính là thu thập dữ liệu, xử lý dữ liệu thô và gán nhãn. Cụ thể được trình bày trong hình dưới đây :



**Hình 1.** Quy trình xây dựng bộ dữ liệu

**Thu thập dữ liệu** Bộ dữ liệu được thu thập từ ứng dụng Google photos và sử dụng công cụ Google-play-scaper. Kết quả thu được 95318 điểm dữ liệu ở thời điểm ban đầu.

**Xử lý dữ liệu thô** Sau khi được thu thập, các dòng dữ liệu trùng lặp được loại bỏ các dòng dữ liệu trùng lặp. Sau đó chúng tôi chọn ra 5000 dòng dữ liệu dài nhất để tiến hành gán nhãn.

### 3.2 Hướng dẫn gán nhãn

Có 2 loại nhãn riêng biệt là nhãn khía cạnh và nhãn đánh giá. Với mỗi nhãn khía cạnh riêng biệt sẽ được gán với một nhãn đánh giá tương ứng để tạo thành một nhãn kép. Đối với một điểm dữ liệu, có thể gán một hoặc nhiều nhãn kép tương ứng với các khía cạnh mà điểm dữ liệu đó đề cập tới.

**Nhãn khía cạnh** : là những nhãn đề cập đến khía cạnh của sản phẩm

**Giao diện:** Đề cập tới cách sắp xếp, bố trí, thẩm mỹ, tính trực quan của ứng dụng và ảnh trong ứng dụng.

VD: Cần thêm cách sắp xếp mới nhất trước, chứ ko phải theo ngày Chính sửa: Những vấn đề về tính năng cắt, ghép, dán, chỉnh sửa ảnh video.

VD: Sao giờ xóa ảnh đã đồng bộ đi mà ảnh trên thu mục ảnh trên máy nó cũng bị xóa theo luôn vậy?

**Tính năng khác:** Đề cập về những tính năng khác như tải ảnh về máy, các thao tác trong ứng dụng, thùng rác, tải ảnh, giải phóng bộ nhớ, trợ lý ảo, tự động sắp xếp album, nhận diện khuôn mặt, bảo mật, chụp ảnh,...

VD: Tại sao không tải ảnh từ gg ảnh về điện thoại được. Sao lưu: Đề cập về vấn đề lưu trữ, đồng bộ hoá ảnh, video của người dùng. Lưu giữ được dữ liệu người dùng hoặc ngược lại như mất ảnh, dung lượng,..., đồng bộ lên bị mờ ảnh, video.

VD: Các video tải lên xem không được nét như cũ nữa, ảnh thì vẫn bình thường mong ad sửa lại, lúc trước thì cái gì tải lên xem lại cũng nét mới gần đây là bị hiện tượng video coi lại không nét như trước nữa

**Cập nhật:** Đề cập đến các bản cập nhật của ứng dụng, những vấn đề trước và sau về mặt thời gian của ứng dụng.

VD: Rất thích nhưng cập nhật lên bản mới sao lưu chậm mà ko hiện thị hình ảnh đã sao lưu từ trước gì hết.

**Độ ổn định:** Đề cập về tình trạng ổn định của ứng dụng như nhanh hay chậm, giật lag (bao gồm cả ứng dụng và vấn đề internet),... xem video, ảnh không được. Các loại lỗi của ứng dụng như lặp ảnh, lỗi tương tác,...

VD: Sao thùng rác của tôi k có j mà nó cứ ghi là thùng rác đầy là sao Mong ad mau khắc phục lỗi ạ

**Chung:** Những nhận xét đánh chung chung, không nhận xét hay khen - chê chi tiết về một khía cạnh nào cụ thể.

VD: tiện lợi, cảm ơn, rất tốt, ứng dụng hay, rất tệ, ứng dụng tệ hại.

**Khác:** Những vấn đề liên quan tới ứng dụng nhưng không liên quan tới những khía cạnh trên như dung lượng ứng dụng, tài khoản người dùng,... Những đánh giá không liên quan.

VD: google rất vô lý tôi quên mật khẩu đăng nhập google đã lấy mã xác minh ở email khác từ google gửi về để xác minh tài khoản thì nó bảo là chúng tôi không thể xác minh tài khoản này thuộc về bạn Mong nhà sx xem lại rất chi là vô lý

**Nhân cảm xúc :** Là những nhân đánh giá được về quan điểm, cảm nhận của khía cạnh mà câu đánh giá nhắc tới. Bao gồm 3 nhân:

**Positive:** Nhận xét tích cực, hài lòng, khen, cảm xúc vui thích.

VD: Rất tuyệt vời, Đẹp - Hoàn Hảo - Rất Có Ích. Thanks so much.

**Negative:** Nhận xét tiêu cực, không hài lòng, chê, cảm xúc khó chịu, có những từ ngữ tục tĩu mang tính tiêu cực.

VD: Tự nhiên xóa 1 ảnh mà mất hết cả album ảnh. Điên quá

**Neutral :** Các đánh giá không mang tính chất tích cực hay tiêu cực quá rõ ràng (trung dung), góp ý xây dựng ứng dụng. VD: Cần thêm tính năng có thể lưu dc ảnh vào máy

### 3.3 Kết quả

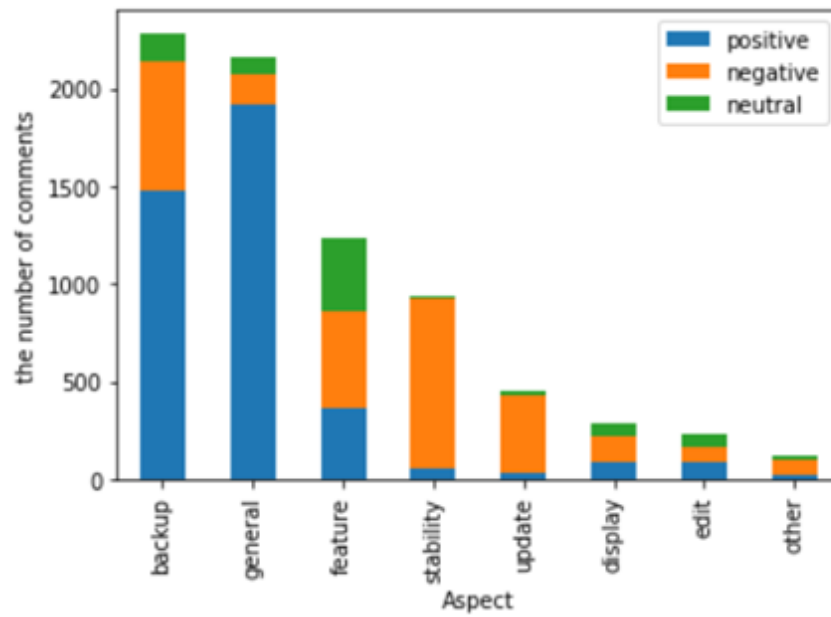
Chúng tôi thu được bộ dữ liệu với số lượng từng nhãn được trình bày trong bảng dưới đây

	Chung	Giao diện	Chỉnh sửa	Tính năng	Sao lưu	Cập nhật	Độ ổn định	Khác
Tích cực	1918	88	91	362	1478	27	49	26
Tiêu cực	155	131	74	495	662	404	872	75
Trung tính	84	69	71	377	144	25	19	15
Tổng	2157	288	236	1234	2284	458	940	116

**Bảng 1.** Thống kê bộ dữ liệu

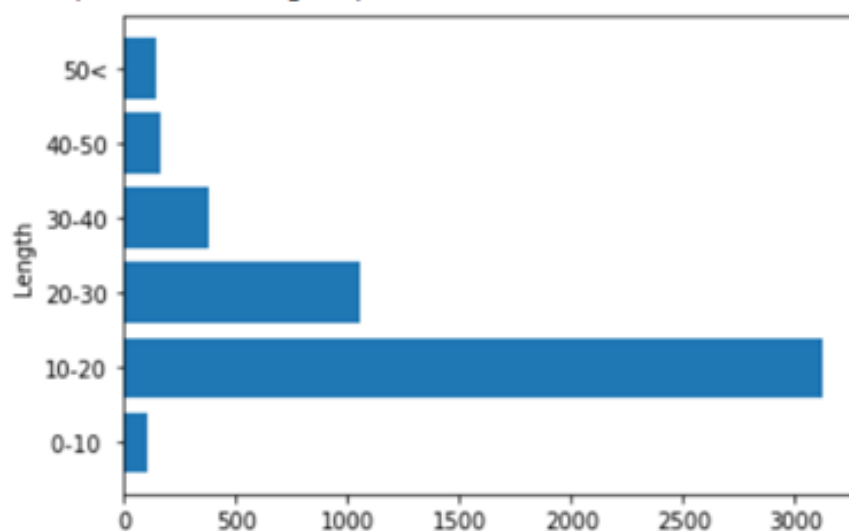
### 3.4 Phân tích bộ dữ liệu

Để có một cái nhìn tổng quan về bộ dữ liệu, chúng tôi tiến hành phân tích bộ dữ liệu. Hình thể hiện sự phân bố của các nhãn cảm xúc trên từng khía cạnh. Tuy nhiên, số lượng nhãn cảm xúc tích cực có sự chênh lệch lớn hơn so 2 nhãn còn lại, tập trung chủ yếu vào khía cạnh lưu trữ (backup) và chung (general). Nhãn cảm xúc tiêu cực (negative) phân bố tương đối đều trên các khía cạnh. Còn nhãn cảm xúc trung tính (neutral) ít hơn số với 2 nhãn còn lại và phân bố không đều. Các nhãn khía cạnh có sự chênh lệch tương đối lớn.



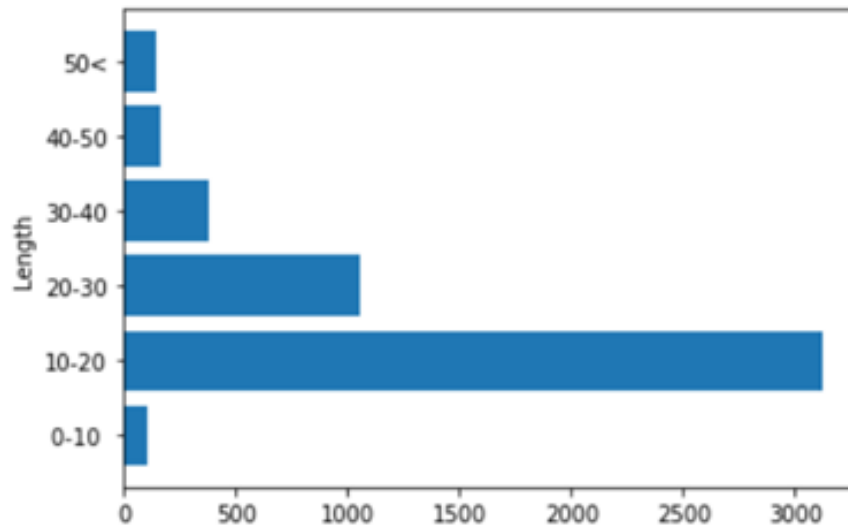
**Hình 2.** Phân bố của nhãn cảm xúc trên từng khía cạnh.

Ngoài ra, chúng tôi đã liệt kê ra số lượng câu trong bộ dữ liệu. Hình cho thấy độ dài của các câu đánh giá. Có thể dễ dàng nhìn thấy, hầu hết các câu đánh giá đều có nội dung từ 10-20 từ vựng, chiếm hơn một nửa số với các câu đánh giá khác.



**Hình 3.** Phân bố độ dài câu đánh giá

Bên cạnh đó, chúng tôi cũng liệt kê những từ lặp lại nhiều trong câu. Qua hình có thể thấy những từ ngữ được lặp lại nhiều lần nhất trong nội dung câu đánh giá thường là những từ liên quan mật thiết tới ứng dụng như “ảnh”, “dùng”, và những tính năng được quan tâm của ứng dụng như sao lưu và lưu trữ.



**Hình 4.** Những từ xuất hiện nhiều nhất trong các câu đánh giá

## 4 Tiền xử lý dữ liệu

Để làm sạch dữ liệu, chúng tôi đã thực hiện một số xử lý trước khi đưa vào mô hình huấn luyện. Quá trình tiền xử lý gồm các bước như sau:

Bước xử lý	Mô tả
1	Xoá các ký tự kéo dài
2	Sửa lỗi chính tả
3	Chuyển biểu tượng cảm xúc thành "tích cực" hoặc "tiêu cực"
4	Đưa chữ viết hoa về chữ viết thường
5	Chuẩn hoá Unicode
6	Loại bỏ dấu câu và các ký tự đặc biệt

**Bảng 2.** Các bước tiền xử lý

Trong các câu bình luận, số lượng các câu chứa những từ có ký tự kéo dài như "hayyy" là khá nhiều. Vì vậy chúng tôi đã sử dụng bước 1 để xử lý, đưa "hayyy" về "hay", hoặc "chánnn" về "chán".

Tiếp theo ở bước 2, chúng tôi rà soát các trường hợp sai chính tả trong các câu bình luận. Sau đó sẽ thực hiện sửa lỗi chính tả. Các từ "ko", "khg", "khong" sẽ sửa thành "không". Hoặc "chăm Zn" sẽ sửa thành "trăm cảm". Tương tự đối với các từ sai chính tả khác, sẽ sửa lại thành từ viết đúng, ví dụ như bảng sau:



Từ sai chính tả	Từ viết đúng
ko, khg	không
chằm Zn, chằm zn, trầm zn	trầm cảm
ứng dụg, ứg dụg	ứng dụng
cũg, cũh	cũng

**Bảng 3. Sửa lỗi chính tả cho các từ**

Đối với các biểu tượng cảm xúc, chúng tôi sẽ quy về dạng cảm xúc mà biểu tượng đó thể hiện. Ví dụ, biểu tượng ☹ sẽ được quy về "tích cực". Còn biểu tượng ☺ sẽ được quy về "tiêu cực". Ở bước 4, các từ viết hoa sẽ được chuyển về dạng viết thường, để đồng bộ, giúp mô hình huấn luyện tốt hơn. Ví dụ, câu "Ứng dụng này Đỉnh của Đỉnh" sẽ chuyển thành "ứng dụng này đỉnh của đỉnh".

Hiện nay, có 2 loại mã Unicode được sử dụng phổ biến, Unicode tổ hợp và Unicode dựng sẵn. Nếu không xử lý vấn đề này, khi đưa vào mô hình, máy tính sẽ hiểu đó là các từ khác nhau mặc dù ta đang nhìn thấy chúng chẳng khác nhau gì. Vì thế bước 5 sẽ đưa về 1 chuẩn Unicode dựng sẵn (vì dạng này phổ biến hơn).

Cuối cùng, bước 6 sẽ loại bỏ dấu câu và các ký tự đặc biệt như : " \* @ \$ % & ... Vì các dấu ngắt câu và các ký tự đặc biệt khác không giúp ích trong việc phân loại một văn bản. Do đó, chúng tôi sẽ loại bỏ nó đi.

## 5 Phương pháp thử nghiệm

Trong bài báo này chúng tôi sẽ sử dụng 4 phương pháp để thử nghiệm trên bộ dữ liệu:

- **support vector machine** : SVM là mô hình máy học đơn giản để phân loại văn bản. Trong bài báo này chúng tôi sử dụng mô hình SVM kết TF-IDF embedding để so sánh với các mô hình transfer learning khác
- **BERT-Base, Multilingual Cased** : BERT (Bidirectional Encoder Representations from Transformers) là một mô hình học sẵn được tạo ra bởi Google AI. BERT được ứng dụng vào nhiều bài toán NLP khác nhau: Question Answering, Natural Language Inference,... với kết quả rất tốt.
- **PhoBERT**: Đây là một pre-trained được huấn luyện ành riêng cho tiếng Việt dựa trên kiến trúc và cách tiếp cận giống RoBERTa của Facebook được Facebook giới thiệu giữa năm 2019.
- **vibert4new** : là một pre-trained model được huấn luyện trên 20GB dữ liệu tin tức và đặt được những kết quả cao trên tiếng Việt.

## 6 Kết quả thử nghiệm

Kết quả thử nghiệm được trình bày trong bảng dưới đây

Mô hình	Precision	Recall	F1-score
SVM	41.17	32.74	36.47
PhoBert	77.86	42.91	55.33
vibert4new	76.45	49.87	60.37
Multilingual BERT cased	80.50	50.91	62.38

**Bảng 4. Kết quả thử nghiệm**

Các mô hình cho kết quả khá thấp trên bộ dữ liệu. Trong đó mô hình Multilingual BERT cho kết quả cao nhất với 62.38 điểm F1-score tiếp đến là các mô hình vibert4new, PhoBert và cuối cùng là SVM.

Dưới đây là kết quả trên từng nhãn của mô hình có kết quả cao nhất là Multilingual BERT cased

	precision	recall	f1-score	support
backup_negative	0.69	0.38	0.49	72
backup_neutral	0.00	0.00	0.00	11
backup_positive	0.91	0.88	0.89	152
display_negative	0.00	0.00	0.00	7
display_neutral	0.00	0.00	0.00	4
display_positive	0.00	0.00	0.00	7
edit_negative	0.00	0.00	0.00	6
edit_neutral	0.00	0.00	0.00	5
edit_positive	0.00	0.00	0.00	13
feature_negative	0.00	0.00	0.00	43
feature_neutral	1.00	0.02	0.04	46
feature_positive	0.67	0.06	0.11	35
general_negative	0.00	0.00	0.00	17
general_neutral	0.00	0.00	0.00	6
general_positive	0.78	0.74	0.76	187
other_negative	0.00	0.00	0.00	6
other_neutral	0.00	0.00	0.00	2
other_positive	0.00	0.00	0.00	4
stability_negative	0.68	0.60	0.64	90
stability_neutral	0.00	0.00	0.00	1
stability_positive	0.00	0.00	0.00	3
update_negative	0.97	0.80	0.88	40
update_neutral	0.00	0.00	0.00	2
update_positive	0.00	0.00	0.00	3

**Hình 5. Kết quả trên từng nhãn của mô hình Multilingual BERT cased**

Dựa vào bảng kết quả trên ta có thể thấy được mô hình dự đoán có kết quả bằng không trên nhiều nhãn, chỉ 6 nhãn là có kết quả khác không đó

là : backup-negative, backup-positive, feature-neutral, feature-positive, general-positive, updatenegative, stability-negative.

Một số lý do có thể giải thích cho kết quả trên đây :

- Dữ liệu có sự chênh quá lớn giữa các nhãn, phần lớn đều tập chung về hai khía cạnh là chung và sao lưu. Số lượng nhãn positive chiếm đa số trong dữ liệu.
- Phần lớn những nhãn mô hình cho về giá trị không đều có quas ít dữ liệu dẫn đến mô hình không thể học được các đặc trưng cần thiết để dự đoán.
- Độ dài của các câu bình luận quá ngắn, độ dài trung bình của các câu trong tập dữ liệu là gần 14 tiếng dẫn đến việc thiếu hụt thông tin đó việc phân biệt giữa các nhãn cảm xúc đem lại kết quả không cao

## 7 Kết luận và hướng phát triển

Trong bài báo này chúng tôi đã thử nghiệm các mô hình Bert khác nhau trên bộ dữ liệu cảm xúc cho các bình luận tiếng Việt của ứng dụng Google Photos tuy nhiên kết quả còn hạn chế. Trong đó mô hình Bert-base-multilingual có kết quả cao nhất (0.62 f1-score). Dữ liệu chênh lệch lớn giữa các nhãn khiến cho các mô hình hoạt động không tốt. Dự đoán trên các nhãn chiếm tỉ lệ thấp Không chính xác.

Trong tương lai chúng tôi hy vọng có thể cải thiện kết quả trên bộ dữ liệu bằng việc nghiên cứu và cải tiến các phương pháp tiền xử lý cùng với đó là cải thiện chất lượng của bộ dữ liệu và cuối cùng là thử nghiệm trên các mô hình SOTA để đem lại kết quả tốt hơn.

## Tài liệu

1. M. Hu and B. Liu, “Mining Opinion Features in Customer Reviews”, in Proceedings of Nineteenth National Conference on Artificial Intelligence(AAAI-2004), 2004
2. Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013.
3. Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014.
4. N. T. T. Thuy, N. X. Bach, and T. M. Phuong, “Crosslanguage aspect extraction for opinion mining,” in 2018 10th International Conference on Knowledge and Systems Engineering (KSE), IEEE, 2018, pp. 67–72
5. Kiritchenko, X. Zhu, C. Cherry, and S. Mohammad, “NRC-Canada-2014: Detecting aspects and sentiment in customer reviews,” in Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), Dublin, Ireland: Association for Computational Linguistics, Aug. 2014, pp. 437–442. [Online]. Available: <https://www.aclweb.org/anthology/S14-2076>
6. T. Dang, V. Nguyen, N. Kiet, and N. Ngan, “A transformation method for aspect-based sentiment analysis,” Journal of Computer Science and Cybernetics, vol. 34, no. 4, pp. 323–333, 2019. [Online]. Available: <http://vjs.ac.vn/index.php/jcc/article/view/13162>
7. J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “Bert: Pre-training of deep bidirectional transformers for language understanding,” arXiv preprint arXiv:1810.04805, 2018.

8. J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," arXiv preprint arXiv:1810.04805, 2018.
9. Khang Phuoc-Quy Nguyen and Kiet Van Nguyen. Exploiting Vietnamese Social Media Characteristics for Textual Emotion Recognition in Vietnamese. Khang Phuoc-Quy Nguyen and Kiet Van Nguyen. arXiv:2009.11005v3, 2020.
10. Dang Van Thin et al. Deep Learning for Aspect Detection on Vietnamese Reviews. 5th NAFOSTED Conference on Information and Computer Science, 2018.
11. H. S. Le, T. V. Le, and T. V. Pham, "Aspect Analysis for Opinion Mining of Vietnamese Text", in 2015 International Conference on Advanced Computing and Applications (ACOMP), 2015.
12. L. Mai and B. Le, "Aspect-Based Sentiment Analysis of Vietnamese Texts with Deep Learning", in Intelligent Information and Database Systems - 10th Asian Conference, ACIIDS 2018, Dong Hoi City, Vietnam, March 19-21, 2018, Proceedings, Part I, 2018.
13. T. D. Van, K. V. Nguyen, and N. L.-T. Nguyen, "NLP@UITat VLSP 2018: A Supervised Method for Aspect Based Sentiment Analysis", in Proceedings of the Fifth International Workshop on Vietnamese Language and Speech Processing (VLSP), 2018.
14. Do, H.T.T., Huynh, H.D., Van Nguyen, K., Nguyen, N.L.T., Nguyen, A.G.T.: Hate speech detection on vietnamese social media text using the bidirectional-lstm model. arXiv preprint arXiv:1911.03648, 2019.
15. Nguyen, L.T., Van Nguyen, K., Nguyen, N.L.T.: Constructive and toxic speech detection for open-domain social media comments in vietnamese. arXiv preprint arXiv:2103.10069, 2021.
16. Ho, V.A., Nguyen, D.H.C., Nguyen, D.H., Thi-Van Pham, L., Nguyen, D.V., Van Nguyen, K., Nguyen, N.L.T.: Emotion recognition for vietnamese social media text. In: International Conference of the Pacific Association for Computational Linguistics. pp. 319–333. Springer, 2019.
17. Maria Pontiki and Dimitris Galanis and Haris Papageorgiou and Suresh Manandhar and Ion Androutsopoulos, "Semeval-2015 task 12: Aspect based sentiment analysis", in Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015). The Association for Computational Linguistics, 2015.
18. Pontiki, M., Galanis, D., Pavlopoulos, J., Papageorgiou, H., Androutsopoulos, I., Manandhar, S.: SemEval-2014 task 4: Aspect based sentiment analysis. In: Proceedings of SemEval 2014). pp. 27–35. Association for Computational Linguistics, Dublin, Ireland, 2014.