

# ĐỀ XUẤT HỆ THỐNG NHẬN DIỆN CẢM XÚC ĐỂ GỢI Ý VÀ PHẢN HỒI TIẾNG VIỆT TRONG HỘI THOẠI MẠNG SỬ DỤNG MÔ HÌNH PhoBERT

Nguyễn Hữu Phát\*, Hoàng Phi Long†

\*Khoa Điện Tử, Trường Điện-Điện Tử, Đại học Bách Khoa Hà Nội

† Trường Điện-Điện Tử, Đại học Bách Khoa Hà Nội

**Tóm tắt**—Bài báo đề xuất một hệ thống nhận diện cảm xúc nhằm gợi ý cho con người và tự động phản hồi trong hội thoại với mục đích tăng trải nghiệm của người dùng và phản hồi ngay lập tức. Chúng tôi đề xuất xây dựng một hệ thống nhận diện cảm xúc của người dùng bao gồm 6 loại cảm xúc (disgust, pleasure, sad, fear, anger, other) bằng mô hình PhoBERT. Bên cạnh đó, chúng tôi sử dụng bộ dữ liệu được xây dựng trong bài báo [1] với mục tiêu đối chiếu và chứng minh sự cải thiện từ mô hình mà chúng tôi đề xuất. Đầu vào của hệ thống là đoạn văn bản. Đầu ra của hệ thống là nhãn cảm xúc tương ứng với đầu vào. Kết quả mô phỏng của hệ thống cho độ chính xác lên đến 99.8% với bộ dữ liệu huấn luyện và 96.4% với bộ dữ liệu kiểm thử. Kết quả này cải thiện được khoảng 15% so với bài báo trước đó cho thấy chúng tôi có thể xây dựng một hệ thống nhận diện cảm xúc chính xác hơn cho người dùng và có khả năng ứng dụng trong thực tế.

**Từ khóa**—NLP, PhoBERT, sentiment analysis, deep learning, emotional classification.

## I. ĐẶT VẤN ĐỀ

Trong những năm gần đây, Việt Nam được coi là một nước có thị trường thương mại điện tử phát triển nhanh nhất ở Đông Nam Á với tỷ lệ tăng trưởng trung bình hàng năm lên đến 18% [2].

Trong một sàn thương mại điện tử, có rất nhiều yếu tố cần thiết để giữ chân cũng như tăng tỷ lệ duy trì của khách hàng. Trong những yếu tố đó bao gồm hai yếu tố quan trọng, đó là trải nghiệm người dùng và tốc độ phản hồi của người bán hàng. Tuy nhiên, hiện tại phần lớn các sàn thương mại điện tử tập trung vào gợi ý sản phẩm để tăng doanh số và sử dụng chatbot để phản hồi ngay lập tức mà còn thiếu sự quan tâm tới cảm xúc của người dùng. Chính vì vậy, chúng tôi đề xuất xây dựng một hệ thống nhận diện cảm xúc của người dùng trong hội thoại để giải quyết vấn đề trên.

Hệ thống mà chúng tôi đề xuất gồm hai phần chính cần quan tâm cụ thể là:

- Đầu tiên, chúng tôi thực hiện tiền xử lý bộ dữ liệu để loại bỏ các dữ liệu không cần thiết.

- Thứ hai, chúng tôi sử dụng mô hình PhoBERT để huấn luyện với dữ liệu đã qua tiền xử lý ở trên.

Phần còn lại của bài báo bao gồm bốn phần được tổ chức như sau. Trong phần II, chúng tôi sẽ thảo luận về các nghiên cứu liên quan. Trong phần III, chúng tôi sẽ trình bày về hệ thống được đề xuất. Phần IV sẽ đánh giá hệ thống được đề xuất và phân tích kết quả. Cuối cùng, chúng tôi sẽ đưa ra kết luận và những định hướng nghiên cứu trong tương lai.

## II. CÁC NGHIÊN CỨU LIÊN QUAN

Hiện nay, đã có nhiều nghiên cứu liên quan đến chủ đề nhận diện cảm xúc của con người. Ở Việt Nam, có một vài bộ dữ liệu được xây dựng nhằm huấn luyện và xây dựng ra một mô hình nhận diện cảm xúc. Các tác giả của những bộ dữ liệu trên sử dụng những mô hình khác nhau để huấn luyện và đã cho những kết quả đáng kể.

Trong những nghiên cứu về tiếng Việt, với hai nhãn positive và negative trong bài báo [3], tác giả sử dụng hai bộ dữ liệu là Ntc-sv gồm 50,760 mẫu với độ chính xác cao nhất là 91.15% và Vreview gồm 57,612 mẫu với độ chính xác lên đến 88.22% sử dụng mô hình BERT-RCNN. Với ba nhãn positive, negative và neutral, tác giả trong [4] sử dụng bộ dữ liệu trích xuất từ những đánh giá của người Việt về 50 khách sạn ở Việt Nam từ Agoda. Bộ dữ liệu trên gồm số câu tương ứng với ba nhãn trên là 1980,777,547. Bài báo trên đạt được độ chính xác lên đến 76.8% sử dụng mô hình SVM. Với sáu nhãn disgust, pleasure, sad, fear, anger, other, trong [1] chúng tôi xây dựng bộ dữ liệu bao gồm 7,242 mẫu và đạt được độ chính xác cao nhất là 86.8% trên tập huấn luyện và 81.4% trên tập kiểm thử.

Trong những nghiên cứu quốc tế, với hai nhãn positive và negative trong bài báo [5], tác giả thử nghiệm với nhiều bộ dữ liệu và cho độ chính xác lên đến 83.9% với mô hình SVM (Support Vector Machine). Trong [6], các tác giả đã có một bài viết đánh giá phân tích tình cảm và phát hiện cảm xúc từ văn bản trên các bài báo khác nhau để có thể so sánh chúng một cách khách quan. Trong bài báo này tác giả đã đưa ra được rất nhiều bài báo có kết quả khá tốt trong đó có bài nghiên cứu “Comparative Analyses of BERT, RoBERTa, DistilBERT, and XLNet for Text-based Emotion Recognition” với bộ dữ liệu gồm 7 nhãn tương ứng với 7666 câu đạt được độ chính xác cao nhất lên đến 74.31% với mô hình RoBERTa. Trong [7], tác giả sử dụng bộ dữ liệu CrowdFlower gồm 40,000 câu với 6 nhãn cảm

## Đề xuất hệ thống nhận diện cảm xúc đề nghị và phản hồi trong hội thoại mạng

xúc và đạt được độ chính xác lên đến 63.2%, bộ dữ liệu khác gồm 1,387,787 câu với 7 nhãn cảm xúc đạt được độ chính xác lên đến 80.4% với mô hình được thiết kế dựa trên mạng RNN.

Như đã phân tích ở trên, có một vài bộ dữ liệu đã được xây dựng nhằm giải quyết chủ đề nhận diện cảm xúc của con người. Tuy nhiên, với sáu nhãn cảm xúc trở lên, độ chính xác cao nhất đối với tiếng Việt chỉ đạt được đến 81.4% và với tiếng Anh là 80.4%. Chính vì vậy, chúng tôi đề xuất xây dựng một hệ thống nhận diện cảm xúc sử dụng mô hình PhoBERT cùng với bộ dữ liệu gồm sáu nhãn cảm xúc ở trên nhằm mục đích cải thiện khả năng gợi ý và phản hồi trong hội thoại mạng. Trọng phạm vi hiểu biết của chúng tôi chưa tìm thấy hệ thống tương tự như hệ thống chúng tôi đề xuất.

- Trong bước dự đoán, mô hình đã được huấn luyện sẽ được sử dụng để phân loại cảm xúc của dữ liệu chưa được gán nhãn. Đầu ra có thể là cảm xúc của đầu vào hoặc tỉ lệ của sáu nhãn cảm xúc đối với đầu vào.

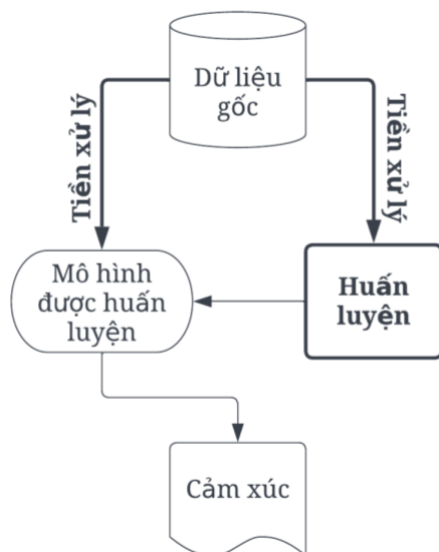
*Đóng góp chính của chúng tôi đối với bài báo này là thay đổi nguyên tắc tiền xử lý và tinh chỉnh mô hình như trong hình 1.*

### B. Chi tiết hệ thống

Hình 2 khái quát chi tiết cách thức huấn luyện mô hình nhận diện cảm xúc và thực hiện dự đoán. So với [1], chúng tôi đã cải thiện quá trình tiền xử lý để dữ liệu đầu vào của mô hình ít nhiễu và mang nhiều ý nghĩa hơn. Ngoài ra, chúng tôi sử dụng một phiên bản khác của PhoBERT là PhoBERT<sub>large</sub> đồng thời tinh chỉnh các siêu tham số khác giúp mô hình được huấn luyện tốt hơn. Những thay đổi trên sẽ được trình bày ở phần C.

## III. HỆ THỐNG ĐỀ XUẤT

### A. Tổng quan về hệ thống



Hình 1. Mô hình tổng quan hệ thống đề xuất.

Hình 1 là mô hình tổng quan các khối chính của hệ thống đề xuất dựa trên tài liệu [1]. Hệ thống gồm hai bước chính là huấn luyện và dự đoán.

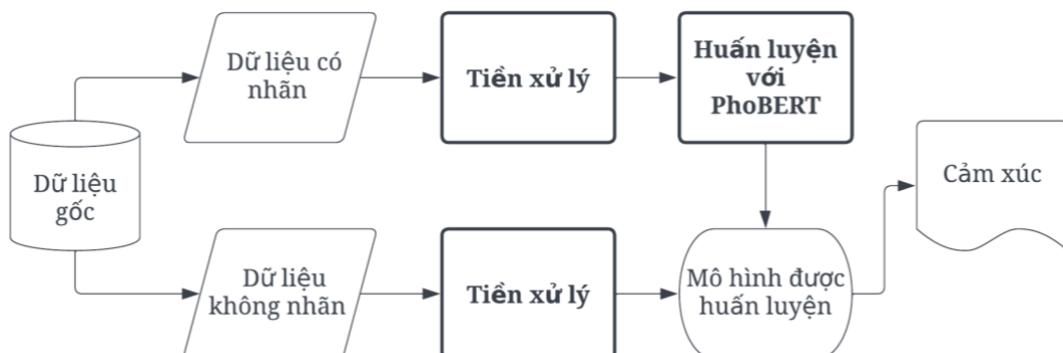
- Trong bước huấn luyện, chúng tôi sẽ tiền xử lý những dữ liệu đã được gán nhãn và huấn luyện mô hình với PhoBERT.

### 1. Nhãn của các cảm xúc

Theo [1], bộ dữ liệu được đánh nhãn dựa trên phân tích cơ bản về cảm xúc của con người trong [8]–[11]. Có sáu nhãn cảm xúc cơ bản được miêu tả tương ứng như sau.

- Disgust: Nó thường xảy ra để phản ứng với tình huống khiến bạn không thoải mái hoặc không mong muốn. Nó có thể khiến chúng ta cảm thấy không thích, không đồng ý, phản kháng, và hoảng sợ.
- Pleasure: Nó là một cảm giác hài lòng.
- Sad: Nó được định nghĩa như một trạng thái cảm xúc ngăn ngại được cá nhân hoá bởi cảm giác bức bối, buồn bã, tuyệt vọng, mất hứng thú, và trầm cảm.
- Fear: Một cảm xúc mạnh mẽ quan trọng trong sinh tồn. Khi chúng ta đối mặt với một vài dạng nguy hiểm hoặc đe dọa, nó có thể có hại nếu nguy hiểm mang tính vật lý hoặc tinh thần
- Anger: Một cảm xúc mạnh mẽ cụ thể được cá nhân hoá bởi cảm giác thù địch, kích động tới bức bối, và phản kháng. Phụ thuộc vào cường độ, mức độ thay đổi từ hơi không thoải mái tới giận dữ và tức giận.
- Other: là những cảm xúc khác hoặc không có biểu hiện gì.

Bảng I dưới đây mô tả một số mẫu dữ liệu cùng với nhãn cảm xúc tương ứng.



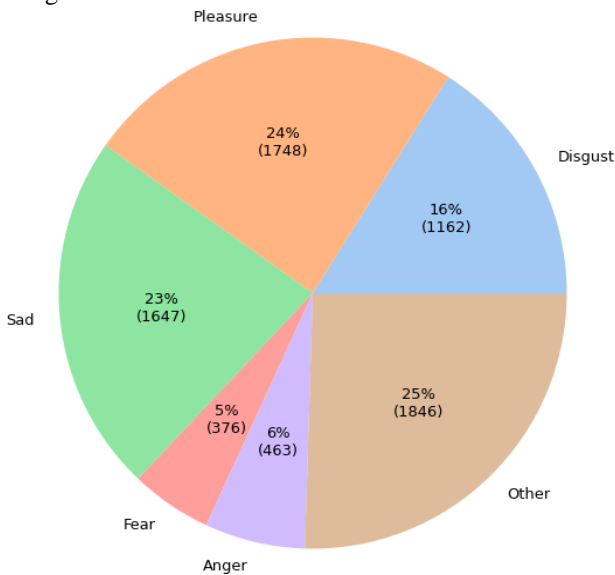
Hình 2. Sơ đồ hoạt động chi tiết của mô hình

Bảng 1. Các mẫu dữ liệu cùng với nhãn tương ứng

STT	Dữ liệu	Nhãn
1	Chán ông chồng này lắm rồi, suốt ngày rượu chè, chả giúp được gì cả.	Disgust
2	Tháng này giảm đc hẳn 2kg	Pleasure
3	Tạm biệt nhé :((	Sad
4	Hôm qua tao bị chó đuổi chạy mất cả dép!!	Fear
5	Láo thật	Anger
6	Thứ 2 là ngày đầu tuần.	Other

## 2. Dữ liệu đào tạo

Bộ dữ liệu mà chúng tôi sử dụng gồm 6 nhãn cảm xúc với số lượng mẫu dữ liệu mỗi nhãn tương ứng được mô tả trong hình 3.



Hình 3. Minh họa bộ dữ liệu đào tạo

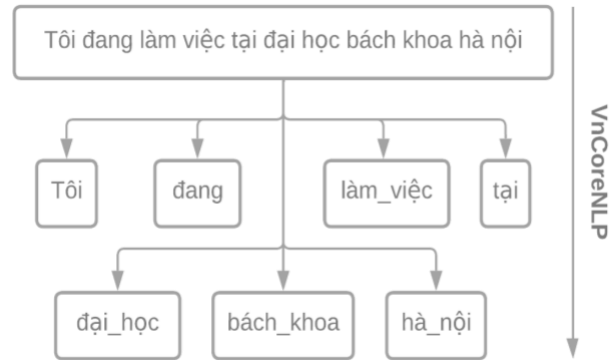
## C. Huấn luyện

Quá trình huấn luyện bao gồm hai bước là tiền xử lý và huấn luyện sẽ được trình bày sau đây.

### 1. Tiền xử lý

Tiền xử lý là một bước vô cùng quan trọng trong xử lý ngôn ngữ tự nhiên (NLP) nói chung hay nhận diện cảm xúc nói riêng. Dữ liệu trong bộ dữ liệu trích xuất được ở dạng thô, vì vậy cần được xử lý, làm sạch và biến đổi trước khi huấn luyện chúng. Quá trình này bao gồm **chín bước** sau.

- **Tách từ:** Là quá trình xác định ranh giới của các từ trong câu, cũng có thể hiểu là quá trình xác định các từ đơn, từ ghép... có trong câu. Đối với xử lý ngôn ngữ tự nhiên, để có thể xác định cấu trúc ngữ pháp của câu, từ loại của một từ trong câu thì cần phải xác định được đâu là từ trong câu. Vấn đề này đơn giản đối với con người nhưng lại vô cùng khó giải quyết đối với máy tính. Trong bài báo này, chúng tôi sử dụng bộ công cụ VnCoreNLP [12] để tách từ. Quá trình này được mô tả trong hình 4 dưới đây.



Hình 4. Quá trình tách từ bằng bộ công cụ VnCoreNLP

- **Loại bỏ từ dừng:** Từ dừng là những từ không bổ sung nghĩa cho câu. Chúng có thể được bỏ qua mà không làm mất đi ý nghĩa của câu. Đối với một số công cụ tìm kiếm, từ dừng là một số từ chức năng ngắn, phổ biến và sẽ bị loại bỏ hoàn toàn. Trong bài báo này, chúng tôi sử dụng bộ từ điển Vietnamese-Stopword [13] để loại bỏ từ dừng.
- **Loại bỏ dấu câu:** Dấu câu là các ký hiệu được quy ước để hỗ trợ cho việc đọc và hiểu văn bản. Tuy nhiên, đối với máy tính, việc loại bỏ dấu câu sẽ giúp máy tính đối xử với các từ trong văn bản một cách công bằng hơn. Các dấu câu bị loại bỏ bao gồm [.,\/#!\$%^&\\*::;{}|=|\_~()]
- **Loại bỏ URL:** URL là các đường dẫn đến các địa chỉ web, không mang ý nghĩa cảm xúc.
- **Loại bỏ biểu tượng cảm xúc:** Biểu tượng cảm xúc là một hình ảnh đại diện của các biểu hiện trên gương mặt bằng cách sử dụng những ký tự. Trong văn bản, biểu tượng cảm xúc có thể bị sai lệch và cần bị loại bỏ để đối xử với các từ trong văn bản một cách công bằng hơn.
- **Loại bỏ các số:** Các số không mang ý nghĩa cảm xúc nên cần bị loại bỏ.
- **Sửa lại teencode:** Teencode là một thuật ngữ dùng để chỉ một dạng chính tả không chính thức khác. Công dụng chính của dạng chữ viết này là dùng để mã hóa. Do dữ liệu được trích xuất từ mạng xã hội và các bài viết nên chứa một lượng teencode khá lớn, chính vì vậy, teencode cần được sửa đúng để được đối xử đúng như những từ đúng chính tả khác. Trong bài báo này, chúng tôi sử dụng danh sách một số teencode phổ biến thống kê từ dữ liệu [14] để sửa lại teencode.
- **Loại bỏ từ hiếm/từ sai:** Từ hiếm/sai là những từ nằm ngoài từ điển. Thường không mang ý nghĩa trong câu. Trong bài báo này, chúng tôi sử dụng từ điển Viet74K [15] để loại bỏ những từ nằm ngoài từ điển Việt Nam.

## Đề xuất hệ thống nhận diện cảm xúc đề nghị và phản hồi trong hội thoại mạng

- Chuẩn hoá viết thường: Đối với máy tính, “Tôi” và “tôi là hai từ khác nhau do nó phân biệt chữ hoa và chữ thường. Tuy nhiên trong ngôn ngữ tự nhiên, nó lại mang cùng một ý nghĩa. Chính vì vậy, văn bản cần phải được chuẩn hoá viết thường.

### 2. Huấn luyện

Hiện nay, để giải quyết những vấn đề liên quan đến xử lý ngôn ngữ tự nhiên (NLP) nói chung và nhận diện cảm xúc nói riêng, có rất nhiều mô hình huấn luyện như mạng nơ ron hồi quy (RNN) hay mạng bộ nhớ ngắn-dài (LSTM) được sử dụng. Tuy nhiên, những mô hình này không mang lại kết quả tốt đối với bài toán nhận diện cảm xúc trong tiếng Việt mà chúng tôi muốn giải quyết. Chính vì vậy, trong bài báo này, chúng tôi quyết định sử dụng mô hình PhoBERT [16] phiên bản **PhoBERTlarge**.

Mô hình PhoBERT có một số đặc điểm nổi bật như sau.

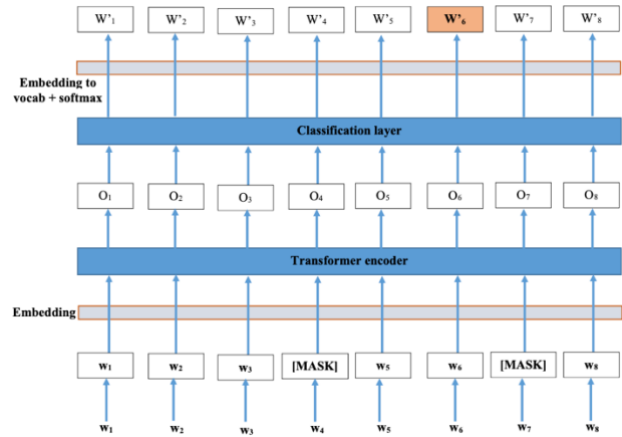
- Là một pre-trained được huấn luyện đơn ngôn ngữ cho ngôn ngữ Tiếng Việt, việc huấn luyện dựa trên kiến trúc và cách tiếp cận giống RoBERTa của Facebook được giới thiệu giữa năm 2019. RoBERTa được cải tiến hơn so với mô hình BERT.
- Có hai phiên bản là PhoBERT<sub>base</sub> với 12 transformers block và PhoBERT<sub>large</sub> với 24 transformers block.
- PhoBERT được train trên khoảng 20GB dữ liệu bao gồm 1GB dữ liệu Vietnamese Wikipedia và 19GB dữ liệu còn lại được lấy từ Vietnamese news. Đây là một lượng dữ liệu khá ổn để train mô hình có kiến trúc như BERT.
- Mô hình BERT [17]–[20] sử dụng Transformer là một mô hình attention (attention mechanism) học mối tương quan giữa các từ (hoặc 1 phần của từ) trong một văn bản. Transformer gồm có 2 phần chính: Encoder và Decoder, encoder thực hiện đọc dữ liệu đầu vào và decoder đưa ra dự đoán. BERT chỉ sử dụng Encoder của mô hình transformer.
- Khác với các mô hình directional (các mô hình chỉ đọc dữ liệu theo 1 chiều duy nhất: trái→phải, phải→trái) đọc dữ liệu theo dạng tuần tự, Encoder đọc toàn bộ dữ liệu trong 1 lần, việc này làm cho BERT có khả năng huấn luyện dữ liệu theo cả hai chiều, qua đó mô hình có thể học được ngữ cảnh (context) của từ tốt hơn bằng cách sử dụng những từ xung quanh nó (phải và trái).

Mô hình PhoBERT được huấn luyện bao gồm hai bước là Masked LM (MLM) và Next Sentence Prediction (NSP) sẽ được trình bày sau đây.

#### 2.1. Masked LM (MLM)

Trước khi đưa vào BERT, 15% số từ trong văn bản được thay thế bởi token [MASK], khi đó mô hình sẽ dự đoán từ được thay thế bởi [MASK] với ngữ cảnh là các từ không bị thay thế. Quá trình xử lý được mô tả trong hình 5 gồm những bước sau.

- Thêm một lớp phân loại với đầu vào là đầu ra của Encoder.
- Nhân các vector đầu ra với ma trận embedding để đưa chúng về không gian từ vựng (vocabulary dimensional).
- Tính toán xác suất của mỗi từ trong tập từ vựng sử dụng hàm softmax.



Hình 5. Masked LM trong mô hình BERT [1]

#### 2.2. Next Sentence Prediction (NSP)

Tại bước này, đầu vào của mô hình là một cặp câu. Mô hình sẽ dự đoán câu thứ hai có phải câu tiếp theo của câu thứ nhất không. Trong quá trình huấn luyện, 50% dữ liệu đầu vào là cặp câu trong đó câu thứ hai đúng là câu tiếp theo của câu thứ nhất, 50% còn lại thì câu thứ hai được chọn ngẫu nhiên từ tập dữ liệu. Một số nguyên tắc xử lý dữ liệu được mô tả trong hình 6 và trình bày sau đây.

- Chèn token [CLS] vào trước câu đầu tiên và [SEP] vào cuối mỗi câu.
- Các token trong từng câu được đánh dấu là A hoặc B.
- Chèn thêm vector embedding biểu diễn vị trí của token trong câu.



Hình 6. NSP trong mô hình BERT

Các bước xử lý trong Next Sentence Prediction được trình bày sau đây.

- Toàn bộ câu đầu vào được đưa vào Transformer.
- Chuyển vector đầu ra của [CLS] về kích thước 2x1 bằng một lớp phân loại.
- Tính xác suất là câu tiếp theo bằng softmax.

### 3. Thuật toán tối ưu

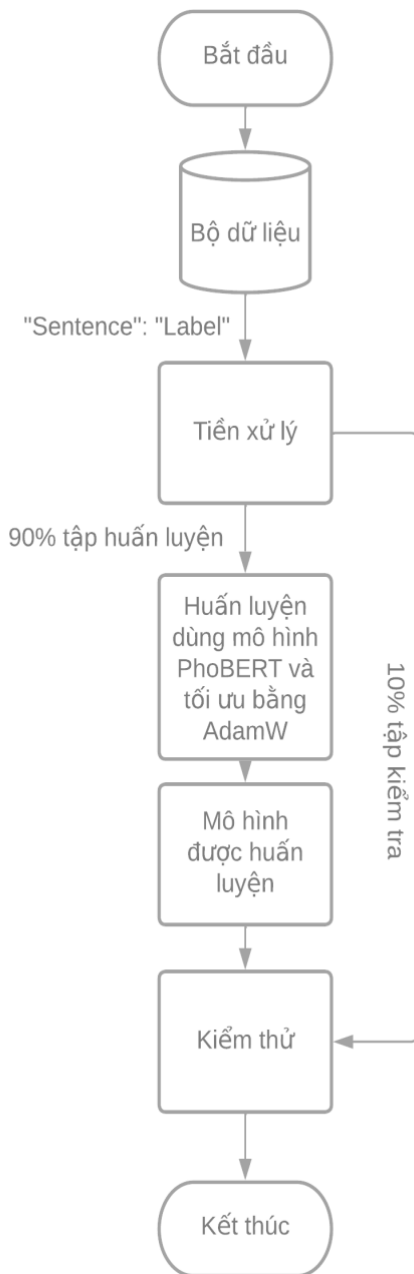
Có rất nhiều thuật toán tối ưu có thể sử dụng với PhoBERT. Tuy nhiên, trong bài báo này, chúng tôi quyết định sử dụng thuật toán tối ưu là AdamW [21] với các siêu tham số như sau.

- Lr: **1e-6**
- Betas (b1,b2): **(0.9, 0.999)**
- Eps: **1e-6**

- Weight\_decay: **0.01**
- Correct\_bias: **True**

#### 4. Lưu đồ thuật toán huấn luyện

Hình 7 cho thấy lưu đồ thuật toán chúng tôi sử dụng để huấn luyện. Chúng tôi đưa đầu vào huấn luyện là bộ dữ liệu với nhãn cảm xúc tương ứng. Tiếp theo, chúng tôi tiền xử lý dữ liệu đầu vào như đã trình bày ở trên. Sau đó, chúng tôi chia bộ dữ liệu với **90%** dành cho huấn luyện và **10%** dành cho kiểm thử. Tiếp theo, chúng tôi sử dụng mô hình PhoBERT để huấn luyện và tối ưu mô hình bằng thuật toán AdamW. Cuối cùng, chúng tôi kiểm thử mô hình với 10% dữ liệu đã chia ở trên. Chi tiết hơn về kết quả sẽ được trình bày ở phần tiếp theo.



Hình 7. Lưu đồ thuật toán huấn luyện

## IV. KẾT QUẢ

### A. Tiêu chí đánh giá

Để đánh giá độ chính xác của mô hình, chúng tôi sử dụng công thức sau.

$$Accuracy = \frac{1}{n} \sum_{i=1}^{n-1} (\hat{y}_i = y_i) \quad (1)$$

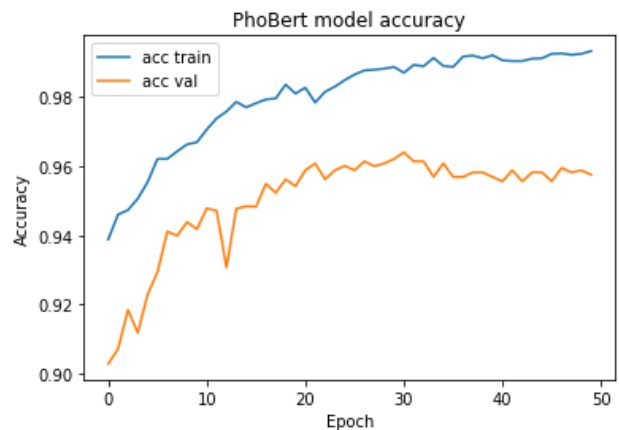
Trong đó,  $n$  là số lượng nhãn,  $\hat{y}_i$  là giá trị dự đoán bởi mô hình, và  $y_i$  là giá trị thực tế tương ứng.

Để đánh giá độ chính xác, chúng tôi sử dụng GoogleColab Pro với cấu hình là Intel(R) Xeon(R) CPU @ 2.00GHz, 4 lõi, 24 GB RAM. Bên cạnh đó, chúng tôi sử dụng bộ dữ liệu ở bài báo [1] với tỷ lệ khác là **90%** cho tập huấn luyện và **10%** cho tập kiểm tra nhằm mục đích nhiều dữ liệu hơn cho việc huấn luyện.

Để thử nghiệm huấn luyện và đạt được kết quả tốt nhất, chúng tôi thực hiện thuật toán huấn luyện đã trình bày ở trên nhiều lần. Trong mỗi lần thực hiện, chúng tôi thay đổi một tham số trong số các siêu tham số như tốc độ học, batch size, độ dài lớn nhất của mỗi mẫu đầu vào để tìm ra được giá trị tối ưu nhất cho siêu tham số đó và lặp đi lặp lại cho đến khi tìm được giá trị tối ưu nhất cho tất cả các siêu tham số có thể tinh chỉnh.

### B. Kết quả

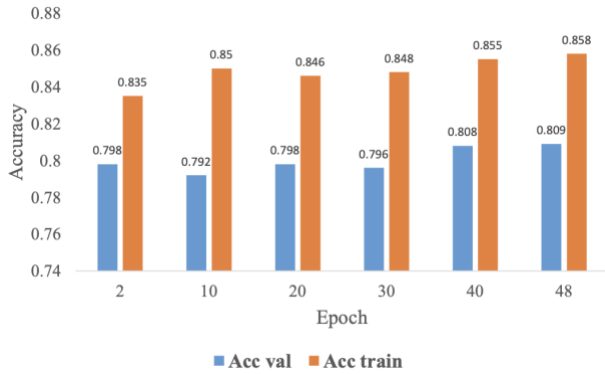
Kết quả huấn luyện mô hình mà chúng tôi đề xuất so với mô hình ở bài báo [1] và mô hình nhận diện cảm xúc trong tiếng Anh ở bài báo [7] được thể hiện trong hình 8, hình 9 và bảng II. Thời gian huấn luyện mô hình mà chúng tôi đề xuất so với mô hình ở bài báo [1] được thể hiện trong hình 10 và hình 11. Với mô hình PhoBERT, chúng tôi đạt được độ chính xác cao nhất với bộ dữ liệu huấn luyện là 99.8% và với bộ dữ liệu kiểm thử là 96.4%. Kết quả này cải thiện được khoảng 15% so với bài báo [1] trước đó và cao hơn so với bài báo [7] khoảng 16%. Về thời gian thực hiện huấn luyện, do sử dụng phiên bản PhoBERT<sub>large</sub> có 24 transformers block, nhiều hơn so với PhoBERT<sub>base</sub> có 12 transformers block nên thuật toán trở nên phức tạp và mất nhiều thời gian hơn để huấn luyện lên đến 350 phút, nhiều hơn so với thời gian huấn luyện trong [1] 170 phút. Tuy nhiên, độ chính xác của mô hình được huấn luyện được cải thiện khá tích cực. Chính vì vậy, chúng tôi sử dụng mô hình PhoBERT để thực hiện ứng dụng trên website nhằm gợi ý và phản hồi trong hội thoại mạng.



Hình 8. Độ chính xác của kết quả huấn luyện mô hình PhoBERT



## Đề xuất hệ thống nhận diện cảm xúc đề gợi ý và phản hồi trong hội thoại mạng



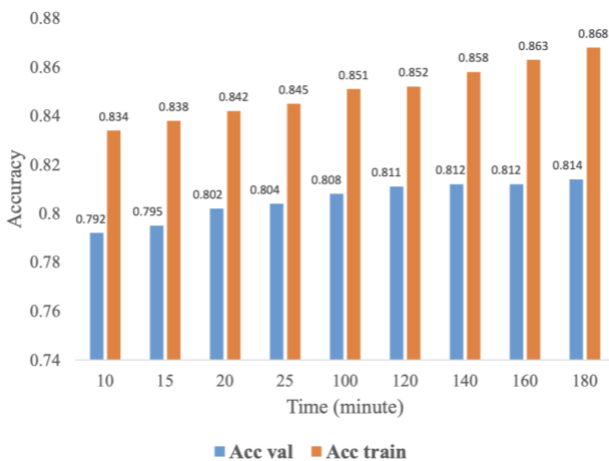
Hình 9. Độ chính xác của kết quả huấn luyện mô hình PhoBERT trong bài báo [1]



Hình 10. Thời gian huấn luyện mô hình PhoBERT

Bảng II. Độ chính xác của kết quả huấn luyện trong bài báo [7]

Bộ dữ liệu	Số lượng nhãn	Độ chính xác
Wang et al	7	80.4%
CrowdFlower	6	63.2



Hình 11. Thời gian huấn luyện mô hình PhoBERT trong bài báo [1]

### C. Thảo luận

Mặc dù nhận được kết quả khá tích cực cho hệ thống và cải thiện chính xác hơn so với bài báo trước, tuy nhiên, thời gian huấn luyện mô hình lên đến 350 phút, lâu hơn so với bài báo trước là 180 phút và lâu hơn nhiều so với những mô hình khác như RNN, LSTM. Trong một số trường hợp chạy thực tế, những văn bản có nhãn “Fear” hay “Anger” vẫn còn thiếu chính xác do sự mất cân bằng trong bộ dữ liệu. Để cải thiện những hạn chế này, hướng tiếp theo chúng tôi sẽ thực hiện các bước như đơn giản hoá mô hình, tìm những mô hình khác có thể cho ra kết quả tương đương hoặc tốt hơn để thay thế hay bổ sung dữ liệu để đạt được sự cân bằng tốt hơn.

### V. KẾT LUẬN

Bài báo tập trung vào việc cải thiện độ chính xác trong tác vụ nhận diện cảm xúc sử dụng mô hình PhoBERT và thuật toán tối ưu AdamW so với bài báo [1] trước và đã đạt được kết quả tốt hơn khoảng 15%.

Tuy nhiên hệ thống vẫn còn nhược điểm như thời gian huấn luyện dài hay dự đoán sai đối với một số nhãn. Lý do có thể là mất cân bằng dữ liệu và mô hình PhoBERT là mô hình có độ phức tạp cao. Trong tương lai, chúng tôi sẽ tiếp tục bổ sung dữ liệu để cân bằng bộ dữ liệu cũng như cố gắng giảm thời gian huấn luyện và cải thiện độ chính xác.

### LỜI CẢM ƠN

Chúng tôi muốn dành lời cảm ơn đặc biệt và trân trọng nhất tới Đại học Bách Khoa Hà Nội (HUST) vì sự hỗ trợ thời gian cũng như cơ sở vật chất vô cùng nhiệt huyết dành cho nghiên cứu này.

### TÀI LIỆU THAM KHẢO

- [1] N. Huu Phat, N. Anh Tuan, H. Phi Long, and D. Dang Dang, “Proposing System to Recognize Emotions in Public Network Using Phobert Deep Learning Model,” in 2022 9th NAFOSTED Conference on Information and Computer Science (NICS), 2022, pp. 276–289.
- [2] N.T Thanh Thu, “Factors Affecting The Behavior Of E-Commerce Using: A Case Study In Vietnam,” in 2022, Journal of Positive School Psychology, vol. 6, 2022.
- [3] Q. T. Nguyen, T. L. Nguyen, N. H. Luong, and Q. H. Ngo, “Fine-Tuning BERT for Sentiment Analysis of Vietnamese Reviews,” in Proceedings - 2020 7th NAFOSTED Conference on Information and Computer Science, NICS 2020, Nov. 2020, pp. 302–307, doi: 10.1109/NICS51282.2020.9335899.
- [4] N. T. Duyen, N. X. Bach, T. M. Phuong, “An empirical study on sentiment analysis for Vietnamese,” in 2014 International conference on advanced technologies for communications (ATC), pp 309-314.
- [5] C. S. Khoo and S. B. Johnkhan, “Lexicon-based sentiment analysis: Com-parative evaluation of six sentiment lexicons,” J. Inf. Sci., vol. 44, no. 4, pp. 491–511, Aug. 2018.
- [6] "Nandwani, P., Verma, R. A review on sentiment analysis and emotion detection from text. Soc. Netw. Anal. Min. 11, 81 (2021). <https://doi.org/10.1007/s13278-021-00776-6>".
- [7] A. Seyeditabari, N. Tabari, S. Gholizadeh, W. Zadrozny, Emotion detection in text: focusing on latent representation, 2019.
- [8] K. Cherry, “The 6 types of basic emotions and their effect on hu- man behavior,” in Version 1, 2021, retrieved from official website: <https://www.verywellmind.com>.
- [9] A. Mohanta and V. K. Mittal, “Human emotional states classification based upon changes in speech production features in vowel regions,” in 2017 2nd International

- Conference on Telecommunication and Networks (TEL-NET), 2017, pp. 1-6.
- [10] S. S. Shanta, M. Sham-E-Ansari, A. I. Chowdhury, M. M. Shahriar, and M. K. Hasan, "A comparative analysis of different approach for basic emotions recognition from speech," in *2021 International Conference on Electronics, Communications and Information Technology (ICECIT)*, 2021, pp. 1-4.
- [11] W. S. S. Khine, P. Siritanawan, and K. Kotani, "Generation of compound emotions expressions with emotion generative adversarial networks (emogans)," in *2020 59th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE)*, 2020, pp. 748-755.
- [12] Thanh Vu, Dat Quoc Nguyen, Dai Quoc Nguyen, Mark Dras, and Mark Johnson. 2018. "VnCoreNLP: A Vietnamese Natural Language Processing Toolkit." In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations*, pages 56-60, New Orleans, Louisiana. Association for Computational Linguistics.
- [13] V.-D. Le, "Python vietnamese toolkit," in *Accessed 25, June 2022*, 2022, retrieved from official website: <https://github.com/stopwords/vietnamese-stopwords>.
- [14] N. Van Hieu, "Danh sách một số teencode phổ biến thống kê từ dữ liệu," retrieved from official website: <https://gist.github.com/nguyenvanhieuvn/7d9441c10b3c2739499fc5a4d9ea06fb>
- [15] H. N. Duc, "Vietnamese word list," retrieved from official website: <http://www.informatik.uni-leipzig.de/~duc/software/misc/wordlist.html>
- [16] D. Q. Nguyen and A. Nguyen, "Phobert: Pre-trained language models for vietnamese," Jan. 2020, pp. 1037-1042.
- [17] S. Liu, H. Tao, and S. Feng, "Text classification research based on bert model and bayesian network," in *2019 Chinese Automation Congress (CAC)*, 2019, pp. 5842-5846.
- [18] A. Gillioz, J. Casas, E. Mugellini, and O. A. Khaled, "Overview of the transformer-based models for nlp tasks," in *2020 15th Conference on Computer Science and Information Systems (FedCSIS)*, 2020, pp. 179-183.
- [19] J. Dong, F. He, Y. Guo, and H. Zhang, "A commodity review sentiment analysis based on bert-cnn model," in *2020 5th International Conference on Computer and Communication Systems (ICCCS)*, 2020, pp. 143-147.
- [20] M. Ramina, N. Darnay, C. Ludbe, and A. Dhruv, "Topic level summary generation using bert induced abstractive summarization model," in *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*, 2020, pp. 747-752.
- [21] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101, 2017.

## PROPOSING EMOTIONAL CLASSIFICATION SYSTEM FOR RECOMMENDATION AND RESPONDING IN VIETNAMESE NETWORK COMMUNICATION USING PHOBERT MODEL

### Abstract:

The article proposes a system of emotional classification applying to recommend user chatting and automatically respond in network communication for improve user experience and instantly respond. We propose to use PhoBERT model to build an emotional classification system including 6 classes of emotion (disgust, pleasure, sad, fear, anger, other). Besides, we use the dataset built in the article [1] which is published earlier. We use this dataset for comparison and prove our improvement of our proposed model. The input of model

is text and the output of model is respective emotion label. The result have shown an accuracy of 99.8% on the training dataset and 96.4% on the validation dataset. This result also show that our proposed model is more accurate than the before paper by 15%. Therefore, we can build an emotion classification system more accuracy for user that applicable in real life.

**Keywords**—NLP, PhoBERT, sentiment analysis, deep learning, emotional classification.



**Nguyen Huu Phat**, nhận bằng kỹ sư (2003), thạc sỹ (2005) ngành Điện tử và Viễn thông tại Đại học Bách Khoa Hà Nội (HUST), Việt Nam và bằng tiến sĩ (2012) về Khoa học Máy tính tại Viện Công nghệ Shibaura, Nhật Bản. Hiện tại, đang là giảng viên tại Viện Điện tử Viễn thông, HUST, Việt Nam. Các nghiên cứu gồm xử lý hình ảnh và video, mạng không dây, big data, hệ thống giao thông thông minh (ITS), và internet của vạn vật (IoT). Ông đã nhận được giải thưởng bài báo hội nghị tốt nhất trong SoftCOM (2011), giải thưởng tài trợ sinh viên tốt nhất trong APNOMS (2011), giải thưởng danh dự của Viện Công nghệ Shibaura (SIT).



**Hoàng Phi Long** Hiện tại là sinh viên Viện Điện tử Viễn thông, Trường Đại Học Bách Khoa Hà Nội. Hiện đang công tác tại trung tâm nghiên cứu Viện Điện Tử Viễn Thông, Đại Học Bách Khoa Hà Nội. Hướng nghiên cứu của anh gồm xử lý ngôn ngữ tự nhiên, hệ thống gợi ý và blockchain.