ASPECT-BASED SENTIMENT ANALYSIS ON STUDENT'S FEEDBACK IN VIETNAMESE

Ton Nu Thi Sau*, Do Phuoc Sang, Pham Thi Thu Trang
Hanoi University of Home Affairs Campus in HCM City

ARTICLE INFO

Received: 29/9/2021
Revised: 18/11/2021
Published: 18/11/2021

KEYWORDS

Vietnamese dataset
Machine learning
Deep learning
Aspect based sentiment analysis
Ensemble architecture

ABSTRACT

In recent years, universities are interested in surveying and analyzing student's feedbacks to improve teaching effectiveness as well as training quality. However, the manual analysis will be costly in terms of effort and time-consuming with the large data. Therefore, in this paper, we introduce a new dataset on student's feedback of aspect categories detection and aspect-sentiment classification tasks. Our data consists of 5,010 sentences which are annotated by 11 pre-defined aspect categories (teacher behavior, teaching skills…) and 3 sentiment polarities (positive, negative, neutral) with annotation agreements of 88.95% and 80.52% according to two tasks. In addition, we present a series of experiments on the dataset based on a combination model BiLSTM-CNN, compared with other machine learning approaches. The experimental results show that our combination method achieves the best scores with the F1-score of 78.93% and 73.78% for the aspect category detection task and aspect-sentiment classification task, respectively. Experimental results demonstrate the effectiveness of our ensemble architecture.

PHÂN TÍCH Ý KIẾN THEO KHÍA CẠNH TRÊN BÌNH LUẬN PHẢN HỒI CỦA SINH VIÊN CHO TIẾNG VIỆT

Tôn Nữ Thị Sáu*, Đỗ Phước Sang, Phạm Thị Thu Trang
Phân hiệu Trường Đại học Nội vụ Hà Nội tại Thành phố Hồ Chí Minh

THÔNG TIN BÀI BÁO

Ngày nhận bài: 29/9/2021
Ngày hoàn thiện: 18/11/2021
Ngày đăng: 18/11/2021

TỪ KHÓA

Dữ liệu tiếng Việt
Máy học
Học sâu
Phân tích ý kiến theo khía cạnh
Mô hình kết hợp

TÓM TẮT

Trong vài năm gần đây, các trường đại học thường khảo sát, thu thập ý kiến của sinh viên để nâng cao hiệu quả giảng dạy và cải thiện chất lượng đào tạo. Tuy nhiên việc phân tích một cách thủ công sẽ tốn nhiều chi phí về công sức và thời gian khi kích thước phản hồi lớn. Do đó, trong bài báo này, chúng tôi giới thiệu một bộ dữ liệu trên phản hồi của sinh viên cho bài toán phát hiện khía cạnh và phân loại cảm xúc theo khía cạnh. Bộ dữ liệu của chúng tôi bao gồm 5010 câu được gán nhãn theo 11 khía cạnh khác nhau (hành vi, kỹ năng giảng dạy…) và theo ba cảm xúc (tích cực, tiêu cực và trung tính) với độ đồng thuận là 88,95% và 80,52% tương ứng hai bài toán. Bên cạnh đó, chúng tôi cũng trình bày một chuỗi thí nghiệm dựa trên bộ dữ liệu dựa trên mô hình kết hợp BiLSTM-CNN so sánh với các mô hình máy học khác. Kết quả nghiên cứu cho thấy rằng phương pháp kết hợp BiLSTM-CNN đạt kết quả tốt hơn các phương pháp khác với chỉ số F1 là 78,93% và 73,78% tương ứng cho bài toán phát hiện khía cạnh và phân loại trạng thái cảm xúc theo khía cạnh. Kết quả thử nghiệm chứng minh tính hiệu quả của kiến trúc tổng thể của chúng tôi.

DOI: https://doi.org/10.34238/tnu-jst.5101

* Corresponding author. *Email: sauvtc@gmail.com*

1. Giới thiệu

Trong các năm gần đây, ngành giáo dục ở Việt Nam đã có những thay đổi đáng kể từ chương trình đào tạo, chất lượng đội ngũ giảng viên cho đến môi trường học tập với mục đích giúp sinh viên tiếp thu kiến thức hiệu quả hơn. Đặc biệt hiện nay chương trình đào tạo, chất lượng đội ngũ giảng viên, cơ sở vật chất,… được các trường đại học rất quan tâm và cố gắng cải thiện cho phù hợp với nhu cầu của người học, đáp ứng sự phát triển của xã hội. Để giải quyết được vấn đề này, các trường đại học thường khảo sát để lấy ý kiến phản hồi của người học liên quan đến chương trình học, nội dung các môn học, hoạt động giảng dạy của giảng viên... Thông thường các ý kiến này được phân tích một cách thủ công bởi các nhân viên. Tuy nhiên, việc phân tích ý kiến phản hồi theo cách thủ công sẽ làm mất nhiều thời gian và không tổng hợp được một cách chính xác các vấn đề mà sinh viên đề cập đến. Bài toán phân tích ý kiến phản hồi theo khía cạnh được các nhà nghiên cứu đặt ra với mục đích nghiên cứu ra các thuật toán, mô hình phân tích các ý kiến một cách tự động với độ chính xác cao. Trong 5 năm trở lại đây, hầu hết các nghiên cứu đều sử dụng bộ dữ liệu được công bố bởi hội thảo SemEval-2016 [1]. Hội thảo này đã công bố tổng cộng 19 bộ dữ liệu trên 8 ngôn ngữ cho 7 lĩnh vực khác nhau. Tuy nhiên, trong đó không có các lĩnh vực giáo dục. Chính vì thế, các nhóm nghiên cứu [2]-[4] đã trình bày các nghiên cứu tập trung vào lĩnh vực miền giáo dục. Cụ thể, tác giả M. Sivakumar và cộng sự [2] đã sử dụng các phương pháp phân lớp và phân cụm truyền thống trên dữ liệu phản hồi sinh viên thu thập ở trang Twitter. Tác giả G. S. Chauhan và cộng sự [3] đã trình bày nghiên cứu ảnh hưởng của khía cạnh trong môi trường dạy học bằng cách sử dụng các mô hình máy học và dựa trên từ vựng. Tác giả Z. Kastrati và cộng sự [4] đã trình bày một kiến trúc tận dụng chiến lược học giám sát kém (weak supervision) dựa trên mô hình CNN để dự đoán các khía cạnh trên ý kiến phản hồi của sinh viên. Kết quả đánh giá trên độ đo F1 đạt 86,13% cho bài toán phát hiện khía cạnh và 82,10% cho bài toán phát hiện cảm xúc cho khía cạnh.

Còn đối với tiếng Việt, bài toán này cũng nhận được nhiều nghiên cứu từ năm 2018 sau cuộc thi shared-task VLSP [5]. Tác giả Nguyễn Thị Minh Huyền và các cộng sự [5] đã tổ chức cuộc thi và sử dụng bộ dữ liệu cho bài toán ABSA đối với miền dữ liệu nhà hàng và khách sạn ở mức độ đoạn. Dựa trên bộ dữ liệu này, tác giả Đặng Văn Thìn và các cộng sự [6] đã sử dụng một phương pháp chuyển bài toán nhiều nhãn thành các bài toán phân lớp nhị phân sử dụng các đặc trưng được rút trích từ ý kiến của người dùng. Sau đó, Đặng Văn Thìn và các cộng sự [7] cũng đề xuất một phương pháp học sâu Deep Convolutional Neural Network để giải quyết bài toán phát hiện khía cạnh trên hai bộ dữ liệu này. Ngoài ra, tác giả Nguyễn Thị Thanh Thúy và các cộng sự [8] cũng trình bày một bộ dữ liệu cho bài toán ABSA ở miền nhà hàng và tận dụng dữ liệu bổ sung từ tiếng Anh để làm giàu dữ liệu. Kết quả thử nghiệm trên phương pháp SVM cho thấy hiệu quả của cách tiếp cận này. Tác giả Trần Thiện Khải và Phan Thị Tươi [9] đã trình bày một kiến trúc kết hợp nhiều mô hình máy học khác nhau cho bài toán phân tích cảm xúc trên dữ liệu tiếng Việt. Tuy nhiên phương pháp của các tác giả cần nhiều tài nguyên về bộ nhớ và thời gian huấn luyện. Gần đây, tác giả Đặng Văn Thìn cùng các cộng sự [10] cũng đã công bố 2 bộ dữ liệu chuẩn ở mức độ câu cho hai miền nhà hàng và khách sạn với kích thước 10,000 câu ý kiến để phục vụ cho nghiên cứu. Từ đó, chúng ta có thể thấy hầu hết các bộ dữ liệu được xây dựng và phát triển cho miền nhà hàng hoặc khách sạn. Từ việc nhận thấy được tầm quan trọng và nhu cầu bài toán cho lĩnh vực giáo dục, chúng tôi tiến hành thu thập và xây dựng bộ dữ liệu đối với ý kiến phản hồi sinh viên ở mức độ câu. Mục tiêu của chúng tôi là nghiên cứu và áp dụng các phương pháp máy học để xây dựng cho việc hỗ trợ phân tích các ý kiến phản hồi của sinh viên một cách tự động. Các đóng góp chính của chúng tôi trong bài báo này được trình bày như sau:

+ Đầu tiên, chúng tôi tiến hành thu thập và gán nhãn thủ công một bộ dữ liệu ở mức độ câu cho ý kiến phản hồi của sinh viên bao gồm 11 khía cạnh khác nhau và 3 mức trạng thái cảm xúc (tích cực, tiêu cực, trung tính).

+ Thứ hai, chúng tôi nghiên cứu và áp dụng các phương pháp máy học khác nhau, bao gồm các phương pháp máy học truyền thống cũng như các mô hình học sâu để giải quyết bài toán này trên bộ dữ liệu đã xây dựng.

Bố cục bài báo của chúng tôi được trình bày như sau: Phần tiếp theo sẽ trình bày phương pháp nghiên cứu. Sau đó là kết quả và bàn luận trình bày ở phần 3. Cuối cùng là trình bày phần kết luận.

2. Phương pháp nghiên cứu

Trong phần này, chúng tôi sẽ trình bày chi tiết quy trình xây dựng và gán nhãn dữ liệu cho hai toán con bài toán ABSA bao gồm: (1) Bài toán phát hiện khía cạnh và (2) Bài toán phát hiện cảm xúc theo khía cạnh cho miền dữ liệu giáo dục. Ngoài ra, chúng tôi cũng trình bày chi tiết các phương pháp thử nghiệm và so sánh với các phương pháp máy học truyền thống và học sâu trên bộ dữ liệu xây dựng.

2.1. Xây dựng và gán nhãn dữ liệu

Qua tìm hiểu chúng tôi nhận thấy hiện nay chưa có bộ dữ liệu chuẩn về ý kiến phản hồi của sinh viên theo khía cạnh. Cho nên mục tiêu của chúng tôi là xây dựng một bộ dữ liệu chuẩn ở mức độ câu phục vụ cho việc nghiên cứu bài toán sử dụng các phương pháp học giám sát. Để thu thập và gán nhãn dữ liệu ý kiến phản hồi của sinh viên, chúng tôi kế thừa và phát triển dựa trên các ý kiến của bộ dữ liệu UIT-VSFC [11]. Chúng tôi tận dụng các ý kiến đã được xử lý và tiến hành xây dựng các hướng dẫn gán nhãn theo 11 khía cạnh và 3 trạng thái cảm xúc khác nhau dựa trên các phân tích thực tế cho hai bài toán khác nhau: (1) Bài toán phát hiện khía cạnh – các khía cạnh khác nhau được đề cập trong ý kiến phản hồi của sinh viên; (2) Bài toán phát hiện cảm xúc theo khía cạnh – đối với mỗi khía cạnh được đề cập sẽ xác định trạng thái cảm xúc (tích cực, tiêu cực, trung tính). Ví dụ, cho một ý kiến "*thầy rất nhiệt tình , nhưng dạy hơi khó hiểu.*", thì kết quả gán nhãn của chúng tôi là "*{Hành vi, positive}, {Kỹ năng giảng dạy, negative}*". Trước khi gán nhãn, chúng tôi xây dựng tài liệu hướng dẫn gán nhãn để hỗ trợ người gán nhãn trong quá trình xây dựng dữ liệu. Sau đó, chúng tôi sẽ tiến hành các giai đoạn gán thử và đánh giá người gán nhãn trên bộ dữ liệu của chúng tôi. Kết quả gán nhãn cuối cùng giữa ba người gán nhãn của chúng tôi đạt độ đồng thuận là 88,95% cho bài toán phát hiện khía cạnh và 80,52% cho bài toán phát hiện cảm xúc trên khía cạnh. Kết quả độ đồng thuận này cho phép chúng tôi tiến hành gán nhãn một cách độc lập. Danh sách các khía cạnh và số lượng tương ứng trong toàn bộ bộ dữ liệu được trình bày ở Bảng 1 và Bảng 2. Kết quả cuối cùng, chúng tôi đã xây dựng được 5010 câu ý kiến phản hồi của sinh viên được gán nhãn theo 11 khía cạnh và 3 trạng thái cảm xúc khác nhau. Bộ dữ liệu của chúng tôi được chia thành ba tập khác nhau là tập huấn luyện, tập phát triển và tập kiểm tra theo tỷ lệ chia 7/1/2.

Bảng 1. *Danh sách và thống kê số lượng các khía cạnh được trong bộ dữ liệu của chúng tôi*

Ký hiệu	Khía cạnh	Diễn giải
#aspect1	Kỹ năng giảng dạy	Các ý kiến đề cập đến cách tổ chức dạy học, phương pháp dạy học lý thuyết, thực hành của giảng viên.
#aspect2	Kinh nghiệm	Các ý kiến đề cập đến kinh nghiệm thực tiễn và việc đưa nội dung thực tiễn lồng ghép vào bài giảng của giảng viên
#aspect3	Hành vi	Các ý kiến đề cập đến các hành vi, thái độ của giảng viên trong giảng dạy và giao tiếp với người học.
#aspect4	Bài tập	Các ý kiến đề cập đến bài tập, số lượng bài tập và các loại bài tập của giảng viên,…
#aspect5	Chấm điểm	Các ý kiến đề cấp đến hoạt động chấm điểm của giảng viên.
#aspect6	Cung cấp tài liệu	Các ý kiến đề cập đến việc cung cấp tài liệu, giáo trình, giáo án của giảng viên.
#aspect7	Kiến thức	Các ý kiến đề cập đến mức độ hiểu biết của giảng viên về nội dung giảng dạy, kiến thức đã cung cấp cho sinh viên.
#aspect8	Chương trình học	Các ý kiến đề cập đến chương trình học, môn học.
#aspect9	Thiết bị dạy học	Các ý kiến đề cập đến trang thiết bị dạy học như phòng học, máy chiếu, quạt, đèn.
#aspect10	Đề xuất	Các ý kiến đề cập đến đề xuất, mong muốn của sinh viên gửi đến giảng viên và nhà trường.
#aspect11	Nói chung	Các ý kiến đề cập đến vấn đề chung của giảng viên và hoặc không thuộc các khía cạnh trên.

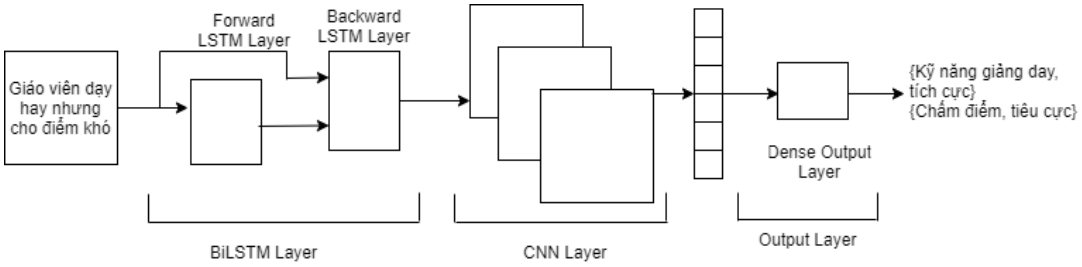
Nhìn vào Bảng 1, chúng ta dễ dàng nhận thấy được sự mất cân bằng giữa các khía cạnh với nhau, cụ thể đối với các khía cạnh liên quan đến nhận xét của sinh viên về giảng viên nhiều như "*kỹ năng giảng dạy*", "*hành vi*" hay "*bài tập*". Còn đối với các khía cạnh khác ít được sinh viên đề cập đến như là "*chấm điểm*", "*thiết bị dạy học*", "*đề xuất*". Điều này có thể giải thích được, bởi vì hầu hết các khảo sát là khảo sát của sinh viên đánh giá về chất lượng giảng dạy của môn học, nên hầu hết các ý kiến phản hồi của sinh viên sẽ nhận xét đến giảng viên là điều có thể chấp nhận được. Đối với bài toán này, việc mất cân bằng giữa các khía cạnh là một điều không thể tránh khỏi vì dữ liệu chúng tôi được thu thập từ những ý kiến thực tế. Vì vậy, sự chênh lệch giữa các khía cạnh này cũng là một thách thức trong bộ dữ liệu của chúng tôi xây dựng.

Bảng 2. *Danh sách và thống kê số lượng các trạng thái cảm xúc Tương ứng với từng khía cạnh trong bộ dữ liệu của chúng tôi*

Khía cạnh	Nhãn cảm xúc			Tổng
	Tích cực	Tiêu cực	Trung tính	
Kỹ năng giảng dạy	1148	536	15	1699
Kinh nghiệm	102	18	0	120
Hành vi	1530	434	6	1970
Bài tập	151	122	4	277
Chấm điểm	24	37	1	62
Cung cấp tài liệu	66	89	0	155
Kiến thức	155	57	2	214
Chương trình học	17	59	2	78
Thiết bị dạy học	59	2	3	64
Đề xuất	121	2	29	152
Nói chung	232	285	50	567

2.2. Kiến trúc mô hình

Sau khi xây dựng tập dữ liệu ý kiến phản hồi của sinh viên, chúng tôi tiến hành cài đặt các phương pháp dựa trên cách tiếp cận các mô hình máy học truyền thống và các mô hình học sâu. Kiến trúc mô hình tổng quát được thí nghiệm trong bài báo này được trình bày ở Hình 1. Sau đó, bài báo trình bày một mô hình kết hợp giữa hai mô hình mạng hồi quy hai chiều là Bidirectional Long short-term memory và mô hình mạng tích chập Convolutional Neural Network – viết tắt là BiLSTM-CNN.



Hình 1. *Kiến trúc mô hình kết hợp BiLSTM-CNN cho bài toán tích ý kiến theo khía cạnh*

Mô hình của chúng tôi được mô tả ở Hình 1 bao gồm các thành phần chính như sau: Lớp đầu vào (Input), lớp nhúng từ (Embedding), lớp mạng hồi quy LSTM hai chiều (BiLSTM), Lớp tích chập (Convolution), lớp gộp (Pooling), lớp phân loại (Fully connected) và lớp đầu ra (Output). Trong đó, chi tiết các thành phần chính được trình bày như sau:

+ **Lớp đầu vào**: Các phản hồi sau khi qua bước tiền xử lý sẽ được biểu diễn thành các véc tơ số với chiều dài cố định với chiều của vectơ cố định là bình luận dài nhất. Các bình luận không đủ độ dài sẽ được tự động thêm giá trị <PAD>.

+ Lớp nhúng từ: Mỗi từ vựng sẽ được chuyển thành một vectơ đại diện thông tin biểu diễn của chúng. Các công trình nghiên cứu trước đây đã chứng minh việc sử dụng các bộ nhúng từ (pre-trained word embedding) đem lại hiệu quả tốt hơn so với việc khởi tạo các vector này một cách

ngẫu nhiên. Chính vì thế, trong bài báo này, chúng tôi sử dụng bộ nhúng từ đã được huấn luyện sẵn dành[1] cho tiếng Việt được huấn luyện trên miền dữ liệu tin tức để rút trích các vectơ từ vựng.

+ Lớp BiLSTM: Tiếp theo, chúng tôi sử dụng một mô hình mạng hồi quy LSTM hai chiều để khai thác thông tin mối liên hệ của các từ vựng theo ngữ cảnh trước và sau trong câu bình luận.

+ Lớp tích chập: Dựa trên các véc tơ biểu diễn từ lớp BiLSTM, chúng tôi sử dụng nhiều bộ lọc (filter) với các kích thước khác nhau để rút trích các đặc trưng cục bộ của bình luận. Cụ thể, kích thước bộ lọc được sử dụng trong lớp này có kích thước là 2,3 và 4. Các giá trị này cho phép mô hình rút trích ra các đặc trưng cục bộ 2-gram, 3-gram và 4-gram.

+ Lớp gộp: Ở tầng kiến trúc này, chúng tôi sử dụng kỹ thuật Global Max Pooling cho mỗi lớp tích chập để rút trích ra các đặc trưng quan trọng của bình luận để làm véc tơ biểu diễn cho toàn bộ đầu vào.

+ Lớp phân loại: Sau khi rút trích ra các đặc trưng quan trọng biểu diễn đầu vào, chúng tôi đưa các đặc trưng này qua lớp phân loại với hàm kích hoạt RELU để xác định xem nhãn khía cạnh và trạng thái cảm xúc tương ứng được đề cập bình luận trong đầu vào.

+ Lớp đầu ra: Mỗi khía cạnh và trạng thái cảm xúc tương ứng sẽ được biểu diễn thành một one-hot véc tơ có độ dài là 4 phần tử đại diện cho các thông tin: None, positive, neutral, negative. Chúng tôi sử dụng một bộ phân lớp với hàm kích hoạt softmax tương ứng mỗi khía cạnh để tính toán giá trị phân bố xác suất của từng nhãn phân loại.

$$\text{softmax} = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \quad (1)$$

$$\text{Loss} = -\sum_{i=1}^{4} y_i \,.\log \hat{y}_i \quad (2)$$

Bộ dữ liệu phản hồi sinh viên của chúng tôi là một bộ dữ liệu không chứa nhiều lỗi ngữ pháp, từ vựng. Tuy nhiên, để tăng độ chính xác cho mô hình, chúng tôi vẫn tiến hành các bước xử lý dữ liệu trước khi huấn luyện. Các bước tiền xử lý được trình bày như sau:

+ Bước 1: Xóa các thông tin dư thừa trong bình luận như nhiều khoảng trắng, dấu chấm câu hoặc các icon trong bình luận và áp dụng các biểu thức chính quy để thay thế các dữ liệu số thành ký từ "num".

+ Bước 2: Sau đó, chúng tôi sử dụng thư viện Pyvi[2] để tách đầu vào thành các từ vựng bởi vì một từ vựng trong tiếng Việt được cấu tạo từ một hoặc nhiều âm tiết.

+ Bước 3: Bước cuối cùng là chuyển tất cả các từ vựng trong chuỗi đầu vào thành chữ thường để giảm kích thước từ vựng trong bộ dữ liệu và nâng cao hiệu quả.

2.2.1. Mô hình so sánh

Trong bài báo này, chúng tôi cũng nghiên cứu và cài đặt các phương pháp máy học truyền thống như Support Vector Machine, Naive Bayes hay Neural Network kết hợp với các đặc trưng thủ công. Bên cạnh đó, chúng tôi cũng nghiên cứu các mô hình học sâu như mạng hồi quy Long short-term Memory, mạng tích chập Convolution Neural Network trên bộ dữ liệu gán nhãn của chúng tôi. Chi tiết thông số các mô hình so sánh được chúng tôi trình bày như sau:

- Support Vector Machine (SVM) [6], [8]: SVM là một trong những phương pháp máy học truyền thống đạt hiệu quả tốt đối với các bài toán xử lý ngôn ngữ. Chúng tôi sử dụng mô hình Linear SVM với thông số khởi tạo giá trị C=0,1.

- Naive Bayes (NB): Đây cũng là một phương pháp phân loại tốt cho dữ liệu văn bản, tuy nhiên bởi vì véc tơ biểu diễn cho các đặc trưng có xu hướng rời rạc, do đó chúng tôi sử dụng mô hình Naive Bayes đa thức để cài đặt thí nghiệm.

- Neural Network (NN): Mạng nhân tạo với một lớp ẩn duy nhất với 128 node được sử dụng hàm kích hoạt ReLu, hàm tối ưu hóa Adam, giá trị $\alpha = 0{,}001$ và tối đa 300 lần lặp.

[1] https://github.com/sonvx/word2vecVN
[2] https://github.com/trungtv/pyvi

- CNN: Mạng tích chập CNN [12] là một trong những mô hình học sâu có hiệu quả đối với các bài toán phân loại văn bản. Chính vì thế, chúng tôi sử dụng mạng CNN như là một mô hình so sánh chuẩn để đánh giá hiệu quả.

- LSTM: Tương tự như mô hình CNN thì mô hình mạng hồi quy LSTM cũng là mô hình học sâu chuẩn, do đó chúng tôi cũng cài đặt mô hình mạng hồi quy LSTM [13] với các thông số chuẩn.

Đối với các mô hình máy học truyền thống, chúng tôi sẽ tiến hành rút trích các đặc trưng thủ công từ vựng và áp dụng kỹ thuật TF-IDF để biểu diễn các đặc trưng văn bản thành các vectơ số để đưa vào các mô hình huấn luyện các bộ phân lớp.

2.2.2. Chi tiết cài đặt

Đối với mô hình kết hợp BiLSTM-CNN, chúng tôi sử dụng mô hình mạng hồi quy 2 chiều LSTM với giá trị mỗi số chiều của chiều ẩn là 128 chiều. Số lượng bộ lọc trong mỗi lớp tích chập của chúng tôi có 128 bộ lọc với kích thước kernel tương ứng 2,3,4 từ vựng với hàm kích hoạt ReLU. Giá trị tốc độ học của hàm tối ưu Adam được chọn với giá trị 0,001. Giá trị batch size để huấn luyện mô hình được gán là 32. Đối với mô hình học sâu CNN thì chúng tôi sử dụng 3 bộ lọc tích chập khác nhau với kích thước tương tự như mô hình kết hợp là 3 lớp tích chập với kernel là 2,3,4. Còn đối mô hình LSTM thì số mỗi số chiều của chiều ẩn có giá trị là 128. Cả hai mô hình CNN và LSTM đều sử dụng một bộ nhúng từ word2vec[3] đã huấn luyện trên tập dữ liệu các bài báo tin tức với số chiều của mỗi véc-tơ là 300 chiều. Các mô hình máy học truyền thống như Naïve Bayers, SVM, Neural Network thì chúng tôi áp dụng kỹ thuật Grid Search để lựa chọn ra các tham số mô hình trên tập phát triển của bộ dữ liệu.

2.2.3. Độ đo đánh giá

Để đánh giá hiệu quả của các phương pháp khác nhau, chúng tôi sử dụng các độ đo chuẩn cho bài toán này là độ chính xác, độ phủ và chỉ số F1-score được tính theo phương pháp micro bởi vì tỷ lệ mất cân bằng giữa các nhãn khía cạnh với nhau. Công thức tính độ chính xác, độ phủ và chỉ số F1 micro được trình bày như sau:

$$P = \frac{\sum_{i=1}^{C}|A \cap B|}{\sum_{i=1}^{C}|A|} \quad (3)$$

$$R = \frac{\sum_{i=1}^{C}|A \cap B|}{\sum_{i=1}^{C}|B|} \quad (4)$$

$$F_1 = \frac{2 * P * R}{P + R} \quad (5)$$

Trong đó: A là phân lớp được hệ thống dự đoán ra, B là phân lớp đích (phân lớp được người dùng gán nhãn), C là tổng số lượng nhãn khía cạnh (C=11 trong trường hợp dữ liệu chúng tôi).

3. Kết quả và bàn luận

Ở trong phần này, chúng tôi sẽ trình bày kết quả nghiên cứu của phương pháp thử nghiệm và so sánh kết quả với các mô hình máy học truyền thống và mô hình học sâu khác trên bộ dữ liệu đã xây dựng. Bảng 3 và Bảng 4 trình bày kết quả thực nghiệm các mô hình trên tập kiểm tra tương ứng với hai bài toán là: Phát hiện khía cạnh và Phát hiện khía cạnh cùng với trạng thái cảm xúc tương ứng theo các độ đo như: độ chính xác, độ phủ và chỉ số F1. Nhìn một cách tổng quan giữa hai bài toán, chúng ta dễ dàng nhận thấy được sự hiệu quả của phương pháp kết hợp BiLSTM-CNN liên quan đến chỉ số F1, cụ thể đối với bài toán phát hiện khía cạnh, mô hình chúng tôi đạt độ chính xác là 78,78%, độ phủ là 79,08%, còn độ đo F1 là 78,93%. Còn đối với bài toán phát hiện khía cạnh và trạng thái cảm xúc tương ứng, thì mô hình này đạt kết quả độ chính xác là 73,64%, độ phủ là 73,93% và độ đo F1 là 73,78%. Ở đây, chúng ta thấy rằng kết quả của bài toán thứ hai lúc nào cũng sẽ thấp hơn bài toán đầu tiên với mục tiêu của bài toán thứ hai là xác định các khía cạnh và trạng thái cảm xúc tương ứng, do đó khi tính toán độ đo, chúng ta sẽ tính đúng một mẫu khi mô hình vừa xác định chính xác cả hai nhãn khía cạnh và trạng thái cảm xúc. Đối với ba phương pháp máy học truyền thống như SVM, NB và NN, chúng ta thấy được sự hiệu quả của mô hình SVM so với hai

[3] https://github.com/sonvx/word2vecVN

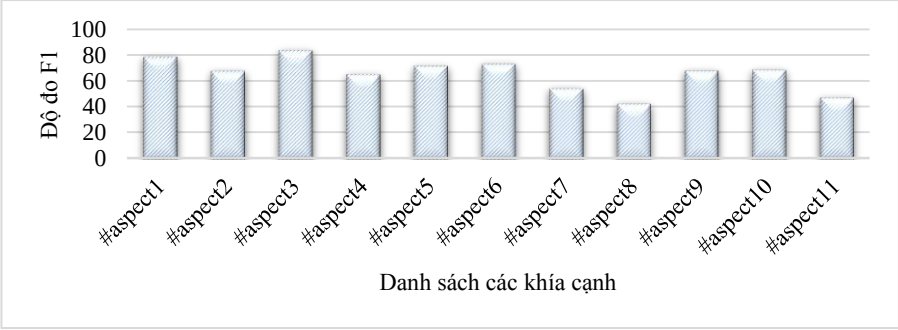
phương pháp còn lại. Kết quả này cho thấy rằng SVM vẫn là một mô hình hiệu quả nhất trong các phương pháp máy học cổ điển. Còn đối với hai mô hình học sâu là CNN và LSTM thì chúng ta thấy có sự hiệu quả cao hơn +0,61% và +1,27% của kiến trúc CNN. Tuy nhiên sự chênh lệch này không đáng kể giữa hai mô hình. Dựa vào kết quả chúng ta vẫn thấy được sự vượt trội của các mô hình học sâu so với các mô hình máy học truyền thống. Cụ thể mô hình CNN cao hơn mô hình SVM là +0,9% cho bài toán phát hiện khía cạnh, và +3,48% cho bài toán phát hiện khía cạnh và trạng thái cảm xúc. Còn mô hình đề xuất thử nghiệm của chúng tôi thì cao hơn mô hình CNN lần lượt là +2,82% và + 1,26% tương ứng cho hai bài toán. Kết quả mô hình kết hợp CNN và BiLSTM cao hơn hai mô hình học sâu CNN và LSTM bởi vì chúng tôi sử dụng mô hình BiLSTM để học biểu diễn theo ngữ cảnh hai chiều của câu đầu vào, sau đó dùng kỹ thuật CNN để rút trích các đặc trưng theo từng bộ lọc trên biểu diễn của BiLSTM. Điều này giúp mô hình có nhiều thông tin và tăng độ hiệu quả hơn khi sử dụng hai mô hình một cách riêng lẻ.

Bảng 3. *Kết quả thí nghiệm các phương pháp cho bài toán phát hiện khía cạnh trên tập kiểm tra*

Phương pháp	Độ chính xác (%)	Độ phủ (%)	Chỉ số F1 (%)
NB	57,75	61,75	59,69
NN	68,70	75,37	71,88
SVM	68,41	**83,51**	75,21
LSTM	73,25	77,90	75,50
CNN	72,60	79,98	76,11
BiLSTM-CNN	**78,78**	79,08	**78,93**

Bảng 4. *Kết quả thí nghiệm các phương pháp cho bài toán phát hiện khía cạnh và trạng thái cảm xúc tương ứng trên tập kiểm tra*

Phương pháp	Độ chính xác (%)	Độ phủ (%)	Chỉ số F1 (%)
NB	51,76	55,34	53,49
NN	61,18	67,12	64,01
SVM	62,80	**76,66**	69,04
LSTM	68,52	74,21	71,25
CNN	69,17	76,21	72,52
BiLSTM-CNN	**73,64**	73,93	**73,78**



Hình 2. *Kết quả chi tiết từng khía cạnh và trạng thái cảm xúc của mô hình kết hợp BiLSTM-CNN trên tập kiểm tra*

Hình 2 mô tả kết quả chi tiết độ đo F1 của các khía cạnh trong tập dữ liệu kiểm tra của mô hình đề xuất cho bài toán phát hiện khía cạnh và cảm xúc tương ứng. Nhìn vào Hình 2, chúng ta thấy được sự hiệu quả của mô hình đối với các khí cạnh như "*Hành vi*", "*Kỹ năng giảng dạy*", "*Cung cấp tài liệu*" với độ đo F1 lần lượt là 84,10%, 78,99% và 73,68%. Trong khi đó, các khía cạnh như "*Chương trình học*", "*Nói chung*", "*Kiến thức*" với độ đo F1 lần lượt là 42,86%, 47,71% và 54,76%. Kết quả này có thể giải thích bởi vì số lượng các khía cạnh này thường là các khía cạnh có số lượng ý kiến ít trong dữ liệu. Do đó, để nâng cao hiệu quả của các khía cạnh này, chúng tôi sẽ cố gắng bổ sung các dữ liệu bằng cách gán nhãn thêm hoặc áp dụng các phương pháp tăng cường dữ liệu. Do đó, các nghiên cứu trong tương lai khi sử dụng bộ dữ liệu của chúng tôi cần tập trung chú ý các nâng cao hiệu quả các khía cạnh này để tăng hiệu quả tổng quan của toàn hệ thống.

4. Kết luận

Trong bài báo này, chúng tôi đã trình bày một nghiên cứu về bài toán Phân tích cảm xúc theo khía cạnh trên ý kiến phản hồi của sinh viên với các mục tiêu đã đạt được như sau: (1) Thu thập, xây dựng và gán nhãn thủ công một bộ dữ liệu với kích thước 5010 câu ý kiến bao gồm 11 khía cạnh và mỗi khía cạnh sẽ được gán bởi 3 trạng thái cảm xúc khác nhau; (2) Chúng tôi cũng đã cài đặt các phương pháp máy học, học sâu trên bộ dữ liệu xây dựng để làm nền tảng cho sự phát triển bài toán này ở các công trình tiếp theo. Kết quả thực nghiệm đã minh chứng mô hình kết hợp của chúng tôi BiLSTM-CNN cho kết quả hiệu quả hơn so với các mô hình khác với chỉ số F1 là 78,93% cho bài toán phát hiện khía cạnh và 73,78% cho bài toán phát hiện khía cạnh và trạng thái cảm xúc tương ứng. Trong sự phát triển tương lai của nghiên cứu, chúng tôi sẽ tập trung gán nhãn bổ sung thêm để tăng số lượng dữ liệu và nghiên cứu các phương pháp để nâng cao hiệu suất của mô hình. Bên cạnh đó, bộ dữ liệu gán nhãn của chúng tôi cũng sẽ được công bố cho cộng đồng nghiên cứu để thúc đẩy phát triển lĩnh vực này trong tiếng Việt.

Lời cám ơn

Bài báo là sản phẩm nghiên cứu của đề tài "Xây dựng phần mềm phân tích tự động ý kiến phản hồi của sinh viên về chất lượng đào tạo ở Phân hiệu Trường Đại học Nội vụ Hà Nội tại Thành phố Hồ Chí Minh", mã số của đề tài ĐTCT.2022.133 được tài trợ bởi Trường Đại học Nội vụ Hà Nội.

TÀI LIỆU THAM KHẢO/ REFERENCES

[1] M. Pontiki, D. Galanis, H. Papageorgiou, I. Androutsopoulos, S. Manandhar, M. Al-Smadi, and G. Eryiğit, "SemEval-2016 task 5: Aspect based sentiment analysis," *In International workshop on semantic evaluation*, 2016, pp. 19-30.

[2] M. Sivakumar and U. Srinivasulu Reddy, "Aspect based sentiment analysis of students opinion using machine learning techniques," *In 2017 International Conference on Inventive Computing and Informatics (ICICI)*, IEEE, 2017, pp. 726-731.

[3] G. S. Chauhan, P. Agrawal, and Y. K. Meena, "Aspect-based sentiment analysis of students' feedback to improve teaching–learning process," *In Information and Communication Technology for Intelligent Systems*, Springer, Singapore, 2019, pp. 259-266.

[4] Z. Kastrati, A. S. Imran, and A. Kurti, "Weakly supervised framework for aspect-based sentiment analysis on students' reviews of MOOCs," *IEEE Access*, vol. 8, pp. 106799-106810, 2020.

[5] T. M. H. Nguyen, V. H. Nguyen, T. Q. Ngo, X. L. Vu , M. V. Tran, X. B. Ngo, and A. C. Le, "VLSP shared task: sentiment analysis," *Journal of Computer Science and Cybernetics*, vol. 34, no. 4, pp. 295-310, 2018.

[6] V. T. Dang, D. N. Vu, V. K. Nguyen, and L. T. N. Nguyen, "A transformation method for aspect-based sentiment analysis," *Journal of Computer Science and Cybernetics*, vol. 34, no. 4, pp. 323-333, 2018.

[7] V. T. Dang, D. N. Vu, V. K. Nguyen, and L. T. N. Nguyen, "Deep learning for aspect detection on vietnamese reviews," *In 5th NAFOSTED Conference on Information and Computer Science (NICS)*, IEEE, 2018, pp. 104-109.

[8] T. T. T. Nguyen, X. B. Ngo, and M. P. Tu, "Leveraging Foreign Language Labeled Data for Aspect-Based Opinion Mining," 2020 RIVF International Conference on Computing and Communication Technologies (RIVF), IEEE, 2020.

[9] K. T. Tran and T. T. Phan, "Deep learning application to ensemble learning—the simple, but effective, approach to sentiment classifying," *Applied Sciences 9*, no. 13, p. 2760, 2019.

[10] V. T. Dang, L. T. N. Nguyen, T. M. Truong, L. S. Le, and T. D. Vo, "Two New Large Corpora for Vietnamese Aspect-based Sentiment Analysis at Sentence Level," *Transactions on Asian and Low-Resource Language Information Processing*, vol. 20, no. 4, pp. 1-22, 2021.

[11] V. K. Nguyen, V. D. Nguyen, X. V. P. Nguyen, T. H. T. Truong, and L. T. N. Nguyen, "UIT-VSFC: Vietnamese students' feedback corpus for sentiment analysis," *In 10th International Conference on Knowledge and Systems Engineering (KSE)*, IEEE, 2018, pp. 19-24.

[12] Y. Kim, "Convolutional neural networks for sentence classification," *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 1746-1751.

[13] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735-1780, 1997.