

ĐÁNH GIÁ Ý KIẾN BẠN ĐỌC ĐỐI VỚI THƯ VIỆN THÔNG QUA QUÁ TRÌNH PHÂN LOẠI QUAN ĐIỂM

TS Đậu Mạnh Hoàn
Trường Đại học Quảng Bình

Tóm tắt: Phân loại quan điểm là nghiên cứu, phân tích các ý kiến của con người, thái độ, cảm xúc và tình cảm của người đó đối với một nội dung hay thực thể. Các nội dung hay thực thể có thể đại diện cho các cá nhân, tổ chức, sự kiện hoặc chủ đề. Trên cơ sở các vấn đề đưa ra xem xét sẽ được bao phủ bởi các ý kiến nhiều nhất, kết quả phân loại sẽ dự đoán được tình cảm hay ý kiến đó là tích cực hay tiêu cực.

Với mục đích nắm bắt những ý kiến, tình cảm cũng như đánh giá của bạn đọc sau khi sử dụng thư viện trong thời gian qua, chúng tôi thực hiện nghiên cứu thông qua quá trình biên tập dữ liệu bình chọn của bạn đọc trong quá trình khảo sát, sau đó sử dụng các kỹ thuật học máy để phân loại quan điểm nhằm nắm bắt được các thông tin qua dữ liệu đó. Quá trình đánh giá này không bị chi phối bởi các nhà quản lý hay người thực hiện khảo sát, có độ chính xác cao và rất khách quan.

Trong bài viết này, tác giả trình bày kết quả thực hiện của quá trình phân loại quan điểm của bạn đọc đối với thư viện Trường Đại học Quảng Bình, qua đó nắm bắt được các ý kiến của bạn đọc đối với thư viện và rút ra được những kết luận hữu ích cho quá trình quản lý, phục vụ và phát triển thư viện trong thời gian tới.

Từ khóa: Phân loại quan điểm; phân tích tâm lý; khai phá ý kiến.

EVALUATING READER OPINIONS THROUGH THE CLASSIFICATION ANALYSIS OF THEIR VIEWPOINTS

Abstract: Viewpoint classification is the study and analysis of people's opinions, attitudes, emotions, and feelings towards a content or entity. The content or entity may represent individuals, organizations, events or topics. On the basis of the issues under consideration that will be covered by the most popular opinions, the classification results will predict whether that sentiment or opinion is positive or negative.

For the purpose of capturing the opinions, sentiment as well as evaluations of readers after exploiting and using the library in the past time. We do our research through the process of editing your poll data of readers during the past time survey, then using machine learning techniques to classify opinions to capture information through that data. This evaluation process is not governed by managers or survey takers, is highly accurate and very objective.

In this article, the author presents the results of the process of classifying readers' opinions about the library of Quang Binh University, thereby capturing the opinions of readers about the library and drawing useful conclusions for the process of library management, service and development in the future.

Keywords: Opinion classification; sentiment analysis; sentiment mining.

1. TỔNG QUAN

Ngày nay, sự phát triển và ứng dụng của công nghệ thông tin đã thâm nhập vào tất cả các hoạt động của cuộc sống, trong đó có hoạt động giáo dục. Tất cả các hoạt động dịch vụ đều sử dụng và khai thác mạng internet, với nhiều thiết bị truy cập, nhiều trang mạng xã hội được kết nối và chia sẻ thông tin. Người sử dụng thường xuyên chia sẻ quan điểm, ý kiến, suy nghĩ tình cảm của

họ thông qua các trang mạng xã hội, các trang web, các diễn đàn hay các cuộc điều tra, khảo sát. Và như vậy, qua các trang mạng đó chúng ta tìm thấy các ý kiến của họ, thông tin phản hồi về sản phẩm hoặc lời nhận xét, phê bình. Đánh giá được các ý kiến đó sẽ cung cấp nhiều thông tin hữu ích cho những mục đích khác nhau của các tổ chức và nhà quản lý.

Phân loại quan điểm, phân tích tâm lý

hay khai phá ý kiến là các lĩnh vực nghiên cứu để phân tích các ý kiến của con người, thái độ, cảm xúc và tình cảm của người đó đối với một nội dung hay thực thể nào đó. Phân loại quan điểm là trích xuất và xử lý các ý kiến của con người về một thực thể, nội dung nào đó. Quá trình phân tích quan điểm sẽ xác định được xu hướng ý kiến của con người, sau đó chúng ta sử dụng để phân tích nó. Mục tiêu của phân loại quan điểm là tìm xu hướng ý kiến, xác định những tình cảm mà họ thể hiện theo khuynh hướng nào, từ đó đánh giá các ý kiến phân cực của họ đối với thực thể đó. Phân loại quan điểm của người sử dụng là một lĩnh vực nghiên cứu trong khai phá dữ liệu, trong xử lý ngôn ngữ tự nhiên và ngôn ngữ học với mục đích thu thập được những thông tin mong muốn về quan điểm của người dùng [4].

Phân loại quan điểm thông qua phân tích ý kiến của người dùng ngày càng thu hút được nhiều sự chú ý của con người vì lợi ích tuyệt vời của nó với xã hội và cuộc sống. Phân loại quan điểm người dùng trên cơ sở phân tích ý kiến của họ về một nội dung nào đó, những tình cảm có thể được phân loại thành một trong hai loại, chẳng hạn như tích cực và tiêu cực; hoặc rất tốt và tốt; hoặc xấu và rất xấu; hoặc tốt và xấu, Ở khía cạnh này, một nhiệm vụ phân tích ý kiến có thể được hiểu như là một nhiệm vụ phân loại mà mỗi thể loại đại diện cho một tình cảm của người dùng.

Quá trình phân tích quan điểm người dùng có thể giúp các nhà quản lý, các tổ chức xã hội phân tích, tổng hợp, tổ chức và dự báo các định hướng ý kiến của đối tượng. Nó đóng vai trò quan trọng trong nhiều hoạt động như đánh giá ý kiến người dùng, phân tích thị trường sản phẩm và đánh giá quan điểm của người dùng. Việc phân tích quan điểm đi đến phân loại ý kiến người dùng sẽ cung cấp cho các tổ chức xác định chiến lược để nâng cao chất lượng hoạt động. Nó cũng tạo điều kiện cho các nhà hoạch định chính sách hay các chính trị gia phân tích thái độ, sự tiếp nhận của công chúng đối với

các chính sách, dịch vụ công cộng hoặc với các vấn đề chính trị, trung cầu dân ý [1, 3]. Mục đích chính của phân loại quan điểm người sử dụng là xác định sự phân cực của ý kiến người dùng. Trong nghiên cứu này, chúng tôi vận dụng vào bài toán phân loại quan điểm của bạn đọc thông qua quá trình khai thác sử dụng các dịch vụ của họ đối với thư viện Trường Đại học Quảng Bình.

2. VAI TRÒ, Ý NGHĨA CỦA KHAI PHÁ QUAN ĐIỂM CỦA BẠN ĐỌC ĐỐI VỚI HOẠT ĐỘNG THƯ VIỆN

Quản lý hệ thống TT-TV bằng phương pháp truyền thống chỉ có khả năng quản lý ở mức độ thấp, và đó chính là rào cản ảnh hưởng đến sự phát triển của thư viện. Quản lý truyền thống chủ yếu dựa vào kinh nghiệm để đưa ra quyết định, đó là chủ quan, một chiều, thiếu thông tin, và không thể đáp ứng yêu cầu của thời đại nhất là khi mọi hoạt động đều gắn với công nghệ. Từ xu hướng phát triển của nền tri thức số, các nguồn lực của thông tin là rất đa dạng và phong phú, bao gồm các loại tài liệu giấy và số, sách điện tử, tạp chí điện tử, cơ sở dữ liệu đĩa CD, dữ liệu đa phương tiện và dữ liệu web.

Tất cả các hoạt động trong thư viện đều có mục đích cuối cùng là phục vụ bạn đọc một cách tốt nhất. Vì vậy, việc nắm bắt ý kiến, đánh giá, nhận xét và những kiến nghị của bạn đọc đối với thư viện một cách kịp thời là vô cùng quan trọng đối với các nhà quản lý, từ đó có những chính sách, hoạch định, thay đổi để đáp ứng nhu cầu của bạn đọc. Bạn đọc là người trực tiếp khai thác cơ sở vật chất của thư viện, các ý kiến nhận xét, phản hồi về cơ sở vật chất sẽ giúp các nhà quản lý có những điều chỉnh phù hợp. Trong quá trình tìm kiếm tài liệu, khai thác các dịch vụ liên quan đến tài liệu, bạn đọc cũng sẽ thấu hiểu được sự đáp ứng và hiệu quả mang lại của hoạt động đó. Bạn đọc sẽ đánh giá được khả năng tìm kiếm nhanh hay chậm của hệ thống, tính đa dạng của nguồn tài liệu, ... Như vậy, nắm bắt được ý kiến của bạn đọc sẽ giúp các nhà quản lý có

được những thông tin quan trọng để hoạch định và đưa ra những chính sách phát triển hệ thống TT-TV một cách tốt nhất.

Nắm bắt được các ý kiến nhận xét, quan điểm, tình cảm và các kiến nghị của bạn đọc sẽ giúp các nhà quản lý có được các chính sách như:

- Thúc đẩy phát triển văn hóa đọc;
- Phát triển tài nguyên thông tin phù hợp với mục tiêu, nội dung, chương trình, lĩnh vực, ngành đào tạo, nghiên cứu khoa học và phát triển công nghệ của Trường Đại học Quảng Bình;
- Điều chỉnh các chính sách phù hợp để xây dựng và phát triển các nguồn tài liệu nội sinh, cơ sở dữ liệu số, tài liệu dạy học trực tuyến, tài nguyên học liệu mở và các nguồn tài nguyên khác phục vụ cho công tác giảng dạy, học tập và nghiên cứu khoa học. Tổ chức và quản lý các hoạt động TT-TV một cách hiệu quả. Phát huy khả năng ứng dụng công nghệ thông tin vào công tác quản lý và phục vụ bạn đọc, hiện đại hóa thư viện. Tổ chức không gian đọc;
- Hướng dẫn sử dụng sản phẩm thư viện và dịch vụ thư viện;
- Hoàn thiện kỹ năng tìm kiếm, khai thác và sử dụng thông tin;
- củng cố, mở rộng kiến thức cho người học, người dạy và cán bộ quản lý một cách khoa học và kịp thời. Xây dựng, phát triển, xử lý, lưu giữ, bảo quản và kết nối tài nguyên thông tin phù hợp với người sử dụng thư viện. Tổ chức khai thác, sử dụng các sản phẩm thông tin và dịch vụ thư viện phù hợp với bạn đọc của mình. Góp phần hình thành và phát triển kiến thức, kỹ năng, phẩm chất, năng lực của con người trong xã hội số. Truyền bá tri thức, phát huy giá trị văn hóa của dân tộc và nhân loại;
- Phục vụ tối ưu nhu cầu nghiên cứu, học tập, giải trí cho bạn đọc; có các chiến lược liên kết với thư viện trong nước và nước ngoài, đáp ứng một cách tốt nhất các hoạt động khác của một thư viện.

3. HỌC MÁY VÀ ƯU ĐIỂM CỦA HỌC MÁY

Phương pháp Học máy sử dụng các giải thuật học máy nổi tiếng bằng việc sử dụng cú pháp và các đặc trưng trong ngôn ngữ. Các giải thuật học máy đã được chứng minh là những giải thuật phân loại rất hiệu quả cho vấn đề phân loại văn bản. Các giải thuật này phù hợp với bài toán phân loại văn bản vì chúng có khả năng đáp ứng được không gian đầu vào có số chiều rất lớn, các đặc trưng rời rạc, ít liên hệ lẫn nhau, các vectơ tài liệu thưa và các vấn đề phân lớp trong văn bản có thể chia cắt được. Học máy là một lĩnh vực có liên quan đến việc nghiên cứu các thuật toán và kỹ thuật cho phép các máy tính để “học hỏi” tự động từ kinh nghiệm. Học máy bao gồm học có giám sát, học không có giám sát và học bán giám sát [11].

- Học có giám sát

Các phương pháp học có giám sát phụ thuộc vào sự tồn tại của các nhãn huấn luyện văn bản. Có rất nhiều loại học có giám sát trong phân loại văn bản, quá trình phân loại qua phương pháp học có giám sát được thực hiện qua nhiều bước khác nhau: Đầu tiên cần xác định loại của tập dữ liệu cần huấn luyện, sau đó tiến hành thu thập dữ liệu huấn luyện. Việc thu thập dữ liệu huấn luyện có thể thực hiện bằng nhiều cách khác nhau. Bước tiếp theo là lựa chọn và biểu diễn các đặc trưng, bước này có vai trò quyết định hiệu quả của quá trình phân loại. Hầu hết các phương pháp máy học áp dụng cho bài toán phân loại văn bản đều sử dụng cách biểu diễn văn bản dưới dạng vectơ đặc trưng. Điểm khác biệt duy nhất chính là không gian đặc trưng được chọn lựa. Số lượng các đặc trưng không được quá lớn, do sự bùng nổ dữ liệu, không được quá nhỏ sẽ không đảm bảo thông tin, chúng phải đủ lớn để dự đoán chính xác đầu ra. Để giải quyết vấn đề này, thông thường chúng ta sẽ chọn lựa những đặc trưng được đánh giá là hữu ích, bỏ đi những đặc trưng không quan trọng. Trong giai đoạn này, chúng ta sử dụng các phương pháp chọn lựa đặc

trung hiệu quả để giảm chiều của vectơ đặc trưng văn bản mà không mất đi độ chính xác khi phân loại. Bước cuối cùng là xác định cấu trúc của hàm chức năng cần tìm và chọn giải thuật học tương ứng, chạy giải thuật học từ tập huấn luyện thu thập được để cho ra kết quả. Một số phương pháp học có giám sát phổ biến trong phân loại quan điểm như phân loại theo xác suất với một số kỹ thuật phân loại xác suất nổi tiếng như: Phân loại Naïve Bayes, Bayesian Network, Maximum Entropy; Kỹ thuật phân loại tuyến tính. Kỹ thuật này nổi bật với phương pháp phân loại Máy vector hỗ trợ (support vector machines), Mạng nơron nhân tạo (ANN: Artificial Neural Network), Cây quyết định (decision tree), Phân loại dựa vào các luật cơ bản (rule-based), ... Trong nghiên cứu này, chúng tôi sử dụng kỹ thuật học máy mạng nơron nhân tạo để tiến hành nghiên cứu và thực nghiệm.

- Học không có giám sát

Học không giám sát là phương pháp nhằm tìm ra mô hình phù hợp với các tập dữ liệu quan sát. Nó khác biệt với học có giám sát ở chỗ là đầu ra đúng tương ứng cho mỗi đầu vào là không biết trước. Trong học không có giám sát, đầu vào là một tập dữ liệu được thu thập. Học không có giám sát thường xem các đối tượng đầu vào như là một tập các biến ngẫu nhiên. Sau đó, một mô hình mật độ kết hợp sẽ được xây dựng cho tập dữ liệu đó. Học không có giám sát có thể được dùng kết hợp với suy diễn Bayes để cho ra xác suất có điều kiện cho bất kỳ biến ngẫu nhiên nào khi biết trước các biến khác.

- Học bán giám sát

Khi cả dữ liệu huấn luyện và dữ liệu đã dán nhãn không có sẵn, chúng ta có thể sử dụng phương pháp học nửa giám sát. Một phân loại thống kê được huấn luyện, việc huấn luyện có thể sử dụng cả dữ liệu có nhãn và không có nhãn. Kết hợp các mẫu có gắn nhãn và không gắn nhãn để sinh một hàm hoặc một bộ phân loại thích hợp.

- Ưu điểm của học máy

Ngày nay, hầu hết tất cả các tổ chức, công ty khi muốn xem xét đánh giá dữ liệu khách hàng của họ để phục vụ quản lý, phát triển hoạt động hay kinh doanh đều sử dụng học máy và xem học máy trở thành một phần trọng tâm không thể thiếu trong các hoạt động của họ. Học máy đã trở thành một yếu tố tham gia vào quá trình cạnh tranh và nâng cao chất lượng giữa các công ty hay tổ chức với nhau vì hiệu quả mà nó mang lại.

Ưu điểm nổi bật của học máy là xử lý một cách khoa học và khách quan việc tìm cách biến đổi các dữ liệu đầu vào thành các dữ liệu đầu ra có ý nghĩa hơn. Một mô hình học máy sẽ ánh xạ dữ liệu từ một không gian biểu diễn này vào một không gian biểu diễn khác mà trên đó chúng ta có thể hiểu rõ hơn về dữ liệu ban đầu. Học máy làm thay bộ não con người trong việc tìm kiếm các ánh xạ biến đổi biểu diễn ban đầu của dữ liệu thành dạng biểu diễn có ý nghĩa hơn, trừu tượng và bất biến hơn. Sử dụng kỹ thuật học máy rất quan trọng vì nó có khả năng cung cấp cho các tổ chức cái nhìn về xu hướng trong hành vi của người dùng và các mô hình hoạt động dịch vụ, cũng như hỗ trợ sự phát triển của các dịch vụ mới. Một trong những ưu điểm khác của học máy là đạt được tính phổ quát cao, có nghĩa là máy tính có thể xử lý tốt những dữ liệu mà nó có thể chưa gặp bao giờ, nó có thể suy diễn một số nguyên tắc riêng để quyết định xem nội dung đó thuộc loại kết quả nào.

Học máy cũng đồng thời tạo nên tiền đề cho việc biểu diễn sự phân tán của các dữ liệu, tính toán phân tán, tính toán song song, đưa ra các dự đoán, dự báo hay tư vấn,... khi cho phép lưu trữ, xử lý các thuộc tính học được trên nhiều thiết bị tính toán ở nhiều vị trí địa lý khác nhau. Nó cũng có khả năng tạo ra hay cung cấp một công cụ mới cho việc phát triển của các tổ chức trong tương lai khi nắm bắt được các thuộc tính, đặc trưng của dữ liệu trong quá trình biểu diễn dữ liệu cho máy học từ một ứng

dụng này có thể được sử dụng trong các ứng dụng khác một cách hiệu quả.

4. NGHIÊN CỨU PHÂN LOẠI QUAN ĐIỂM TẠI TRUNG TÂM THÔNG TIN-THƯ VIỆN, TRƯỜNG ĐẠI HỌC QUẢNG BÌNH

4.1. Dữ liệu nghiên cứu

Để tiến hành phân tích ý kiến, phân loại quan điểm của bạn đọc, chúng tôi tiến hành khảo sát trực tuyến 849 sinh viên và 136 cán bộ giảng viên của Trường Đại học Quảng Bình với các chủ đề khác nhau.

Bảng 1. Nội dung thông tin khảo sát

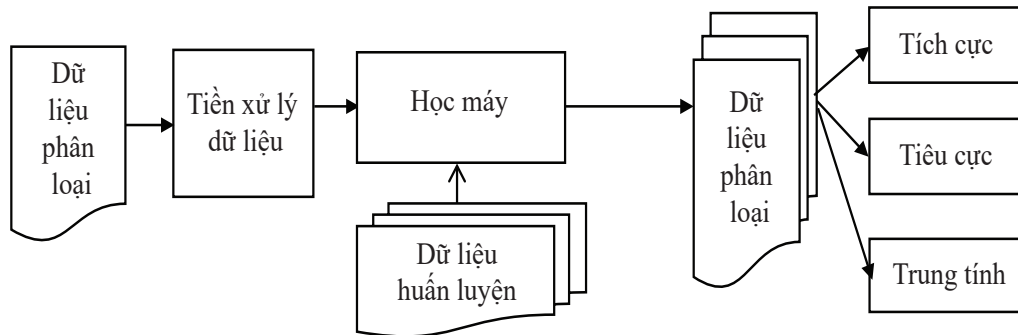
TT	Nội dung khảo sát
1	Mức độ thường xuyên đến Trung tâm Học liệu
2	Mục đích sử dụng Trung tâm Học liệu
3	Sử dụng máy tính tra cứu tài liệu
4	Bạn đọc thường xuyên đến phòng đọc 1
5	Bạn đọc thường xuyên đến phòng đọc 2
6	Bạn đọc thường xuyên đến phòng đọc 3
7	Bạn đọc thường xuyên đến phòng đọc 4
8	Bạn đọc đánh giá dịch vụ thư viện
9	Bạn đọc sử dụng loại hình tài liệu và đánh giá như thế nào
10	Tập huấn sử dụng phần mềm thư viện
11	Truy cập internet và tra cứu tìm kiếm thông tin trên máy tính tại Trung tâm Học liệu như thế nào
12	Yếu tố của phòng đọc tại Trung tâm Học liệu như thế nào
13	Sự đáp ứng của tài liệu tại Trung tâm Học liệu cho nhu cầu tin của bạn đọc
14	Mức độ truy cập vào website của Trung tâm Học liệu
15	Bạn đọc tiếp cận nguồn tin thông qua các kênh nào
16	Số lượng máy tính dùng để tra cứu thông tin và phục vụ học tập ở Trung tâm
17	Phong cách và thái độ phục vụ của cán bộ Trung tâm
18	Sự kết nối internet đến hệ thống máy tính tra cứu tài liệu
19	Tìm kiếm tài liệu, gia hạn, xem thông tin mượn trả tài liệu trên hệ thống
20	Thời gian cho mượn tài liệu của Trung tâm
21	Nguồn học liệu, sản phẩm và dịch vụ của Trung tâm Học liệu phục vụ cộng đồng

Dữ liệu thu thập từ khảo sát sẽ được tiến hành tiền xử lý trên máy tính để phục vụ cho các bước xử lý sau hiệu quả hơn. Dữ liệu thực nghiệm tương ứng có 985 file được tập hợp để xây dựng máy học và kiểm thử. Sau khi tách từ và loại bỏ từ dừng, số từ có được là 74.642 từ. Như vậy, tập ngữ liệu cần mô hình hóa là một ma trận chứa TF*IDF của các từ có kích thước là 985×74642 phần tử. Ngoài ra, chúng tôi cũng thu thập dữ liệu từ các trang web để làm dữ liệu luyện học

trong phạm vi câu văn đánh giá về thư viện, hoạt động giáo dục có 6.500 từ bao gồm: Danh từ: 2.600 từ; Động từ: 1.000 từ; Tính từ: 900 từ; Trạng từ: 900 từ; Từ khác: 11.000 từ. Sau khi tiến hành xử lý dữ liệu, chúng tôi sử dụng 70% dữ liệu có được để làm dữ liệu huấn luyện và dùng 30% dữ liệu còn lại làm dữ liệu kiểm thử.

4.2. Các bước xử lý dữ liệu

Dữ liệu phân loại quan điểm được thực hiện theo tiến trình sau:



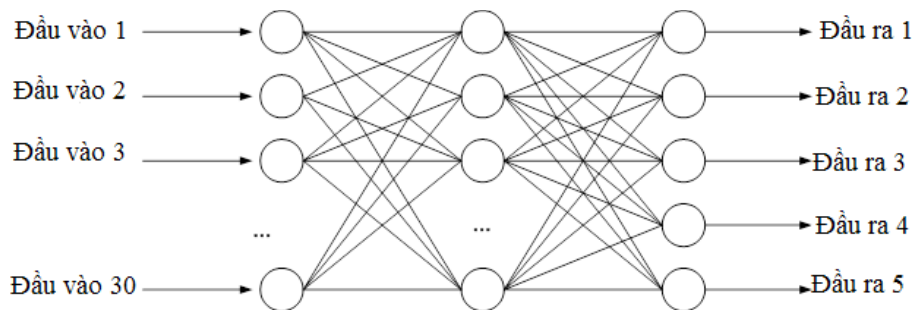
Hình 1. Tiến trình phân loại quan điểm

Trong đó, dữ liệu phân loại là dữ liệu thu thập từ bạn đọc, dữ liệu này được xử lý trước khi đưa vào học máy, sau quá trình xử lý của máy sẽ cho ra dữ liệu đã phân loại, dữ liệu này được phân chia thành 3 loại tích cực, tiêu cực và trung tính.

4.3. Kỹ thuật mạng nơron nhân tạo

Chúng tôi sử dụng kỹ thuật học máy

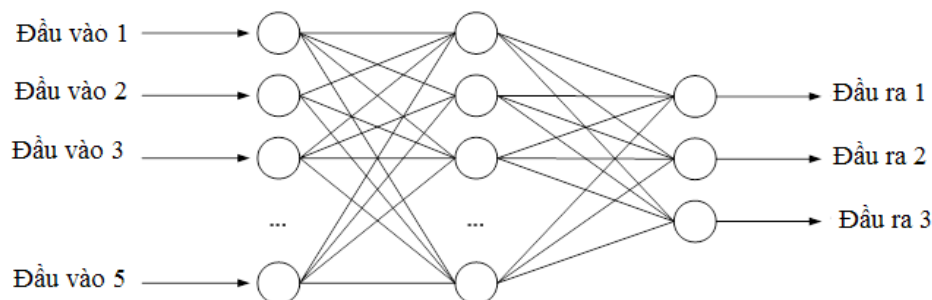
mạng nơron nhân tạo (ANN: Artificial Neural Network), để cho hệ thống tiến hành phân tích, đoán nhận cảm xúc của bạn đọc [5, 6]. Mạng nơron nhân tạo ứng dụng trong phân loại với cấu trúc 03 lớp. Mạng nơron 1: gồm 3 lớp, lớp đầu vào (input) gồm 30 nút, lớp ẩn (hidden) gồm 12 nút, lớp đầu ra (output) gồm 5 nút (vector danh từ, động từ, tính từ, trạng từ và từ khác).



Hình 2. Cấu trúc mạng nơron 1

Mạng Noron 2: Gồm 3 lớp, lớp đầu vào (input) gồm 5 nút (vector danh từ, động từ, tính từ, trạng từ và từ khác), lớp ẩn (hidden)

gồm 7 nút, lớp đầu ra (output) gồm 3 nút (tích cực, không xác định và tiêu cực).



Hình 3. Cấu trúc mạng nơron 2

Dữ liệu luyện học được thu thập từ quá trình khảo sát bạn đọc gồm các thành phần: danh từ, động từ, tính từ, trạng từ. Trong đó, các từ mô tả mức độ về tích cực và tiêu cực được gán nhãn phục vụ quá trình luyện học. Kết quả phân loại được phân thành 03 lớp tích cực, tiêu cực và trung tính, dựa vào kết quả này để phân tích, đánh giá ý kiến của bạn đọc.

4. ĐÁNH GIÁ KẾT QUẢ NGHIÊN CỨU VÀ BÀI HỌC KINH NGHIỆM

Sử dụng kỹ thuật học máy mạng nơron nhân tạo để xử lý và đoán nhận quan điểm của bạn đọc đối với thư viện là phương pháp làm có luận cứ khoa học và có tính chính xác cao thông qua các ưu điểm của kỹ thuật này [10]. Việc phân loại thông qua mạng nơron nhân tạo được xử lý bằng máy là khách quan. Quá trình xử lý dữ liệu hoàn toàn tự động nên kết quả mang lại có tính chính xác cao và minh bạch, không có sự can thiệp của con người. Điều này có ý nghĩa rất quan trọng trong việc lấy ý kiến, đánh giá, nhận xét đối với một tổ chức hay hoạt động cho các nhà quản lý.

Qua quá trình nghiên cứu và hệ thống phân loại, chúng tôi nhận thấy việc phân cực quan điểm tích cực của bạn đọc đối với Trung tâm khá rõ ràng thông qua các câu hỏi khảo sát như:

“Bạn đọc *thường xuyên* đến phòng đọc”- Tỷ lệ đánh giá này chiếm 95%.

“Sự đáp ứng của tài liệu tại Trung tâm Học liệu cho nhu cầu tin của bạn đọc *đầy đủ*”- tỷ lệ đánh giá này chiếm 73%.

“Truy cập internet và tra cứu tìm kiếm thông tin trên máy tính tại Trung tâm Học liệu *tốt*”, “Truy cập internet và tra cứu tìm kiếm thông tin trên máy tính tại Trung tâm Học liệu *rất tốt*” - hai tỷ lệ đánh giá này chiếm 74%.

“Phong cách và thái độ phục vụ của cán bộ Trung tâm *nhật tình*”, “Phong cách và thái độ phục vụ của cán bộ Trung tâm *rất nhật tình*”- hai tỷ lệ đánh giá này chiếm 98%.

“Sự kết nối internet đến hệ thống máy tính tra cứu tài liệu *tốt*”, “Sự kết nối internet đến hệ thống máy tính tra cứu tài liệu *rất tốt*”- hai tỷ lệ đánh giá này chiếm 95%.

Với các nhận xét đó, hệ thống đã đánh giá quan điểm của bạn đọc là tích cực thông qua các nhận xét như “thường xuyên”; “đầy đủ”; “rất tốt”; “rất nhật tình”; “thường xuyên”, ... Tỷ lệ các đánh giá tích cực của bạn đọc qua dữ liệu đều vượt trên mức 73% và số liệu này đã khẳng định ý kiến đánh giá đối với các hoạt động thư viện là tích cực.

Các trường hợp phân cực quan điểm tiêu cực của bạn đọc đối với thư viện cũng được hệ thống đoán nhận rõ ràng nhưng tần suất ý kiến của bạn đọc về quan điểm tiêu cực là rất ít, chiếm tỷ lệ rất thấp như các trường hợp:

“Sự kết nối internet đến hệ thống máy tính tra cứu tài liệu *yếu*”, “Thời gian cho mượn tài liệu của Trung tâm là ít ngày” hay “*Rất ít khi* sử dụng máy tính tra cứu tài liệu”, ... với các từ đoán nhận tiêu cực như “yếu”, “ít ngày”, “rất ít khi” chiếm tỷ lệ dưới 5%.

Các trường hợp đoán nhận trung tính cũng xảy ra nhưng rất hạn chế với các tình huống có từ ẩn ý, từ lóng hay nhận xét chung chung như: “Tìm kiếm tài liệu, gia hạn, xem thông tin mượn trả tài liệu trên hệ thống biết *không rõ lắm*”, với từ đoán nhận “không rõ lắm”.

Như vậy, với đặc trưng của tiếng Việt rất phong phú và đa dạng trong ngữ nghĩa, bạn đọc cũng có nhiều cách biểu đạt và cả cách thức sử dụng ngôn ngữ của người đó, các cách thể hiện tình cảm, quan điểm, cảm xúc của bạn đọc còn phụ thuộc vào bối cảnh và thời gian lúc bình phẩm của những người đó. Trong thực tế, một từ trong một lời bình phẩm được xem là quan điểm tích cực trong tình huống này, song vẫn có thể là tiêu cực trong tình huống khác, nhất là trường hợp nói châm biếm, mỉa mai hay cách dùng từ ẩn dụ. Đôi khi, một nhận xét về một vấn đề hoặc chỉ là một phần của vấn đề cũng có thể gây hiểu lầm cho hệ thống khai phá quan điểm.

Việc sử dụng các kỹ thuật học máy để đánh giá nhận xét của con người đã được minh chứng có hiệu quả trong nhiều ứng dụng [7-9]. Với kết quả nghiên cứu này trên 985 file dữ liệu khi sử dụng phương pháp học máy mạng nơon nhân tạo đã mang lại kết quả đánh giá khách quan và có ý nghĩa trong việc nắm bắt tâm tư, ý kiến, tình cảm và những kiến nghị của bạn đọc đến Thư viện Trường Đại học Quảng Bình. Phương pháp nắm bắt này không bị chi phối bởi tính chủ quan của người khảo sát hay người trực tiếp xử lý. Một quá trình đánh giá hoàn toàn tự động và khách quan thông qua việc trích xuất các thông tin dữ liệu đã mang lại cho các nhà quản lý những thông tin rất hữu ích trong công việc cũng như các định hướng phát triển sau này.

Từ những thông tin nắm bắt được thông qua đánh giá khảo sát sẽ trợ giúp Trung tâm hoàn thiện hơn các chức năng thông tin và thư viện, phục vụ công tác đào tạo, nghiên cứu khoa học, truyền bá pháp lý, tư vấn pháp luật và quản lý của Nhà trường, góp phần thực hiện tốt nhiệm vụ chính trị của Nhà trường.

Tham mưu cho Lãnh đạo Nhà trường về tổ chức và quản lý các hoạt động TT-TV một cách khoa học và hiệu quả hơn; có các chiến lược ứng dụng công nghệ thông tin vào công tác quản lý và phục vụ bạn đọc một cách hợp lý cũng như việc tổ chức thực hiện công tác lưu trữ và khai thác tư liệu thông tin, sách báo, tạp chí, giáo trình, tài liệu điện tử phục vụ cho công tác giảng dạy, học tập và nghiên cứu khoa học một cách hiệu quả hơn.

Tư vấn, cung cấp các dịch vụ TT-TV, tổ chức các khóa bồi dưỡng nhằm nâng cao nghiệp vụ, kiến thức về TT-TV cho đội ngũ nhân sự làm công tác thông tin, tư liệu, thư viện và người dùng tin ở trong và ngoài Nhà trường.

Phát triển thư viện số, lưu giữ, quản lý thông tin, tư liệu nhằm phục vụ công tác đào tạo, nghiên cứu khoa học, triển khai ứng dụng và quản lý, nghiên cứu, thu thập, xử

lý, thông báo, phổ biến và cung cấp tin, tài liệu khoa học, dịch vụ thông tin cho cán bộ, công chức, viên chức, người lao động, học sinh, sinh viên, học viên cao học, nghiên cứu sinh cũng sẽ đạt hiệu quả nhất.

TÀI LIỆU THAM KHẢO

1. Nguyễn Thị Lan Anh (2013). Nghiên cứu thuật toán học máy SVM và ứng dụng trong bài toán khai phá ý kiến phản hồi của khách hàng trên Website. Luận văn thạc sỹ Hệ thống thông tin. Học viện Công nghệ Bưu chính Viễn thông. 2013.
2. Phùng Hữu Đoàn (2013). Khai phá dữ liệu từ các mạng xã hội để khảo sát ý kiến đánh giá xã hội các địa điểm du lịch tại Đà Nẵng, Luận văn thạc sỹ Khoa học máy tính. Đại học Đà Nẵng, năm 2013.
3. Ngô Thị Hoa (2015). Phân loại quan điểm trên phương tiện xã hội đối với dữ thảo đề án quốc gia. Luận văn thạc sỹ Công nghệ thông tin. Đại học Công nghệ - Đại học Quốc gia Hà Nội, 2015.
4. Đậu Mạnh Hoàn (2017). Các hướng tiếp cận chính trong kỹ thuật phân loại tình cảm. Đại học Quảng Bình. 2017.
5. Russell, S. J., & Norvig, P. (2016). Artificial Intelligence: A Modern Approach. Copyright c 2010, 2003, 1995 by Pearson Education, Inc.
6. Simon Haykin (1993). Neural Networks and Learning Machines . Mc Master University, Canada. This third edition.
7. M. Birjali, A. Beni-Hssane et al. (2017). Machine Learning and Semantic Sentiment Analysis base Algorithms for Suicide Sentiment Prediction in Social Networks, Procedia Computer Science 113, pp. 65-72.
8. Quinlan J., (1993). C4.5: Programs for Machine Learning, Morgan Kaufman Publishers, 1993.
9. M. Birjali, A. Beni-Hssane et al. (2017). Machine Learning and Semantic Sentiment Analysis base Algorithms for Suicide Sentiment Prediction in Social Networks, Procedia Computer Science 113, pp. 65-72.
10. Dan W. Patterson (1996). Artificial Neural Networks, Theory and Application. Singapore ; New York : Prentice Hall.
11. Nils J. Nilsson. Introduction to Machine Learning. 1998. Robotics Laboratory Department of Computer Science Stanford University, Stanford, CA 94305.

(Ngày Tòa soạn nhận được bài: 2-7-2022; Ngày phản biện đánh giá: 12-8-2022; Ngày chấp nhận đăng: 15-9-2022).