

ĐỊNH HƯỚNG MỚI TRONG HỆ THỐNG GỢI Ý

Trần Thị Thúy*, Bùi Thị Diễm Trinh**, Trương Quốc Định***

TÓM TẮT

Hầu hết các công trình nghiên cứu về hệ thống gợi ý hiện tại chỉ tập trung vào nhóm giải thuật lọc cộng tác mà chưa quan tâm tới nhóm giải thuật nội dung. Trong khi các hệ thống gợi ý xây dựng bằng nhóm giải thuật cộng tác chỉ tập trung đánh giá và đưa ra kết quả tổng quan nhất về một mục tin. Nhưng trong thực tế, người dùng lại quan tâm tới chi tiết từng tính năng của mục tin đó. Trong nghiên cứu này, chúng tôi đặt sự quan tâm đến việc phân tích tự động các ý kiến của khách hàng về một tính năng cụ thể nào đó của sản phẩm. Với bài toán này, dữ liệu đầu vào cần là một tập các sản phẩm cùng tập các tính năng mà người dùng quan tâm và các tập đánh giá tương ứng với mỗi sản phẩm. Mỗi sản phẩm trong tập sản phẩm đều có các tính năng giống nhau, nó được ghi trong tập tính năng. Ứng với mỗi sản phẩm là một tập các đánh giá của khách hàng về sản phẩm đó. Với mỗi một đánh giá của người dùng, hệ thống sẽ tiến hành trích xuất tự động các thông tin về sản phẩm, tính năng, cụm từ nói lên quan điểm và đồng thời dự đoán mức điểm đánh giá cho tính năng của sản phẩm ứng với quan điểm đó.

Từ khóa: Hệ thống gợi ý, khai phá quan điểm, mô hình từ khóa chủ đề, lọc cộng tác, lọc trên nội dung.

ABSTRACT

Most of the current suggestion system studies focus on collaborative filtering algorithms, but not on the content-based filtering algorithms solution. While the recommender systems suggested by the collaborative filtering algorithm team only focused on evaluating and giving the most general result of a particular item. But in reality, users are more interested in the details of each feature of the item. In this study, we place an interest in automatically analyzing customer feedback on a particular feature of the product. With this problem, the input data needs to be a set of products with the same set of features that the user is interested in and the sets of comments corresponding to each product. Each product in the set has the same features, it is included in the feature set. For each product is a set of customer reviews about the product. For each user comment, the system automatically extracts product information, features, phrases, and predicts the score for the product's performance with that opinion.

Keywords: Recommender systems, opinion mining, topic model, collaborative filtering, Content-based Filtering.

* Khoa Kỹ thuật - Công nghệ, Trường Đại học Cửu Long

** Học viên cao học khóa 23, ngành Hệ thống thông tin trường Đại học Cửu Long

*** Khoa Công nghệ thông tin và Truyền thông, Trường Đại học Cần Thơ

1. Cơ sở khoa học và thực tiễn

2. Hệ thống gợi ý là một dạng của hệ thống lọc thông tin, nó được sử dụng để dự đoán sở thích hay xếp hạng mà người dùng có thể dành cho một mục thông tin nào đó mà họ chưa xem xét tới trong quá khứ (mục thông tin có thể là sản phẩm, bài hát, bộ phim, đoạn video clip, sách, bài báo ...).

Trong thời đại kỷ nguyên số ngày nay, Công nghệ Thông tin (CNTT) đã chứng tỏ được vai trò của mình khi ứng dụng thành công và mang lại nhiều lợi ích trong nhiều lĩnh vực hoạt động, trong đó có lĩnh vực thương mại. Thương mại điện tử những năm gần đây đã phát triển mạnh mẽ, đem lại nhiều lợi ích cho nền kinh tế toàn cầu. Bán hàng qua mạng Internet là một loại hình kinh doanh mới của thương mại điện tử, với hình thức kinh doanh này giúp người dùng có thể dễ dàng mua bán hàng hóa hơn so với hình thức mua bán truyền thống. Tuy nhiên, để giúp người dùng có thể chọn được một sản phẩm họ ưng ý trên một trang web bán hàng trực tuyến là điều không hề dễ dàng khi có quá nhiều sản phẩm cùng được hiển thị. Để loại hình bán hàng trực tuyến thực sự phát triển thì việc tích hợp thêm một hệ thống gợi ý là điều hết sức cần thiết.

Đã có rất nhiều tổ chức, công ty ứng dụng thành công hệ thống gợi ý vào trang bán hàng của mình như trang web bán hàng trực tuyến amazon.com (gợi ý các sản phẩm), movielens.org (gợi ý phim), youtube.com (gợi ý video clip). Ở nước ta hiện nay cũng có các trang web có chức năng gợi ý như mp3.zing.vn, vatgia.com ... Tuy nhiên đa phần các trang chỉ gợi ý dựa trên những thông tin cơ bản như các bài hát có cùng ca sĩ, cùng thể loại hay cùng loại sản phẩm và cùng nhà cung cấp.

Hiện tại, để xây dựng một hệ thống gợi ý có rất nhiều giải thuật được đề xuất, một cách tổng quan có thể chia thành các hướng tiếp cận chính như sau:

- **Nhóm giải thuật lọc cộng tác:** Hoạt động dựa trên hành vi của người dùng trước đó như lịch sử giao dịch, để tìm các quy luật tương tác giữa người dùng và các mục tin. Do đó những hệ thống gợi ý dựa trên cách tiếp cận này đều không quan tâm đến các thuộc tính của mục tin, nó có khả năng khai thác thông tin ngoài phạm vi của các thuộc tính mục tin. Mô hình huấn luyện có thể xây dựng dựa trên hành vi của một người dùng, hoặc hiệu quả hơn, nó có thể từ nhiều người dùng khác nhau nhưng có cùng đặc điểm. Khi làm việc với hành vi của người dùng khác, *lọc cộng tác* sử dụng kiến thức nhóm để tạo ra đề xuất dựa trên những người dùng tương tự. Về bản chất, nó lọc trên những người có cùng sở thích, hay những người có cùng những hành vi tương tự, cùng bấm like, cho điểm đối với cùng một mục tin.

○ **Phương pháp láng giềng:** là dựa trên dữ liệu quá khứ của người dùng “tương tự”, hoặc là dựa trên dữ liệu quá khứ của những mục tin “tương tự”.

• Phương pháp dựa trên người dùng:

• Biểu diễn mỗi người dùng bằng một vector các mục tin đã tương tác, có thể có trọng số.

○ Tính độ tương tự giữa các vector đại diện cho người dùng.

○ Đối với người dùng A, ước tính độ phù hợp của mục tin dựa vào lịch sử của nhóm người dùng tương tự như A

○ Có thể chọn k người dùng gần giống A nhất, hoặc chọn tất cả người dùng nhưng thêm trọng số để ưu tiên những người giống A hơn

● Phương pháp dựa trên mục tin:

○ Biểu diễn mỗi mục tin bằng một vector người dùng.

○ Tính độ tương tự giữa các mục tin

○ Đối với người dùng A, tìm các mục tin tương tự với các mục tin A đã tương tác.

○ Gợi ý mục tin cho A từ các mục tin nói trên, bằng các tiêu chí như trọng số cao, nhiều người tương tác

Trong nhóm giải thuật lọc cộng tác, có một số công trình nghiên cứu ở trong và ngoài nước điển hình như sau:

- Nhóm tác giả (A. Troscher và M. Jahrer, 2010) đã đề xuất hệ thống dự đoán khả năng của học sinh để trả lời một cách chính xác dựa trên kết quả trả lời của học sinh đó trong lịch sử, tác giả đã sử dụng các kỹ thuật lọc cộng tác.

- Tác giả (Nguyễn Hùng Dũng, 2014) đã đề xuất xây dựng hệ thống gợi ý cho bán hàng trực tuyến sử dụng kỹ thuật lọc cộng tác sử dụng phản hồi tường minh từ người dùng thông qua các đánh giá trên sản phẩm. Các tác giả đã tiến hành xây dựng một hệ thống bán hàng trực tuyến, từ đó tích hợp kỹ thuật gợi ý vào hệ thống để tư vấn cho khách hàng những sản phẩm họ có thể yêu thích nhất.

- Tác giả (Luu Nguyễn Anh Thư, 2014) đã đề xuất một giải pháp xây dựng hệ thống gợi ý dành cho bán hàng trực tuyến sử dụng phản hồi tiềm ẩn từ người dùng.

- Tác giả (Phạm Minh Chuẩn, 2015) đã đề xuất phương pháp cải tiến phương pháp phân cụm lọc cộng tác và lập cộng tác tối ưu trọng số sử dụng thuật toán PSO để áp dụng trên bộ dữ liệu hệ gợi ý công việc cho thấy độ chính xác mô hình dự đoán có cải thiện rõ rệt so với phương pháp lọc cộng tác truyền thống đồng thời cũng giải quyết được vấn đề dữ liệu thưa mà phương pháp lọc cộng tác thường gặp phải.

- Tác giả (Nguyễn Tấn Phong, 2014) đã đề xuất sử dụng phản hồi tiềm ẩn từ người dùng (như tỷ lệ thời lượng mà người dùng đã nghe trên tổng thời lượng bài hát) và từ đó đề xuất sử dụng giải thuật xếp hạng cá nhân Bayes để xây dựng “Hệ thống gợi ý bài hát” do kỹ thuật này hoạt động tốt trên dữ liệu phản hồi tiềm ẩn.

- Tác giả (Triệu Vĩnh Viêm, 2013) cũng đã đề xuất xây dựng hệ thống gợi ý phim dựa trên mô hình đã được (Koren, 2010) đề xuất. Mô hình này có khả năng cải tiến được độ chính xác ngang bằng với mô hình nhân tố tiềm ẩn.

○ **Dựa trên mô hình:** là xây dựng các mô hình dự đoán dựa trên dữ liệu thu thập được trong quá khứ, và kỹ thuật phân rã ma trận hay được sử dụng.

Trong nhóm giải thuật dựa trên mô hình, có một số nghiên cứu trong và ngoài nước như:

- Nhóm tác giả (N. Thai-Nghe và cộng sự, 2011) đã đề xuất hệ thống e-learning để gợi ý các đối tượng học tập (bài báo) cho sinh viên, nhóm tác giả đã sử dụng kỹ thuật phân rã ma trận.

- Nhóm tác giả (Nguyễn Thái Nghe, Huỳnh Xuân Hiệp, 2012) đã ứng dụng kỹ thuật phân rã ma trận đa quan hệ trong xây dựng hệ trợ giảng thông minh.

- Tác giả (Huỳnh Lý Thanh Nhân, 2013) đã đề xuất sử dụng kỹ thuật phân rã ma trận thiên vị để dự đoán kết quả học tập của sinh viên, từ đó đưa ra những gợi ý phù hợp cho sinh viên lựa chọn những môn học tự chọn.

o **Gợi ý theo ngữ cảnh** hay còn gọi là hệ thống gợi ý đa chiều:

- Xử lý ngữ cảnh đầu vào: Trong tiếp cận này, thông tin về ngữ cảnh được sử dụng để lựa chọn hoặc xây dựng các thiết lập có liên quan đến dữ liệu. Sau khi lọc thông tin ngữ cảnh, hệ thống đa chiều sẽ trở thành hệ thống hai chiều và ta có thể sử dụng bất kỳ phương pháp gợi ý truyền thống nào trên các dữ liệu đã chọn.

- Xử lý ngữ cảnh đầu ra: Trong tiếp cận này, ngữ cảnh thông tin ban đầu bị bỏ qua, và xếp hạng được dự đoán bằng cách sử dụng bất kỳ phương pháp gợi ý truyền thống trên toàn bộ dữ liệu. Sau đó, các gợi ý được điều chỉnh cho mỗi người sử dụng bằng cách sử dụng các thông tin theo ngữ cảnh

- Mô hình hóa ngữ cảnh: Phương pháp này xây dựng một mô hình dự đoán. Ví dụ, mô hình này có thể đưa ra xác suất một người dùng cụ thể chọn một sản phẩm cụ thể trong một bối cảnh cụ thể là bao nhiêu. Sau đó xác suất này được dùng để điều chỉnh và tìm ra gợi ý phù hợp.

Một số nghiên cứu trong và ngoài nước về gợi ý theo ngữ cảnh như sau:

- Tác giả (Tomáš Horváth và cộng sự, 2011): Context-Aware Factorization Models for Student's Task Recommendation, in Proceedings of UMAP 2011

- Tác giả (Lư Chân Thiện, 2015) đã đề xuất kết hợp các phương pháp. Trước tiên xử

lý thông tin ngữ cảnh đầu vào, kế đến dùng kỹ thuật phân rã ma trận để dự đoán kết quả, sau đó xử lý thông tin ngữ cảnh đầu ra.

- **Nhóm lai ghép:** Là sự kết hợp giữa nhiều phương pháp như dựa trên người dùng, dựa trên mục tin, phân rã ma trận và cũng có thể sẽ bao gồm cả lọc dựa trên nội dung.

- Trong nhóm giải thuật lai ghép, có một số nghiên cứu như: Tác giả (José M. Noguera và cộng sự, 2012) đã áp dụng kỹ thuật lai ghép để xây dựng hệ thống gợi ý du lịch sử dụng hệ thống thông tin địa lý 3D.

- **Nhóm kỹ thuật không cá nhân hóa:** Trong nhóm kỹ thuật này, do chúng khá đơn giản, dễ cài đặt nên thường được các website/ hệ thống tích hợp vào, gồm cả các website thương mại, website tin tức, hay giải trí. Chẳng hạn như trong các hệ thống bán hàng trực tuyến, người ta thường gợi ý các sản phẩm được xem, mua, bình luận ... nhiều nhất; gợi ý các sản phẩm mới nhất; gợi ý các sản phẩm cùng loại, cùng nhà sản xuất; gợi ý các sản phẩm được mua, chọn cùng nhau. Một ví dụ khá điển hình là thông qua luật kết hợp (như Apriori), Amazon đã áp dụng khá thành công để tìm ra các sản phẩm hay được mua cùng nhau. Tuy vậy, bất lợi của các phương pháp này là không cá nhân hóa cho từng người dùng, nghĩa là tất cả các user đều được gợi ý giống nhau khi chọn cùng sản phẩm.

Với nhóm kỹ thuật không cá nhân hóa dựa trên phản hồi tiềm ẩn, có một số nghiên cứu như: Tác giả (Y. Hu, Y. Koren và C. Volinsky, 2008) đã xây dựng hệ thống theo dõi các hành vi người dùng khác nhau, chẳng hạn như lịch sử mua hàng, theo dõi thói quen và hoạt động duyệt web, để mô hình hóa sở thích người dùng.

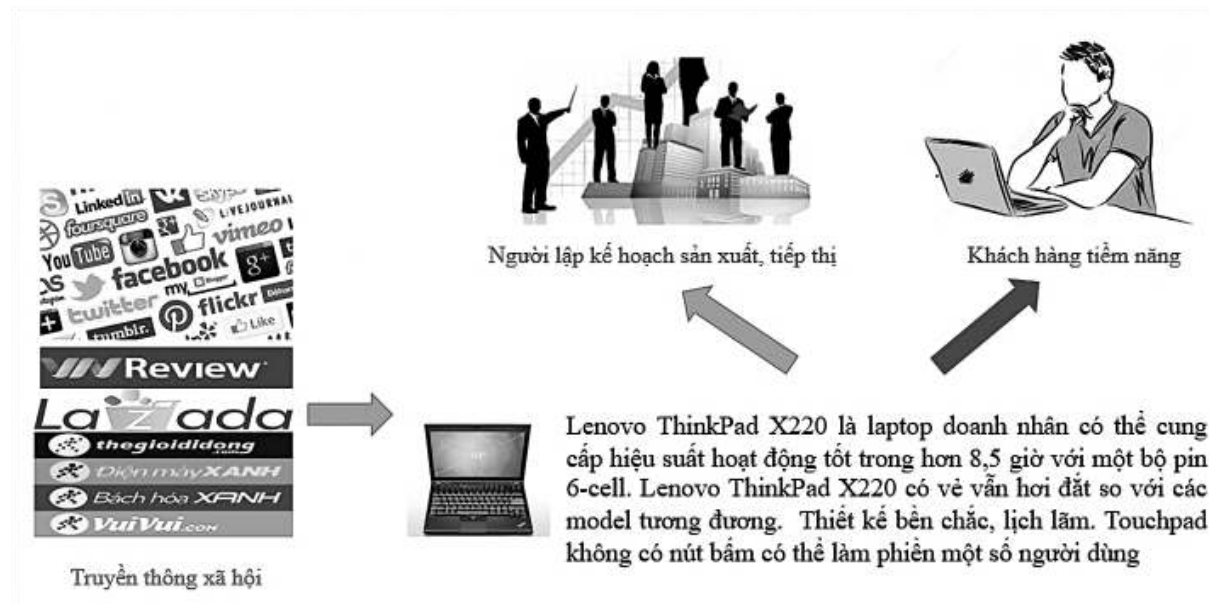
- Nhóm giải thuật lọc trên nội dung:

Gợi ý các mục tin dựa vào hồ sơ của người dùng hoặc dựa vào nội dung/ thuộc tính của những mục tin tương tự như mục tin mà người dùng đã chọn trong quá khứ. Phương pháp này phụ thuộc vào thuộc tính của các mục tin, đặc biệt đối với các mục tin nhiều thuộc tính như các mục tin thuộc lĩnh vực truyền thông, quảng cáo, y tế ... Đặc biệt có thể gợi ý các mục tin mới, thích hợp khi danh sách mục tin được cập nhật liên tục và mục tin mới.

Nhóm tác giả (Hui Li và cộng sự, 2012), đã sử dụng mô hình xác suất và các thuật toán lọc dựa trên nội dung để xây dựng hệ thống dự đoán sở thích của người dùng. Nhóm tác giả

(Robin van Meteren và Maarten van Someren) đã sử dụng các kỹ thuật lọc dựa vào nội dung để xây dựng hệ thống giúp người dùng tìm thông tin họ cần trên internet bằng cách cung cấp cho họ những gợi ý được cá nhân hóa,...

Hầu hết các công trình nghiên cứu về hệ thống gợi ý hiện tại chỉ tập trung vào nhóm giải thuật lọc cộng tác mà chưa quan tâm tới nhóm giải thuật nội dung. Trong khi các hệ thống gợi ý xây dựng bằng nhóm giải thuật cộng tác chỉ tập trung đánh giá và đưa ra kết quả tổng quan nhất về một mục tin. Nhưng trong thực tế, người dùng lại quan tâm tới chi tiết từng tính năng của mục tin đó.



Hình 1. Quan điểm của người dùng có thể khác nhau cho các tính năng khác nhau của cùng 1 sản phẩm

Trong nghiên cứu này, chúng tôi đặt sự quan tâm đến việc phân tích tự động các ý kiến của khách hàng về một tính năng cụ thể nào đó của sản phẩm. Với bài toán này, dữ liệu đầu vào cần là một tập các sản phẩm cùng tập các tính năng mà người dùng quan tâm và các tập

đánh giá tương ứng với mỗi sản phẩm. Mỗi sản phẩm trong tập sản phẩm đều có các tính năng giống nhau, nó được ghi trong tập tính năng. Ứng với mỗi sản phẩm là một tập các đánh giá của khách hàng về sản phẩm đó. Với mỗi một đánh giá của người dùng, hệ thống sẽ

tiến hành trích xuất tự động các thông tin về sản phẩm, tính năng, cụm từ nói lên quan điểm và đồng thời dự đoán mức điểm đánh giá cho tính năng của sản phẩm ứng với quan điểm đó.

Trong phạm vi của nghiên cứu “**Định hướng mới trong hệ thống gợi ý**”, chúng tôi đề xuất ứng dụng mô hình từ khóa chủ đề để giải quyết vấn đề thu thập và phân loại quan điểm của người dùng về tính năng của sản phẩm một cách tốt hơn. Mô hình này có thể trích xuất cả 2 loại thông tin: phân phối đánh giá tổng thể mức độ sản phẩm và từ vựng thể hiện mức độ cảm xúc. Từ đó, mô hình đưa ra

các phán đoán chính xác hơn cho việc đánh giá chức năng sản phẩm. Dựa vào kết quả đó, hệ thống sẽ có các gợi ý phù hợp với yêu cầu người dùng đã đưa ra.

Mô hình chủ đề là một loại mô hình thống kê để phát hiện ra các “chủ đề” trừu tượng xuất hiện trong một bộ sưu tập các tài liệu. Khái niệm đầu tiên về mô hình chủ đề được khởi xướng vào năm 2002 do Griffiths và Steyvers. Khi xem xét tài liệu bất kỳ, sự xuất hiện của các từ trong tài liệu đó sẽ gợi ý cho người đọc về một chủ đề có liên quan, nó cũng có thể xuất hiện nhiều ở một tài liệu khác.

Theatre	Music	League	Prison	Rate	Pub	Market	Railway	Air
Stage	Band	Cup	Years	Cent	Guinness	Stock	Train	Aircraft
Arts	Rock	Season	Sentence	Inflation	Beer	Exchange	Station	Flying
Play	Song	Team	Jail	Recession	Drink	Demand	Steam	Flight
Dance	Record	Game	Home	Recovery	Bar	Share	Rail	Plane
Opera	Pop	Match	Prisoner	Economy	Drinking	Group	Engine	Airport
cast	dance	division	serving	cut	alcohol	news	track	pilot

Bảng 1. Các từ chủ đề trong tập mô tả của Andrews năm 2009.

Mô hình từ khóa chủ đề là sự kết hợp mỗi “từ khóa” với một phân phối xác suất các “chủ đề”, trong đó mỗi “chủ đề” là một phân phối xác suất thông qua các từ. Nếu một văn bản chứa nhiều từ khóa thì văn bản đó sẽ được biểu diễn như là một hỗn hợp xác suất các chủ đề tương ứng với các từ khóa. Một từ khóa có thể dùng chung cho một vài chủ đề khác nhau.

2. Mục đích

Mục tiêu tổng quát là xây dựng được hệ thống gợi ý sản phẩm dựa trên tính năng của sản phẩm được người dùng quan tâm sử dụng kỹ thuật rút trích thông tin và khai phá quan điểm.

Để đạt được mục tiêu tổng quát như trên,

cần đáp ứng được các mục tiêu cụ thể sau:

- Tự động thu thập dữ liệu là các ý kiến bình luận trên các diễn đàn liên quan đến các nhóm sản phẩm (smart phone, laptop, tivi ...)

- Lọc nhiễu đối với dữ liệu thu thập được.

- Tự động phân tích quan điểm thể hiện trong các bình luận để tạo cơ sở cho việc gợi ý, có tính đến yếu tố thời gian.

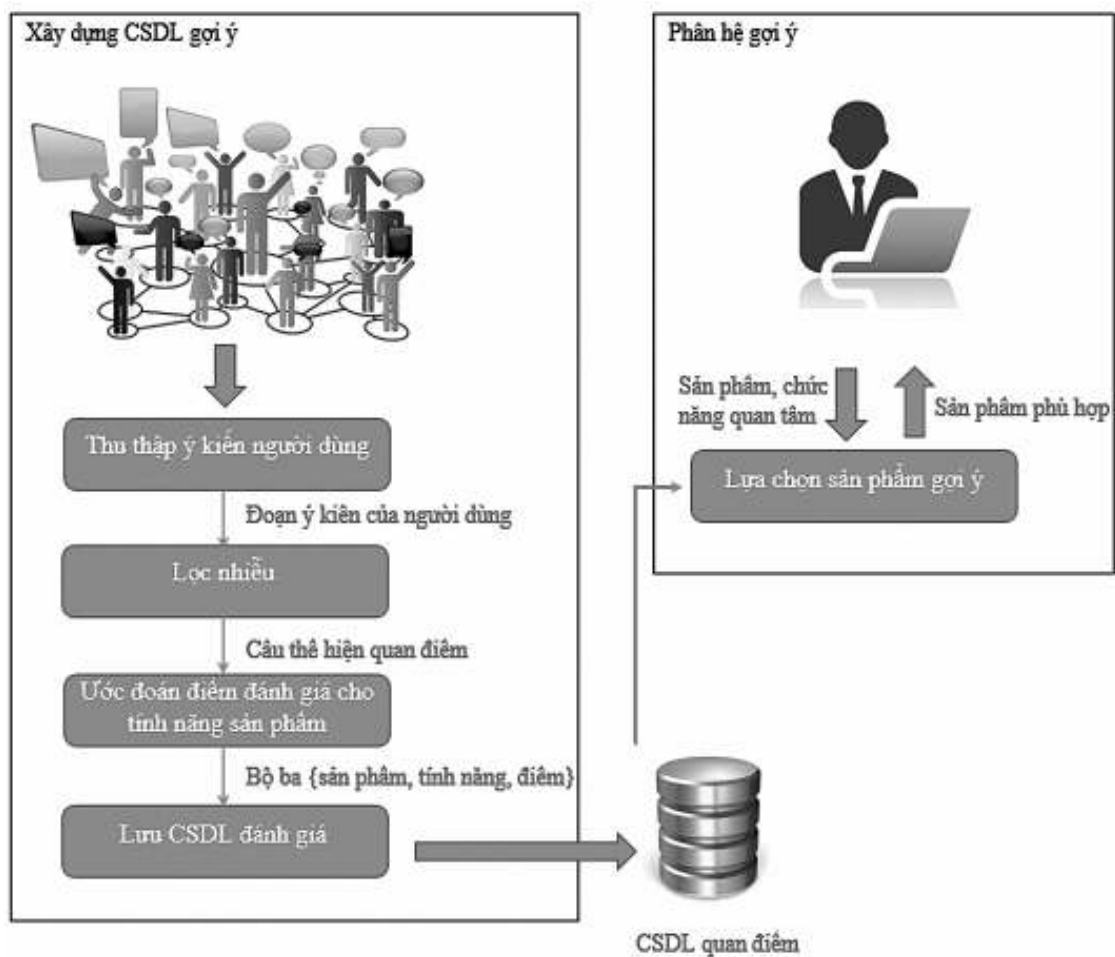
- Phát triển chức năng gợi ý theo tính năng của sản phẩm dựa trên cơ sở dữ liệu quan điểm của người dùng.

Trong phạm vi nghiên cứu này, về phương diện dữ liệu, hệ thống chỉ tập trung vào một

nhóm sản phẩm, dự kiến là nhóm sản phẩm máy tính xách tay.

3. Nội dung và các vấn đề cần giải quyết

Hệ thống mà chúng tôi đề xuất có kiến trúc như hình 2, trong đó bao gồm 2 phân hệ chính: **Xây dựng CSDL gợi ý** và **Gợi ý sản phẩm**.



Hình 2. Kiến trúc hệ thống

Phân hệ xây dựng CSDL gợi ý

Thu thập tự động các ý kiến đánh giá của người dùng

- Ý kiến đánh giá của người dùng sẽ được thu thập một cách tự động từ internet dựa trên các kỹ thuật của lĩnh vực tìm kiếm thông tin.

- Trong quá trình thu thập hệ thống sẽ ghi nhận lại thời gian mà người dùng đưa ra ý

kiến có liên quan phục vụ cho việc xây dựng CSDL gợi ý theo thời gian.

- Xây dựng tính năng cho phép thu thập ý kiến đánh giá từng phần dựa trên kỹ thuật lưu vết địa chỉ ý kiến được thu thập.

- Phương pháp thực hiện:

- Truy vấn các bộ máy tìm kiếm web (ví dụ truy vấn google search bằng cách

truy xuất địa chỉ <https://www.google.com/search?q=CauTruyVan&num=10>).

○ Đọc và phân tích kết quả trả về là trang HTML để thu được dữ liệu.

Lọc nhiễu

- Mục tiêu của giai đoạn này là xử lý bước đầu các ý kiến thu thập được. Đôi khi các ý kiến là một đoạn văn bao gồm nhiều câu và trong đó có những câu không đề cập đến quan điểm về tính năng của sản phẩm. Kết quả của giai đoạn lọc nhiễu là chỉ giữ lại những câu thể hiện quan điểm về một tính năng nào đó của sản phẩm.

- Danh mục các sản phẩm của một nhóm sản phẩm sẽ được cập nhật định kỳ dựa trên danh mục sản phẩm của 1 trang web bán hàng trực tuyến được lựa chọn.

- Sử dụng thư viện vnTagger cho giai đoạn tách từ và gán từ loại.

- Danh mục tính năng của mỗi nhóm sản phẩm được xác định trước với một số từ mô tả ban đầu (topic modelling). Danh mục từ này sẽ được mở rộng thông qua tự diễn đồng nghĩa hoặc học tự động.

Ước đoán điểm đánh giá cho tính năng sản phẩm

- Dữ liệu đầu vào cho giai đoạn ước đoán điểm đánh giá là bộ ba {sản phẩm, tính năng, cụm từ quan điểm} kèm yếu tố thời gian. Hệ thống sẽ thực hiện ước đoán điểm đánh giá (3 hoặc 5 mức điểm) dựa trên mô hình từ khóa chủ đề hoặc kỹ thuật máy học (SVM chẳng hạn).

- Kết quả của giai đoạn này được minh họa như sau:

#id	#sanpham	#hangsanxuat	#dong	#tinhnang	#tucamxuc	#giatri	#diengiai
1	laptop	asus	zenbook	pin	tệ	1	
2	laptop	asus	zenbook	ram	tối ưu	5	
3	laptop	asus	zenbook	ổ cứng	đỏm	2	
4	laptop	asus	zenbook	cpu	tuyệt	5	
5	laptop	asus	zenbook	giá	bèo	4	
6	laptop	asus	zenbook	mẫu mã	lịch lãm	5	
7	laptop	asus	zenbook	thiết kế	bền	3	

Lưu CSDL các quan điểm

- Các quan điểm được thể hiện dưới dạng (sản phẩm, hãng sản xuất, model, tính năng, điểm đánh giá, thời gian) sẽ được lưu vào CSDL quan điểm phục vụ gợi ý sản phẩm cho người dùng.

Phân hệ gợi ý sản phẩm

- Hệ thống tiếp nhận yêu cầu của người dùng về nhóm sản phẩm cũng như danh mục các tính năng mà người dùng quan tâm để đưa ra gợi ý sản phẩm phù hợp.

- Danh mục sản phẩm phù hợp với yêu

cầu cử người dùng được trả về có xếp thứ tự dựa trên điểm đánh giá.

Công thức tính điểm xếp hạng:

$$W_{pi} = \sum_{a_j \in A} W_{p_i a_j}$$

Trong đó:

- W_{pi} là điểm xếp hạng của sản phẩm pi

- a_j là đặc tính mà người dùng quan tâm, trong trường hợp người dùng không xác định bất kỳ đặc tính nào thì hệ thống sẽ xét toàn bộ tất cả các đặc tính của sản phẩm.

- $W_{p_i a_j}$ là điểm xếp hạng của sản phẩm pi đối với tính năng a_j .

Đánh giá tính khả thi của hệ thống

Tính khả thi của hệ thống sẽ được đánh giá từng phần đối với những giai đoạn có thể kiểm chứng được, ví dụ như:

- Tính hiệu quả của giai đoạn lọc nhiễu dựa trên số lượng câu không thể hiện quan điểm của người dùng về tính năng của sản phẩm đã được lọc bỏ.

- Độ chính xác của giai đoạn ước đoán điểm xếp hạng dựa trên phương pháp chuyên gia.

- Thử nghiệm hệ thống trên một tập đối tượng vừa phải người sử dụng để đánh giá tính chính xác của sản phẩm gợi ý.

4. Hướng phát triển của nghiên cứu

- Xây dựng hệ thống phù hợp hơn cho nhiều nhóm sản phẩm với tập đặc tính được mở rộng và không phải mô tả trước.

- Phát triển hoàn chỉnh hệ thống gợi ý để có thể cung cấp dưới dạng dịch vụ.

TÀI LIỆU THAM KHẢO

- [1] A. Töschler and M. Jahrer, "Collaborative filtering applied to educational data mining" in Proceedings of the KDD Cup 2010 Workshop on Improving Cognitive Models with Educational Data Mining, Washington, DC, USA, 2010.
- [2] D. M. Blei, A.Y. Ng, M.I. Jordan. 2003. Latent Dirichlet Allocation. Journal of Machine Learning Research 3(2003) page 993-1022.
- [3] F. Ricci, L. Rokach, B. Shapira and P. B Kantor, 2011. "Recommender Systems Handbook", Springer.
- [4] Griffiths, T. L., & Steyvers, M. (2002). A probabilistic approach to semantic representation. In Proceedings of the 24th Annual Conference of the Cognitive Science Society.
- [5] Huỳnh Lý Thanh Nhân, Nguyễn Thái Nghe. 2013. Hệ thống dự đoán kết quả học tập và gợi ý lựa chọn môn học. Kỷ yếu hội thảo quốc gia lần thứ XVI: Một số vấn đề chọn lọc của CNTT&TT (@2013), trang 110-118. Nhà xuất bản Khoa học kỹ thuật. ISBN: 987-604-67-0251-1
- [6] José M. Noguera, Manuel J. Barranco, Rafael J. Segura, Luís Martínez (2012). A Mobile 3D-GIS Hybrid Recommender System for Tourism. Journal Information Sciences: an International Journal Volume 215, December, 2012, pp 37-52
- [7] Koren, Y., Bell, R., & Volinsky, C. 2009. Matrix factorization techniques for recommender systems. IEEE Computer Society Press, 42(8), 30-37.
- [8] Lư Chân Thiện và Nguyễn Thái Nghe. 2015. MỘT TIẾP CẬN TRONG XÂY DỰNG HỆ THỐNG GỢI Ý THEO NGỮ CẢNH. Kỷ yếu Hội nghị quốc gia lần thứ VIII "Nghiên cứu cơ bản và ứng dụng Công nghệ thông tin" (FAIR 2015)
- [9] Lưu Nguyễn Anh Thư và Nguyễn Thái Nghe. 2015. PHƯƠNG PHÁP XÂY DỰNG HỆ THỐNG GỢI Ý SẢN PHẨM SỬ DỤNG PHẢN HỒI TIỀM ẨN. Kỷ yếu Hội nghị quốc gia lần thứ VIII "Nghiên cứu cơ bản và ứng dụng Công nghệ thông tin" (FAIR 2015)

- [10] M. Steyvers, T. Griffiths. 2006. Probabilistic topic models. In T. Landauer, D McNamara, S. Dennis, and W. Kintsch (eds), *Latent Semantic Analysis: A Road to Meaning*. Laurence Erlbaum 2006.
- [11] N. Thai-Nghe, L. Drumond, T. Horváth, A. Krohn-Grimberghe, A. Nanopoulos, and L. Schmidt-Thieme, “Factorization techniques for predicting student performance,” in *Educational Recommender Systems and Technologies: Practices and Challenges (ERSAT 2011)*, O. C. Santos and J. G. Boticario, Eds. IGI Global, 2011.
- [12] Nguyễn Thái Nghe, Huỳnh Xuân Hiệp (2012): Ứng dụng kỹ thuật phân rã ma trận đa quan hệ trong xây dựng hệ trợ giảng thông minh, Kỷ yếu Hội thảo quốc gia lần thứ XV: Một số vấn đề chọn lọc của CNTT&TT, trang 470-477. Nhà xuất bản Khoa học và Kỹ thuật.
- [13] Nguyen Thai-Nghe, Tomáš Horváth, Lars Schmidt-Thieme (2011): Context-Aware Factorization Models for Student’s Task Recommendation, in *Proceedings of UMAP 2011*
- [14] Phuong Do, Tu Minh Phuong. A keyword-topic model for contextual advertising. In proceeding SoICT ‘12 Proceedings of the Third Symposium on Information and Communication Technology. Pages 63-70.
- [15] X.H. Phan, C.T. Nguyen, D.T. Le, S. Horiguchi, L.M. Nguyen, and Q.T. Ha. 2011. A Hidden Topic-Based Framework toward Building Applications with Short Web Documents. *Knowledge and Data Engineering, IEEE Transactions on* (Volume 23, Issue 7), page 961-976, July 2011.

Ngày nhận bài: 14/4/2018

Ngày gửi phản biện: 15/4/2018