JOURNAL OF SCIENCE OF HNUE DOI: 10.18173/2354-1075.2015-0054
Educational Sci., 2015, Vol. 60, No. 7A, pp. 70-79
This paper is available online at http://stdb.hnue.edu.vn

XÂY DỰNG HỆ THỐNG TƯ VẤN SẢN PHẨM DỰA TRÊN TƯƠNG TÁC NGƯỜI DÙNG

Vương Quang Phương, Hoàng Đỗ Thanh Tùng
Viện Công nghệ Thông tin, Viện Hàn lâm Khoa học và Công nghệ Việt Nam

Tóm tắt. Hệ thống tư vấn hiện được sử dụng rộng rãi trong nhiều lĩnh vực khác nhau, đặc biệt là trong thương mại điện tử. Với các phương pháp truyền thống, việc dự đoán còn nhiều hạn chế về vấn đề người dùng mới, dữ liệu thưa và chỉ dựa trên một tiêu chí, làm giảm khả năng tư vấn của hệ thống. Bài báo đề xuất một hướng tiếp cận kết hợp giữa thuật toán lọc cộng tác trong các hệ tư vấn truyền thống, với sử dụng trọng số dựa trên đánh giá sự quan tâm của người dùng về sản phẩm theo các tương tác (like, comment, click, view sản phẩm) giúp giải quyết vấn đề người dùng mới và dữ liệu thưa, đồng thời tăng khả năng tư vấn có độ chính xác cao hơn.
Từ khóa: Tư vấn sản phẩm, hệ thống tư vấn, tương tác người dùng, trọng số tương tác

1. Mở đầu

Hệ thống tư vấn (Recommender systems – RS) đang từng bước trở thành một lĩnh vực nghiên cứu quan trọng và được ứng dụng khá thành công trong thực tiễn, giúp người dùng đối phó với vấn đề quá tải thông tin. Hiện nay, RS được nghiên cứu và ứng dụng trong nhiều lĩnh vực khác nhau như: giải trí (phim ảnh, âm nhạc,...), giáo dục đào tạo (gợi ý nguồn tài nguyên học tập như sách, báo,..),… và đặc biệt là trong thương mại điện tử.

Thương mại điện tử bao gồm việc sản xuất, quảng cáo, bán hàng và phân phối sản phẩm được mua bán và thanh toán trên mạng Internet, nhưng được giao nhận có thể hữu hình hoặc giao nhận qua Internet dưới dạng số hoá. Giao dịch loại này đã và đang phát triển rất mạnh trên toàn cầu. Kinh doanh ngày nay và trong tương lai sẽ chuyển dần sang online và chiếm phần lớn doanh thu cũng như thị phần của các ngành nghề.

Trên thế giới, đã có nhiều công ti, tổ chức đã áp dụng thành công hệ thống tư vấn nhằm gợi ý các dịch vụ, sản phẩm và các thông tin cần thiết đến người dùng như: website mua sắm trực tuyến Amazon gợi ý cho khách hàng những sản phẩm mà họ có thể quan tâm, YouTube giới thiệu các video clip cho người xem, gợi ý phim ảnh của Netflix.com [1, 3, 5],... Điều này góp phần làm tăng doanh số bán hàng hoặc số lượng truy cập của người dùng vào hệ thống, đồng thời giúp khách hàng có thể tìm kiếm và được tư vấn được những thông tin thú vị hoặc những sản phẩm mà họ mong muốn dễ dàng hơn.

Ở Việt Nam, mặc dù chưa có nền tảng thương mại điện tử và website lớn như vậy, nhưng cũng có những trang web mua bán online quy mô như: vatgia.com, raovat.com, 123mua.vn,

Ngày nhận bài: 8/8/2015. Ngày nhận đăng: 25/11/2015.
Liên hệ: Vương Quang Phương, e-mail: phuongvq88@gmail.com

rongbay.com, thegioididong.com,… Theo báo cáo của Bộ Công Thương dự báo đến năm 2015, dân số nước ta sẽ khoảng 93 triệu dân, với 45% dân số sử dụng internet, xu hướng mua sắm trực tuyến B2C (Business-To-Customer) gia tăng dự kiến sẽ đạt quy mô tối thiểu trên 3,7 tỉ USD.

Ước tính dân số VN năm 2015	Ước tính tỉ lệ dân số truy cập Internet vào năm 2015	Ước tính giá trị mua hàng trực tuyến mỗi năm của một người năm 2015	Ước tính tỉ lệ truy cập Internet tham gia mua sắm trực tuyến		Ước tính doanh số thu được từ TMĐT B2C 2015
			Mức độ tăng trưởng	Tỉ lệ	
93 triệu dân	45%	150 USD	Cao	70%	4,3 tỷ USD
			Trung bình	65%	4,08 tỷ USD
			Thấp	60%	3,7 tỷ USD

Nguồn: Dự báo của Cục TMĐT và CNTT

Hình 1: Ước tính về thương mại điện tử năm 2015 tại Việt Nam

Hiện cũng đã có nhiều hệ thống website có chức năng tư vấn sản phẩm, tuy nhiên đa phần các hệ thống này chỉ gợi ý dựa trên những thông tin cơ bản như: dựa trên nội dung giống nhau (gợi ý các sản phẩm của cùng cùng thể loại, giá tiền, cùng nhà cung cấp, ... (vatgia.com, enbac.com,...). Qua tìm hiểu những trang web trong nước, vẫn chưa thấy có hệ thống nào có tích hợp kĩ thuật đánh giá sự quan tâm của người dùng về sản phẩm dựa trên các tương tác, trong hệ thống tư vấn gợi ý sản phẩm.

Hệ thống tư vấn thường sử dụng kĩ thuật lọc công tác (collaborative filtering) để đưa ra các dự đoán về sở thích của người dùng (user) đối với các mục tin (items – như sản phẩm, sách, báo, phim,...) mà hệ thống cho là phù hợp nhất thông qua việc sử dụng những đánh giá cho điểm trong quá khứ của người dùng. Lọc cộng tác thường được tiếp cận theo lọc cộng tác dựa vào bộ nhớ, cụ thể là tiếp cận dựa trên người dùng (user-based) và dựa trên các mục tin (item-based). Tuy nhiên chúng có những hạn chế:

- Người dùng mới không nhận được sự tư vấn, do chưa có lịch sử đánh giá sản phẩm để hệ thống tư vấn tiến hành dự đoán.

- Ma trận user – item thưa thì hệ thống tư vấn sẽ hoạt động kém hiệu quả.

Trong bài báo này, chúng tôi giới thiệu một giải pháp trong xây dựng hệ thống tư vấn sản phẩm dựa vào tương tác từ người dùng dựa trên kĩ thuật lọc cộng tác.

Chúng tôi đề xuất sử dụng các tương tác thể hiện sự quan tâm của khách hàng đối với sản phẩm (như comment, like, click, view) và kết hợp với kĩ thuật lọc cộng tác để xây dựng một hệ thống tư vấn sản phẩm cho người dùng. Hệ thống có khả năng giải quyết các hạn chế về vấn đề người dùng mới và ma trận thưa của kĩ thuật lọc cộng tác truyền thống.

2. Nội dung nghiên cứu

2.1. Một số khái niệm

2.1.1. Hệ thống tư vấn

Hệ thống tư vấn thường dựa vào ba khái niệm cơ bản, đó là người dùng (user, tập U), mục tin (item, tập I - như sản phẩm) và những phản hồi rui (feedback/ xếp hạng - rating) của người dùng u trên mục tin i. Có 2 dạng bài toán chính trong RS là dự đoán xếp hạng (rating prediction) dựa vào các giá trị xếp hạng trước đó (là những phản hồi tường minh của người dùng, như xếp hạng từ

1 đến 5, hoặc thích/không thích,...) để dự đoán các giá trị xếp hạng trong tương lai, và dạng gợi ý mục tin (item recommendation) dựa vào các tương tác (như số lần click chuột, thời gian xem sản phẩm, số lượng like, comment cho sản phẩm)… nghĩa là những thông tin mà người dùng không trực tiếp đánh giá/xếp hạng).

	i_1	i_2	…	i_k	…	i_n
U_1	5	?	…	3	…	4
U_2	?	?	…	4	…	5
…	…	…	…	…	…	…
U_k	2	5	…	?	…	3
…	…	…	…	…	…	…
U_m	5	4	…	2	…	?

Hình 2: Ma trận user-item-rating

Trong hệ thống dự đoán, thông tin thường được biểu diễn trong một ma trận user - item. Ở đó, mỗi dòng là một user, mỗi cột là một item, và mỗi ô là một giá trị xếp hạng biểu diễn phản hồi/đánh giá của user trên item tương ứng. Các ô có giá trị là những item mà các user đã xếp hạng trong quá khứ. Những ô trống là những item chưa được xếp hạng (điều đáng lưu ý là mỗi user chỉ xếp hạng cho một vài item trong quá khứ, do vậy có rất nhiều ô trống trong ma trận này chứa giá trị rỗng – còn gọi là ma trận thưa – sparse matrix). Nhiệm vụ chính của hệ thống tư vấn là dựa vào các ô đã có giá trị trong ma trận này, để dự đoán các ô còn trống, sau đó sắp xếp kết quả dự đoán (ví dụ, từ cao xuống thấp) và chọn ra Top-N items theo thứ tự, từ đó gợi ý chúng cho người dùng.

Độ tương quan Pearson [2]

Độ tương quan Pearson đo mức độ quan hệ tuyến tính tồn tại giữa hai đối tượng. Công thức tính độ tương quan giữa 2 người dùng ui và uk như sau:

$$sim_{ik} = corr_{ik} = \frac{\sum_{j=1}^{l}(r_{i,j} - \overline{r_i})(r_{k,j} - \overline{r_k})}{\sqrt{\sum_{j=1}^{l}(r_{i,j} - \overline{r_i})^2 \sum_{j=1}^{l}(r_{k,j} - \overline{r_k})^2}}$$

- l là số sản phẩm mà cả hai người dùng ui và uk cùng đánh giá.
- $\overline{r_i}, \overline{r_k}$ là đánh giá trung bình của người dùng u_i và u_k.
- $r_{i,j}$ là giá trị của ô i, j trong ma trận user - item
- Giá trị của độ tương quan Pearson nằm trong đoạn [-1,1].

Dự đoán đánh giá [2]

Trong thuật toán lọc cộng tác cơ bản thì quan điểm của một người dùng về một sản phẩm nào đó sẽ được coi là có liên quan tới quan điểm của các láng giềng của anh ta về cùng sản phẩm đó. Do đó dự đoán mức độ quan tâm của một người dùng cho một sản phẩm sẽ tính bằng một tổng có trọng số những đánh giá tường minh của các láng giềng anh ta cho sản phẩm đó.

Công thức tính dự đoán mức độ quan tâm của người dùng ua đối với một sản phẩm ij như sau:

$$p_{a,j} = \overline{r_a} + \frac{\sum_{i=1}^{l} (r_{i,j} - \overline{r_i}) * sim_{ai}}{\sum_{i=1}^{l} |sim_{ai}|}$$

- $\overline{r_a}$: Đánh giá trung bình của người dùng a.
- $r_i j = R[i, j]$ trong ma trận user item.
- l: số lượng người đánh giá cho sản phẩm j.

2.1.2. Tương tác người dùng

Nguồn dữ liệu

Mạng xã hội

Mạng xã hội là dịch vụ nối kết các thành viên cùng sở thích trên internet lại với nhau với nhiều mục đích khác nhau không phân biệt không gian và thời gian. Mạng xã hội có những tính năng như chat, e-mail, phim ảnh, voice chat, chia sẻ file, blog và xã luận. Mạng đổi mới hoàn toàn cách cư dân mạng liên kết với nhau và trở thành một phần tất yếu mỗi ngày cho hàng trăm triệu thành viên khắp thế giới. Các dịch vụ này có nhiều phương cách để các thành viên tìm kiếm bạn bè, đối tác: dựa theo group (ví dụ như tên trường hoặc tên thành phố), dựa trên thông tin cá nhân (như địa chỉ e-mail), hoặc dựa trên sở thích cá nhân (như thể thao, phim ảnh, sách báo, hoặc ca nhạc), lĩnh vực quan tâm (như kinh doanh, mua bán),...

Về mặt bản chất, mạng xã hội chính là hình thức làm marketing truyền miệng trên môi trường Internet. "Tiếng lành đồn xa, tiếng dữ đồn xa", với khả năng lan truyền thông tin cực nhanh. Vì thế, thâm nhập các mạng xã hội tập trung đông đảo nhóm khách hàng mục tiêu để hiểu rõ hơn nhu cầu của họ, đồng thời hạn chế tối đa các rủi ro cho doanh nghiệp là việc tất yếu hiện nay.

Website thương mại điện tử

Là trang thông tin điện tử được thiết lập để phục vụ một phần hoặc toàn bộ quy trình của hoạt động mua bán hàng hóa hay cung ứng dịch vụ, từ trưng bày giới thiệu hàng hóa, dịch vụ đến giao kết hợp đồng, cung ứng dịch vụ, thanh toán và dịch vụ sau bán hàng. Dù có độ phủ người dùng thấp hơn mạng xã hội, tuy nhiên việc tự xây dựng các website thương mại điện tử cũng mang lại lợi ích dễ làm chủ cho các doanh nghiệp.

Các loại tương tác chính

Tương tác dạng văn bản (Comment): các tương tác phản hồi dạng văn bản được sử dụng phổ biến nhất bởi đa phần người dùng. Chúng được sử dụng để nêu lên các quan điểm của người dùng về các vấn đề, trong đó có các sản phẩm được giới thiệu hoặc quảng cáo. Đây là cách tường minh nhất để thể hiện ý kiến của mình.

Tương tác dạng click chuột: khác với tương tác dạng comment có thể nhìn thấy trực tiếp bằng mắt thường, tương tác click chuột sẽ không hiển thị rõ ràng như vậy. Đây cũng là một trong những loại tương tác quan trọng thể hiện trên website nói chung và thương mại điện tử nói riêng. Bất cứ lúc nào khách hàng đặt con trỏ của họ lên các đường link về sản phẩm của trang web là họ đang tương tác với sản phẩm đó, cho dù họ có nhấp vào hay không. Nếu họ nhấp vào các phần nhất định của sản phẩm thì đó là tương tác dạng click.

Tương tác like: đây là một cách dễ dàng để thể hiện cho mọi người biết rằng chúng ta thích một thứ gì đó (ví dụ sản phẩm được quảng cáo), mà không muốn để lại một lời comment. Hiện nay trên các trang mạng xã hội hoặc website thương mại điện tử, hình thức tương tác này là rất phổ biến.

Khả năng hiển thị (view của người dùng): bằng cách phân tích dữ liệu khả năng hiển thị và tần suất hiển thị, các doanh nghiệp có thể ước tính số người đã xem sản phẩm của họ và số lần người dùng xem quảng cáo sản phẩm đó trong một khoảng thời gian nhất định. Nếu sản phẩm được hiển thị càng nhiều, khả năng tiếp cận người dùng sẽ lớn hơn, mang lại hiệu quả trong việc tiếp thị sản phẩm và tạo ấn tượng với người dùng.

Tầm quan trọng của các tương tác

Xét trên khía cạnh khách hàng và sản phẩm, người dùng có thể thể hiện sự quan tâm của mình thông qua các tương tác như like, click, view trang web có sản phẩm đó. Hoặc sau khi mua và có những trải nghiệm thực tế, người dùng đưa ra những phản hồi đánh giá hoặc tư vấn cho những khách hàng khác. Hiện nay, có hàng triệu những tương tác như vậy trên các website giới thiệu sản phẩm hoặc bán hàng trực tuyến.

Các dữ liệu này có ý nghĩa rất quan trọng đối với các doanh nghiệp sản xuất hoặc phân phối. Họ có thể biết được cảm nhận của khách hàng về sản phẩm, các đặc tính, chức năng, giá cả, chất lượng dịch vụ ... có đủ sức thu hút được khách hàng hay không, từ đó có những thay đổi cho phù hợp.

Với người tiêu dùng, lợi ích là nhận được sự tư vấn từ chính những người đã từng sử dụng, có được nhiều đánh giá thực tế và cái nhìn tổng quát về sản phẩm trước khi đưa ra các quyết định mua sắm.

Mặc dù có giá trị như vậy, nhưng với những hàng hóa được nhiều khách hàng quan tâm, lượng tương tác có thể rất nhiều, việc thu thập các thông tin và đưa ra những gợi ý cho người dùng là vấn đề cần thiết. Do đó, nhu cầu hiện nay là cần có một công cụ để thu thập và phân tích các tương tác, đưa ra những kết quả mà người dùng mong muốn một cách nhanh chóng.

2.2. Đề xuất phương pháp tư vấn dựa trên sự quan tâm của khách hàng

2.2.1. Tính điểm các tương tác

Nếu chỉ dùng công thức tính ra điểm độ quan tâm như phương pháp truyền thống thì trong thực tế chúng tôi thấy các vấn đề phát sinh:

- Khi người dùng mới đăng kí, họ sẽ chưa nhận được sự tư vấn ngay. Nguyên nhân của vấn đề này là do chưa có dữ liệu lịch sử về các đánh giá trong quá khứ, làm cho thuật toán lọc cộng tác không hoạt động được.

- Nếu người dùng mới, có thời gian sử dụng ít và lượng đánh giá sản phẩm còn nhỏ, các gợi ý có thể không sát với mong muốn của họ.

Như đã nêu ở phần trên, hiện nay có rất nhiều loại tương tác khác nhau có thể được thu thập. Chúng thể hiện rất tốt sự quan tâm và đánh giá của người dùng đối với sản phẩm. Từ đó chúng tôi đưa ra ý tưởng kết hợp tính điểm các tương tác này để trở thành một trọng số trong kết quả tư vấn đầu ra, giúp giải quyết các vấn đề:

- Người dùng mới đăng kí, chưa có lịch sử đánh giá hàng hóa, nhưng dựa vào trọng số tương tác, hệ thống tư vấn hoàn toàn có thể đưa ra được gợi ý những sản phẩm được những người dùng khác đánh giá tốt nhất.

- Khi ma trận user – item có dữ liệu thưa, thì trọng số cũng giúp cải thiện được kết quả đầu ra. Gợi ý những sản phẩm có được đánh giá cao nhất.

- Các sản phẩm được đánh giá cao nhất được tính toán tổng hợp từ nhiều người dùng và nhiều tiêu chí, chứ không chỉ tập trung ở một khía cạnh nào. Tạo sự khách quan về chất lượng trong hệ thống tư vấn.

Để thực hiện được công việc tính điểm này, có thể sử dụng nhiều phương pháp. Đơn giản nhất là có thể cho điểm số từng loại tương tác rồi cộng điểm lại. Ở đây chúng tôi sẽ sử dụng phương pháp tính điểm số dựa trên phân lớp dữ liệu. Tức là, từ tập dữ liệu thô chúng tôi thu nhận, sẽ được đưa vào một thuật toán phân lớp để đưa ra đánh giá của người dùng. Thuật toán phân lớp K-Nearest Neighbors (KNN) được lựa chọn thực hiện.

Thuật toán KNN [4]

KNN được sử dụng rất phổ biến trong lĩnh vực Data Mining. KNN là phương pháp để phân lớp các đối tượng dựa vào khoảng cách gần nhất giữa đối tượng cần xếp lớp với tất cả các đối tượng trong Training Data.

Một đối tượng được phân lớp dựa vào k láng giềng của nó. K là số nguyên dương được xác định trước khi thực hiện thuật toán. Khoảng cách Euclidean thường được dùng để tính khoảng cách giữa các đối tượng. Không gian Euclide n chiều là một tập hợp các điểm, trong đó mỗi điểm có thể được thể hiện như $(x(1), x(2), ..., x(n))$, trong đó $x(i)(i = 1, 2, ..., n)$ là một số thực, được gọi là tọa độ x của thứ i, hai điểm x và $y = (y(1), y(2), ..., y(n))$ khoảng cách giữa $d(x, y)$ được tính bằng công thức sau:

$$d(x, y) = ||x - y|| = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}.$$

Thuật toán KNN dùng trong phân lớp được mô tả như sau:

- Xác định giá trị tham số K (số láng giềng gần nhất).

- Tính khoảng cách giữa đối tượng cần phân lớp với tất cả các đối tượng trong dữ liệu huấn luyện.

- Sắp xếp khoảng cách theo thứ tự tăng dần và xác định k láng giềng gần nhất với đối tượng cần phân lớp.

- Lấy tất cả các lớp của k láng giềng gần nhất đã xác định.

- Dựa vào phần lớn lớp của láng giềng gần nhất để xác định lớp cho đối tượng.

Xác định trọng số

Sau khi phân loại các sản phẩm vào trong các lớp, điểm trọng số cho các lớp tương ứng được quy ước như sau.

Lớp	Trọng số
Không tốt	0.0
Trung bình	0.5
Khá tốt	1.0
Tốt	1.5
Rất tốt	2.0

2.2.2. Ứng dụng vào hệ thống tư vấn

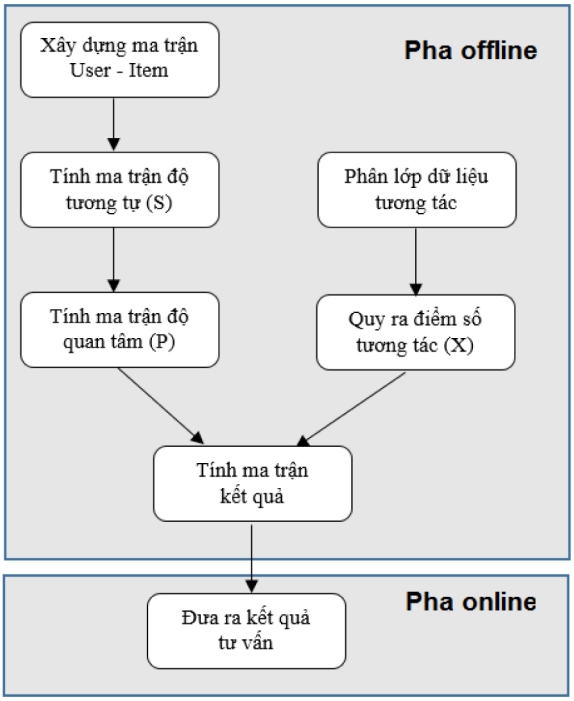
Hệ thống tư vấn sử dụng phương pháp lọc cộng tác kết hợp với tính điểm các tương tác, hoạt động của hệ thống gồm hai pha:

- Pha offline: Tiến hành tính độ quan tâm của người dùng và điểm số các tương tác.

- Pha online: Đưa ra tư vấn và dự đoán đánh giá cho người dùng dựa vào việc các dữ liệu

đã được tính toán ở pha offline.

Cụ thể các bước thực hiện chính được mô hình hóa:



Hình 3: Mô hình hoạt động của hệ thống tư vấn

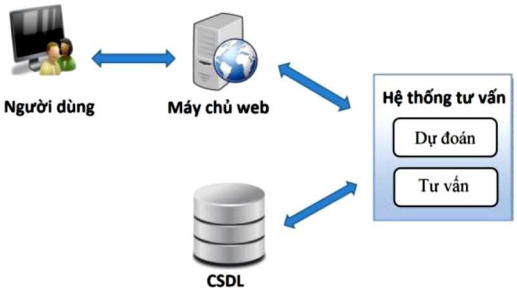
2.3. Thực nghiệm

Nguồn dữ liệu

Trong rất nhiều các website thương mại diện tử tại Việt Nam, chúng tôi muốn lựa chọn một nguồn sao cho lấy được thông tin có chất lượng. Để làm được điều đó thì cần đến các thước đo quy chuẩn để so sánh các website với nhau. Sau khi đánh giá một số website thương mại điện tử phổ biến tại Việt Nam dựa trên công cụ Alexa (www.alexa.com), chúng tôi quyết định lựa chọn vatgia.com để tiến hành thực nghiệm trong bài viết này. Với số lượng sản phẩm rất lớn, đa dạng về chủng loại và lượng khách hàng truy cập nhiều.

Xây dựng hệ thống

Chúng tôi tiến hành xây dựng hệ thống tư vấn sản phẩm được xây dựng dưới dạng một website cung cấp thông tin về sản phẩm giúp người dùng có thể chọn các hàng hóa mà mình yêu thích.



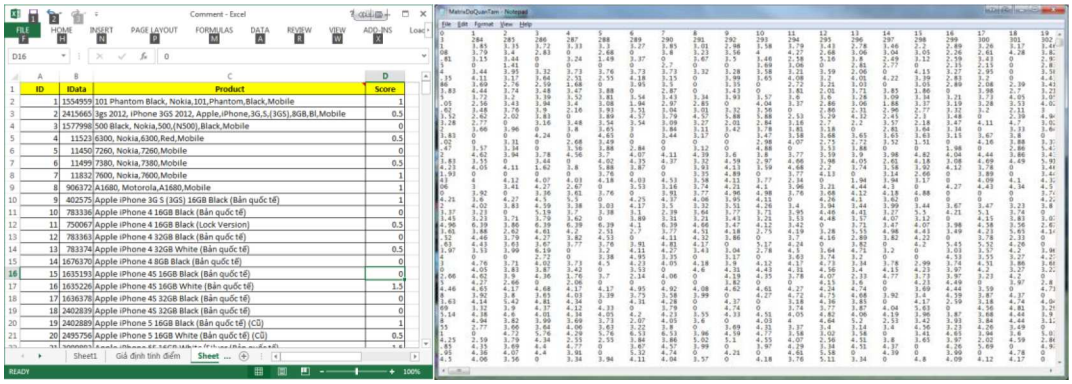
Hình 4: Kiến trúc hệ thống

Mô hình hệ thống gồm các thành phần: người dùng, máy chủ web, module tư vấn và CSDL. Người dùng truy cập vào website có thể xem thông tin sơ bộ về các sản phẩm. Khi đăng nhập vào hệ thống, người dùng sẽ có quyền đánh giá các sản phẩm và dùng chức năng tư vấn các

mặt hàng mình có thể yêu thích. Module tư vấn sẽ chạy trước các tính toán trong pha offline, đầu ra là ma trận kết quả tư vấn. Khi người dùng kích hoạt chức năng này, hệ thống sẽ chỉ việc vào CSDL và lấy ra các tư vấn cần thiết, nên tốc độ đưa ra kết quả là rất nhanh.

- Input: Trong thử nghiệm này, chúng tôi chạy chương trình trên bộ dữ liệu có 1058 user và 386 sản phẩm (trong danh mục mặt hàng điện thoại di động).

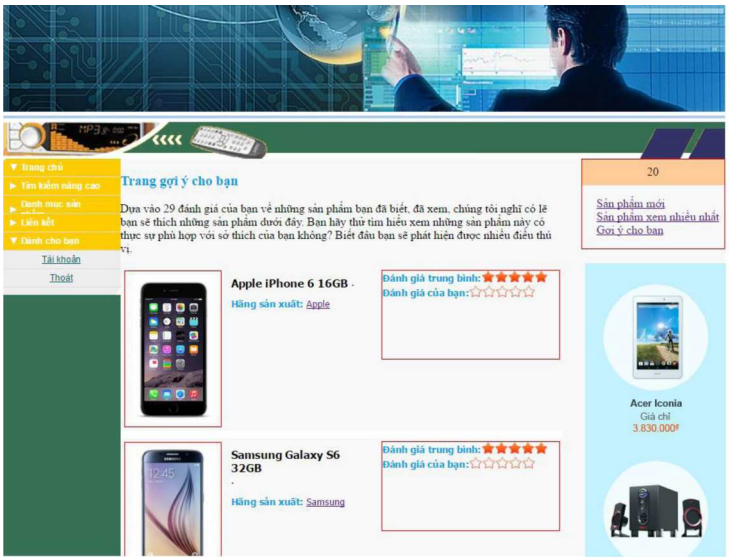
- Output: ma trận điểm trọng số từng sản phẩm và ma trận kết quả dự đoán.



Điểm trọng số *Ma trận kết quả dự đoán*

Hình 5: Ma trận đầu ra

Hệ thống tư vấn sản phẩm được phát triển trên môi trường .NET, dùng hệ quản trị cơ sở dữ liệu SQL Server của Microsoft, có thể dễ dàng triển khai trong thực tế. Giao diện chính của hệ thống:



Hình 6: Giao diện chính hệ thống tư vấn

2.4. Kết quả

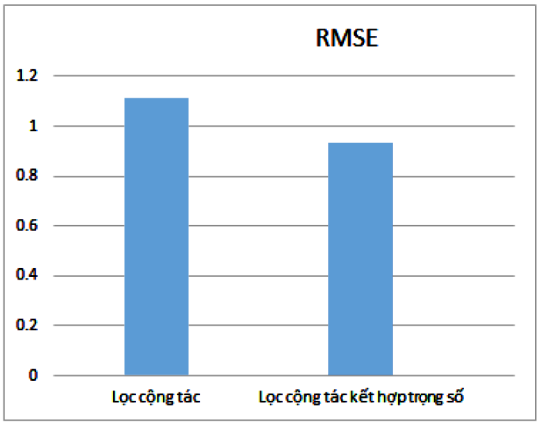
Để đánh giá hiệu quả của hệ thống, chúng tôi sử dụng độ đo mà cộng đồng người dùng trong lĩnh vực hệ thống gợi ý thường sử dụng là Root Mean Squared Error (RMSE). RMSE được

xác định qua công thức:

$$RMSE = \sqrt{\frac{\sum_{u,i,r} \in D^{test}(r - \hat{r}_{u,i})^2}{|D^{test}|}}$$

Với: $D^{\text{test}} \subseteq (UxIxR)$ là tập dữ liệu dùng để đánh giá; (u, i, r) tương ứng là user, item và rating; r là giá trị đánh giá thực tế, còn $r^{\wedge}(u, i)$ là giá trị dự đoán tương ứng của user u trên item i. Kết quả thu được:

Phương pháp	RMSE
Lọc cộng tác	1.114771
Lọc cộng tác kết hợp trọng số	0.934826



Hình 7: So sánh RMSE giữa 2 phương pháp

Sự khác biệt giữa đề xuất trong bài báo này:

- Người dùng mới vẫn nhận được kết quả tư vấn, những tư vấn này dựa trên các dữ liệu khách quan của các người dùng khác, nên có độ tin cậy cao hơn.

- Khi người dùng có ít đánh giá, tức trường hợp ma trận thưa, trọng số tương tác sẽ bổ trợ cho kết quả tư vấn, đưa ra kết quả chính xác hơn.

3. Kết luận

Với lượng dữ liệu tương tác (like, comment, click, view) của người dùng với các sản phẩm rất lớn hiện nay, đây là nguồn dữ liệu quan trọng để lĩnh vực khai phá dữ liệu và các hệ tư vấn có thể khai thác. Trong bài báo này, chúng tôi đã thu thập lại những dữ liệu tương tác. Từ đó, đề xuất một hướng tiếp cận kết hợp giữa kết hợp giữa thuật toán lọc cộng tác trong các hệ tư vấn truyền thống với sử dụng trọng số dựa trên đánh giá sự quan tâm của người dùng về sản phẩm theo các tương tác. Tính toán thực nghiệm cho thấy, hệ thống đưa ra các tư vấn tốt hơn phương pháp truyền thống, đặc biệt là với trường hợp người dùng mới đăng kí, chưa có lịch sử đánh giá và dữ liệu đánh giá còn thưa thớt.

TÀI LIỆU THAM KHẢO

[1] Andreas Toscher, Michael Jahrer, Robert Legenstein, 2008. *Improved Neighborhood-Based Algorithms for Large-Scale Recommender Systems*. Las Vegas, USA, Availble from http://www.commendo.at/UserFiles/commendo/File/kdd08paper.pdf.

[2] Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl, 2001. *Item-Based Collaborative Filtering Recommendation Algorithms*. Hong Kong.
Availble from http://grouplens.org/site-content/uploads/Item-Based-WWW-2001.pdf

[3] Greg Linden, Brent Smith, and Jeremy York, 2003. *Amazon.com Recommendations Item-to-Item Collaborative Filtering*. Amazon.com, IEEE Internet Computing, pp. 76-80.

[4] Hao Zhang, Alexander C. Berg, Michael Maire, Jitendra Malik, 2006. *SVM-KNN: Discriminative Nearest Neighbor Classification for Visual Category Recognition*. CVPR (2) 2006: 2126-2136.

[5] James Davidson, Benjamin Liebald, Junning Liu, Palash Nandy, Taylor Van Vleet, 2010. *The YouTube Video Recommendation System*. RecSys '10 Proceedings of the fourth ACM conference on Recommender systems, Pages 293-296, ACM New York, NY, USA.

[6] L. Backstrom and J. Leskovec, 2011. *Supervised Random Walks: Predicting and Recommending Links in Social Networks*. Proc. Fourth ACM Int'l Conf. Web Search and Data Mining (WSDM '11), pp. 635-644.

[7] L. Zhang, D. Agarwal, and B.-C. Chen, 2011. *Generalizing Matrix Factorization through Flexible Regression Priors*. Proc. Fifth ACM Conf. Recommender Systems (RecSys '11), pp. 13-20.

[8] N. Suguna, and Dr. K. Thanushkodi, 2010. *An Improved k-Nearest Neighbor Classification Using Genetic Algorithm*. IJCSI International Journal of Computer Science Issues, Pages 18-21, Vol. 7, Issue 4, No 2.

[9] Nguyễn Thái Nghe, Nguyễn Tấn Phong, 2014. *Một giải pháp trong xây dựng Hệ thống gợi ý bài hát*. Tạp chí Khoa học Trường Đại học Cần Thơ, Số 34, 2014, Trang 81-91.

[10] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, Lars Schmidt-Thieme, 2009. *BPR: Bayes Personalized Ranking from implicit feedback*. In Proceedings of the 25th International Conference on Uncertainty in Articial Intelligence, AUAI Press, pp. 452-461.

[11] Ungar, L.H. and D.P. Foster, 1998. *Clustering methods for collaborative filtering*. In Proceedings of the Workshop on Recommendation Systems, AAAI Press, Menlo Park, California, pp. 452-461.

ABSTRACT

Building a product recommendation based on user interactions

Today, recommender systems (RS) are widely used in many different fields, particularly l in e-commerce. With the traditional method, predictions are affected by many restrictions in some issue of new users, sparse data and data based on only one criterion, all of which reduce the recommend ability of a system. This paper proposes an approach that combines a collaborative filtering algorithm in the traditional recommender systems with the use of weights based on assessed interest of the user about the product as indicated by interactions (like, comment, click, view) to help solve the issues of new users and sparse data, increasing recommend ability with higher accuracy.

Keywords: Product recommendation, recommender systems, user interaction, weight of interaction.