



Tạp chí Khoa học Trường Đại học Cần Thơ
website: sj.ctu.edu.vn



HỆ THỐNG GỢI Ý SẢN PHẨM TRONG BÁN HÀNG TRỰC TUYẾN SỬ DỤNG KỸ THUẬT LỌC CỘNG TÁC

Nguyễn Hùng Dũng¹ và Nguyễn Thái Nghe²

¹ Trung tâm Công nghệ Phần mềm, Trường Đại học Cần Thơ

² Khoa Công nghệ Thông tin & Truyền thông, Trường Đại học Cần Thơ

Thông tin chung:

Ngày nhận: 04/12/2013

Ngày chấp nhận: 28/04/2014

Title:

Building a product
recommender system in an
online shopping system using
collaborative filtering
technique

Từ khóa:

Lọc cộng tác, hệ thống gợi ý

Keywords:

Collaborative filtering,
recommender systems

ABSTRACT

Recommender system(s) can recommend suitable items to users by using data about their past behavior to predict the future items that the users may like. In the recommender system, collaborative filtering is a technique which is used to assess the user's interest on new items (products). This technique is successfully applied in many applications. In the collaborative filtering systems, the users' preference on new items is predicted based on the past user-item-preference (or user-item-rating) data. In this paper, we introduce the collaborative filtering technique which is based on neighborhood model to recommend products in an online shopping system. We implement this technique and compare its quality to the other baselines. Finally, we build an online shopping system to integrate the collaborative filtering technique to our system so that it can recommend suitable products to the users.

TÓM TẮT

Hệ thống gợi ý có thể đưa ra những mục thông tin phù hợp cho người dùng bằng cách dựa vào dữ liệu về hành vi trong quá khứ của họ để dự đoán những mục thông tin mới trong tương lai mà người dùng có thể thích. Trong hệ thống gợi ý, lọc cộng tác là một kỹ thuật được dùng để đánh giá độ quan tâm của người dùng trên các sản phẩm mới. Kỹ thuật này được áp dụng thành công trong nhiều ứng dụng. Trong các hệ thống lọc cộng tác, sở thích của người dùng trên các sản phẩm mới được dự đoán dựa trên dữ liệu về sở thích của người dùng – sản phẩm (hoặc đánh giá của người dùng trên sản phẩm) trong quá khứ. Trong bài viết này, chúng tôi giới thiệu kỹ thuật lọc cộng tác dựa trên mô hình láng giềng (mô hình lân cận) để gợi ý sản phẩm trong hệ thống bán hàng trực tuyến. Chúng tôi cài đặt kỹ thuật này và so sánh độ tin cậy của nó với các kỹ thuật cơ bản khác. Cuối cùng, chúng tôi xây dựng một hệ thống bán hàng trực tuyến có tích hợp kỹ thuật lọc cộng tác vào hệ thống để nó có thể gợi ý sản phẩm phù hợp cho người dùng.

1 GIỚI THIỆU

Trong những năm gần đây, sự phát triển của thương mại điện tử (E-Commerce) đã đem lại nhiều lợi ích to lớn cho nền kinh tế toàn cầu.

Thông qua thương mại điện tử, nhiều loại hình kinh doanh mới được hình thành, trong đó có mua bán hàng qua mạng. Với hình thức mới này, người tiêu dùng có thể tiếp cận với hàng hóa một cách dễ

dễ dàng và nhanh chóng hơn rất nhiều so với phương thức mua bán truyền thống.

Hiện nay, các hệ thống bán hàng trực tuyến đã tạo nhiều điều kiện thuận lợi để người mua có thể tiếp cận nhiều mặt hàng cùng lúc. Tuy nhiên, việc trình bày và trang trí quá nhiều các mặt hàng trên trang web đã gây ra không ít khó khăn cho người mua. Họ khó có thể chọn ra cho mình một sản phẩm ưng ý nhất.

Để khách hàng có thể đến và mua được một sản phẩm ưng ý thì một lời tư vấn, một sự trợ giúp là rất quan trọng. Trong phương thức bán hàng truyền thống những lời tư vấn như thế từ một người bán hàng sẽ tạo ra một lợi thế rất lớn cho cửa hàng. Do đó, để phương thức bán hàng qua mạng thực sự phát triển thì bên cạnh các lợi thế vốn có của mình việc có thêm một “người trợ giúp” là hết sức cần thiết.

Một hệ thống gợi ý tốt có thể đóng vai trò như một người trung gian hỗ trợ khách hàng đưa ra các quyết định mua hàng đúng đắn. Bằng cách xác định mục đích và nhu cầu của khách hàng, hệ thống có thể đưa ra một tập các gợi ý giúp cho người mua dễ dàng chọn lựa sản phẩm yêu thích hơn. Qua đó hiệu suất của việc mua bán hàng trực tuyến được tăng cao một cách đáng kể.

Lọc cộng tác (collaborative filtering) là một kỹ thuật mạnh và nó đã được áp dụng khá thành công trong nhiều hệ tư vấn. Thực chất, lọc cộng tác là một hình thức tư vấn tự động bằng cách dựa trên sự tương tự giữa những người dùng hoặc giữa những sản phẩm trong hệ thống và đưa ra dự đoán sự quan tâm của người dùng tới một sản phẩm, hoặc đưa ra gợi ý một sản phẩm mới cho người dùng nào đó.

Đã có khá nhiều bài viết về việc sử dụng hệ tư vấn cho lọc cộng tác, chẳng hạn như: bài báo "Lọc cộng tác được áp dụng để khai phá dữ liệu trong giáo dục (Collaborative Filtering Applied to Educational Data Mining)" (A. Töschler và M. Jahrer, 2010); Ứng dụng RS trong dự đoán kết quả học tập của sinh viên, từ đó đưa ra những gợi ý phù hợp theo năng lực của sinh viên (N. Thai-Nghe *et al.*, 2010 và 2011); "Grouplens: an open architecture for collaborative filtering of netnews" (P. Resnick *et al.*, 1994); "An Empirical Analysis of Design Choices in Neighborhood-Based Collaborative Filtering Algorithms" (L. Herlocker *et al.*, 1999); "Hệ thống Ringo (Ringo system)" (U. Shardanand và P. Maes, 1995) mô tả hệ thống tư vấn nhạc dựa trên lọc cộng tác sử dụng độ tương tự

Pearson... Mặc dù vậy, việc xây dựng một hệ thống gợi ý hoàn chỉnh, có tích hợp giải thuật lọc cộng tác để tư vấn cho người dùng vẫn chưa được quan tâm.

Trong bài viết này, chúng tôi sẽ giới thiệu kỹ thuật lọc cộng tác trong hệ thống gợi ý, cài đặt và kiểm nghiệm kỹ thuật này, đồng thời so sánh với các phương pháp khác để thấy mức độ tin cậy của nó. Quan trọng hơn, hiện nay có khá nhiều nghiên cứu về các giải thuật trong hệ thống gợi ý, tuy vậy, việc cài đặt một hệ thống thực sự có tích hợp giải thuật gợi ý vào trong đó thì vẫn còn là lĩnh vực mới mẻ và chưa thấy hệ thống nào ở Việt Nam có khả năng làm như thế. Chúng tôi 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. Việc tích hợp thành công này sẽ mở ra một hướng mới cho các hệ thống bán hàng trực tuyến hiện nay - đa phần chưa cá nhân hóa cho từng người dùng.

Cấu trúc của bài viết được trình bày như sau: chúng tôi bắt đầu giới thiệu về lọc cộng tác - mô hình láng giềng trong phần 2, kể cả việc phân tích ưu điểm và khuyết điểm của lọc cộng tác. Sau đó, chúng tôi trình bày thuật toán của mô hình láng giềng trong lọc cộng tác ở phần 3. Phần 4 sẽ trình bày việc cài đặt và đánh giá giải thuật lọc cộng tác. Phần 5, chúng tôi xây dựng hoàn chỉnh hệ thống gợi ý sản phẩm trong bán hàng trực tuyến có tích hợp kỹ thuật lọc cộng tác được chạy trên nền web. Phần 6, đưa ra kết luận về kết quả nghiên cứu và nêu hướng phát triển tiếp theo của phương pháp đã đề xuất.

2 LỌC CỘNG TÁC

Trong cuộc sống hàng ngày, mọi người thường tin vào những lời giới thiệu từ những người khác thông qua lời nói, thư từ văn bản, các nguồn tin thu được trên các phương tiện thông tin đại chúng. Về bản chất, lọc cộng tác cũng chính là hình thức tư vấn như trên, tuy nhiên, việc này được thực hiện tự động bởi các máy tính. Các nhà phát triển hệ thống gợi ý đầu tiên vào năm 1992 (D. Goldberg *et al.*, 1992) đã đưa ra thuật ngữ “collabrative filtering” (lọc cộng tác). Giả thuyết của lọc cộng tác là: “Nếu người dùng u và u' đánh giá cho n sản phẩm tương tự nhau, hoặc có hành vi tương tự nhau (như: xem, mua, nghe...) thì họ sẽ có các đánh giá tương tự nhau đối với các sản phẩm khác”.

Trong hệ thống gợi ý, người ta thường biểu diễn các đánh giá của người dùng cho các sản

phẩm qua ma trận gồm một tập người dùng U và tập sản phẩm I như Hình 1.

	i_1	i_2	...	i_j	...	i_k
u_1						
....						
u_m						
....						
u_n						

Hình 1: Ma trận $U \times I$

Chúng tôi dùng các ký tự để phân biệt người dùng và các mục tin: u, u' (đại diện cho người dùng) - i, i' (đại diện cho các sản phẩm). Ký hiệu r_{ui} để chỉ mức độ thích của người dùng u cho một sản phẩm i nào đó, chẳng hạn giá trị này trong khoảng từ 1 (không thích) đến 5 (thích nhất) đối với dữ liệu bán hàng, \hat{r}_{ui} là một hàm dùng để dự đoán đánh giá của người dùng u cho sản phẩm i (có thể xem như hàm tiện ích).

Theo mô hình láng giềng, việc dự đoán đánh giá của một người dùng trên một mục tin (item) được dựa trên các đánh giá của những người dùng “lân cận” trên mục tin đó. Vì vậy, một độ đo tương tự (similarity) giữa những người dùng cần được định nghĩa trước khi một tập những người dùng láng giềng gần nhất được chọn ra.

Hệ tư vấn dựa trên lọc cộng tác (hệ lọc cộng tác) sẽ dự đoán hàm tiện ích của những sản phẩm cho những người dùng cụ thể dựa trên trọng số ban đầu của sản phẩm được cung cấp bởi người dùng khác. Giống như trước, hàm tiện ích \hat{r}_{ui} của sản phẩm i cho người dùng u được đánh giá dựa trên những hàm tiện ích $\hat{r}(u_j, i)$ được gán bởi sản phẩm i và người dùng $u_j \in U$, đây là những người mà có cùng sở thích và thị hiếu giống u . Chẳng hạn, trong ứng dụng về tư vấn sách, để tư vấn những sách mới cho người dùng u , hệ tư vấn lọc cộng tác sẽ tìm những người tương tự với người dùng u ; nghĩa là những người dùng khác nhau nhưng có cùng thị hiếu về sách (trọng số về các sách giống nhau là như nhau). Sau đó, chỉ những quyển sách có độ quan tâm nhất sẽ được tư vấn cho người dùng.

Mở rộng của vấn đề này là tìm tập N sản phẩm mà rất có thể người dùng sẽ quan tâm – công việc chính trong các hệ tư vấn hiện nay. Giả sử rằng

chúng ta có một hệ thống dự đoán các đánh giá của người dùng lên các sản phẩm mới, và dựa vào đó ta có thể lấy ra N sản phẩm được dự đoán có trọng số cao nhất. Cuối cùng, hai vấn đề quan trọng nhất đó là làm thế nào để tìm được tập K người dùng tương tự với người dùng u nhất dựa trên thị hiếu về sản phẩm hay làm thế nào để tìm ra tập K sản phẩm tương tự với sản phẩm i nhất để có thể đưa ra các dự đoán trọng số đánh giá của người dùng và sản phẩm tương ứng.

3 THUẬT TOÁN VỀ MÔ HÌNH LĂNG GIỀNG TRONG LỌC CỘNG TÁC USER_KNN (K-NEAREST NEIGHBORHOOD)

3.1 Thuật toán

Phương pháp lọc cộng tác có đặc trưng cơ bản là nó thường sử dụng toàn bộ dữ liệu đã có để dự đoán đánh giá của một người dùng nào đó về sản phẩm mới. Nhờ lợi thế là nó có khả năng đưa trực tiếp dữ liệu mới vào bảng dữ liệu, do đó nó đạt được khá nhiều thành công khi được áp dụng vào các ứng dụng thực tế. Cũng do đó mà các kỹ thuật này thường đưa ra các dự đoán chính xác hơn trong các hệ trực tuyến – nơi mà ở đó luôn có dữ liệu mới được cập nhật.

Thông thường, có hai cách tiếp cận của lọc cộng tác theo mô hình K láng giềng: hệ dựa trên người dùng (User_KNN) – tức dự đoán dựa trên sự tương tự giữa các người dùng và hệ dựa trên sản phẩm (Item_KNN) – dự đoán dựa trên sự tương tự giữa các sản phẩm. Hệ dựa trên người dùng (User_KNN) xác định sự tương tự giữa hai người dùng thông qua việc so sánh các đánh giá của họ trên cùng sản phẩm, sau đó dự đoán đánh giá sản phẩm i bởi người dùng u , hay chính là đánh giá trung bình của những người dùng tương tự với người dùng u . Độ tương tự giữa người dùng u và người dùng u' có thể được tính theo Pearson (L. Herlocker *et al.*, 1999) vì phân tích thực nghiệm cho thấy rằng đối với hệ dựa trên người dùng thì tính độ tương tự theo Pearson sẽ tốt hơn so với một vài cách khác như độ tương tự theo cấp bậc của Spearman (Spearman's rank correlation) hay độ tương tự theo bình phương trung bình (mean squared difference). Công thức tính độ tương tự theo Pearson như sau:

$$\text{sim}_{\text{pearson}}(u, u') = \frac{\sum_{i \in I_{uu'}} (r_{ui} - \bar{r}_u)(r_{u'i} - \bar{r}_{u'})}{\sqrt{\sum_{i \in I_{uu'}} (r_{ui} - \bar{r}_u)^2 \sum_{i \in I_{uu'}} (r_{u'i} - \bar{r}_{u'})^2}} \quad (\text{CT } 1)$$

Ở đây:

$I_{uu'}$ là một tập các item được đánh giá bởi u và u'

\bar{r}_u là giá trị đánh giá trung bình trên tất cả các item của người dùng u .

$\bar{r}_{u'}$ là giá trị đánh giá trung bình trên tất cả các item của người dùng u' .

Đưa ra được những dự đoán hoặc lời gợi ý là một bước quan trọng trong hệ tư vấn lọc cộng tác. Sau khi tính toán độ tương tự giữa các người dùng hay giữa các sản phẩm, chúng ta có thể dự đoán đánh giá của người dùng u trên sản phẩm i theo công thức (P. Resnick *et al.*, 1994) như sau:

$$\hat{r}_{ui} = \bar{r}_u + \frac{\sum_{u' \in K_u} \text{sim}(u, u') \cdot (r_{u'i} - \bar{r}_{u'})}{\sum_{u' \in K_u} |\text{sim}(u, u')|} \quad (\text{CT } 2)$$

Ở đây:

\hat{r}_{ui} chính là dự đoán cho người dùng u trên sản phẩm i

$\text{Sim}(u, u')$ độ tương tự giữa người dùng u và u' .

K_u là số người dùng có độ lân cận gần người dùng u .

Chúng tôi biểu diễn giải thuật lọc cộng tác dựa trên người dùng lân cận gần nhất (User_KNN) sử dụng độ tương tự Pearson bằng ngôn ngữ giả để dự đoán độ thích cho người dùng u trên sản phẩm i như sau:

```

1: procedure USERKNN-CF ( $\bar{r}_u, r, D^{\text{train}}$ )
2: for  $u=1$  to  $N$  do
3:   Tính  $\text{Sim}_{uu'}$ , sử dụng công thức (CT 1)
4: end for
5: Sort  $\text{Sim}_{uu'}$  // sắp xếp giảm dần độ tương tự
6: for  $k=1$  to  $K$  do
7:    $K_u \leftarrow k$  // Các người dùng  $k$  gần nhất của  $u$ 
8: end for
9: for  $i = 1$  to  $M$  do
10:  Tính  $\hat{r}_{ui}$ , sử dụng công thức (CT 2)
11: end for
12: end procedure

```

Ở đây:

\bar{r}_u : đánh giá trung bình của người dùng u trên tất cả các item

r : đánh giá của người dùng trên tập huấn luyện

K : người dùng k gần nhất

N, M : người dùng thứ n và thứ m

D^{train} : tập dữ liệu huấn luyện

3.2 Ưu điểm và khuyết điểm của lọc cộng tác

3.2.1 Ưu điểm

Xét một cách tổng quát thì hệ này không yêu cầu quá nặng vào việc tính toán và do đó có thể đưa ra những tư vấn có độ chính xác cao và nhanh chóng cho một số lượng lớn người dùng. Bởi vì hệ này không yêu cầu mô tả nội dung một cách tường minh mà chỉ dựa vào sự đánh giá của người dùng để ước lượng xem có bao nhiêu người dùng đồng ý và không đồng ý. Đó cũng chính là lý do tại sao mà hệ này có thể đưa ra những gợi ý bất ngờ cho người dùng.

3.2.2 Khuyết điểm

Có một số nhược điểm trong lọc cộng tác nhưng chúng tôi đưa ra và khắc phục 2 nhược điểm cơ bản sau:

- Vấn đề về người dùng mới.
- Vấn đề về item mới.

3.2.3 Cách khắc phục

Để khắc phục vấn đề trên, chúng tôi sẽ kết hợp cả lọc cộng tác (collaborative filtering) và lọc dựa trên một số thuộc tính của người dùng cung cấp để dự đoán sở thích của khách hàng mới. Cụ thể, thông qua việc đăng ký thông tin tài khoản của khách hàng, chúng tôi sẽ thu thập thêm một số thông tin của người dùng mới bằng cách tạo một biểu mẫu (form) cho người dùng nhập một số thông tin cần thiết như: tính cách, nghề nghiệp, thu nhập hoặc một số thông tin về sở thích... Việc ghi nhận lại các thông tin này của người dùng sẽ giúp hệ thống tư vấn tốt hơn cho người dùng mới này. Đối với sản phẩm mới nhập vào, chúng được hiển thị đầu tiên trên trang web và có biểu tượng 'New' để nhận biết đây là sản phẩm mới của hệ thống. Ngoài ra, khi hiển thị chi tiết mỗi sản phẩm, trang web có một không gian để hiển thị các sản phẩm tương tự với sản phẩm mà người dùng đang xem dựa vào một số thuộc tính tương tự. Vì vậy, những sản phẩm mới nhập cũng có thể được gợi ý cho người dùng.

4 KIỂM CHỨNG THỰC NGHIỆM

4.1 Dữ liệu thực nghiệm

Với nghiên cứu này, chúng tôi sử dụng tập dữ liệu MovieLens 100K (đây là dữ liệu chuẩn dùng để đánh giá giải thuật lọc cộng tác, dữ liệu này được tập hợp từ các đánh giá của người dùng tại website <http://movielens.umn.edu>) với khoảng 800

người dùng, 9000 bộ phim và có trên 100.000 đánh giá. Chúng tôi chia tập dữ liệu ra làm 2 phần, lấy ngẫu nhiên 2/3 tập dữ liệu làm dữ liệu đầu vào và 1/3 dùng để test. Trong 2/3 tập dữ liệu đầu vào này, chúng tôi lại chia tiếp 70% tập này làm tập dữ liệu để tìm tham số K. Từ dữ liệu đầu vào hệ thống dự đoán r_{ui} và so sánh dữ liệu với tập test để đo độ lệch của hệ thống. Sau khi tìm được tham số K, chúng tôi lại kiểm tra trên 1/3 tập dữ liệu ban đầu.

4.2 Độ đo

Chúng tôi sử dụng Root Mean Squared Error (RMSE) để đánh giá hiệu quả của thuật toán, RMSE được xác định bằng công thức:

$$RMSE = \sqrt{\frac{1}{|D^{test}|} \sum_{u,i,r \in D^{test}} (r_{ui} - \hat{r}_{ui})^2} \quad (CT\ 3)$$

D^{test} : tập kiểm tra.

r_{ui} : dự đoán của người dùng u trên item i trên tập kiểm tra.

\hat{r}_{ui} : dự đoán của người dùng u trên item i theo công thức (CT 2).

4.3 Tìm tham số và kết quả

Khi huấn luyện, chúng tôi thử tìm kiếm K sử dụng kỹ thuật tìm kiếm lưới (grid search = raw search + smooth search) để đạt được K cho lỗi RMSE trên tập kiểm tra tốt nhất.

Sau đây, chúng tôi thống kê một số kết quả theo từng K láng giềng:

Bảng 1: Kết quả tìm K lần 1 với dữ liệu ngẫu nhiên

Tham số K (dùng 30% của tập dữ liệu đầu vào để xác định các tham số)	RMSE trên tập validation	Thời gian (giờ)
8	0.9364809	10.07
16	0.9216916	10.30
32	0.9184653	11.03
64	0.9188190	11.22

Tiếp theo chúng tôi tìm kiếm mịn hơn trong đoạn K từ 16 đến 32.

Bảng 2: Kết quả tìm mịn K với dữ liệu ngẫu nhiên

Tham số K (dùng 30% của tập dữ liệu đầu vào để xác định các tham số)	RMSE trên tập validation	Thời gian (giờ)
20	0.9190286	10.44
25	0.9183809	10.52
30	0.9182259	10.55

Trong bảng kết quả trên, với K lớn thì lỗi RMSE cho kết quả thấp (không đáng kể) nhưng quá trình tính toán lại mất nhiều thời gian. Từ tập dữ liệu trên chúng tôi chọn K=25 để kiểm thử trên 1/3 tập dữ liệu kiểm thử ban đầu, kết quả thực nghiệm như sau:

Bảng 3: Kết quả tìm K trên tập dữ liệu kiểm tra

Tham số K	RMSE trên tập kiểm tra ban đầu	Thời gian (giờ)
K=25	0.8999636	28.18
K=25	0.8961365	28.32
K=25	0.9068885	28.29

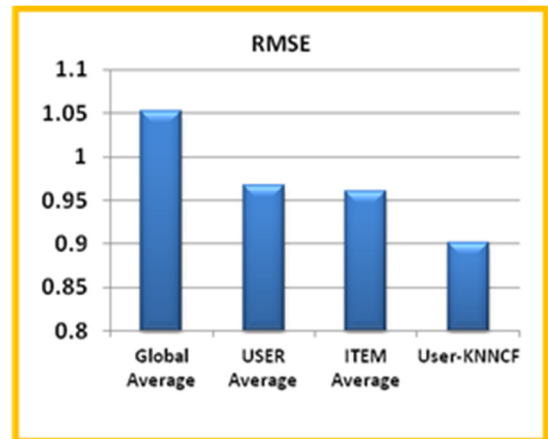
Do tính ngẫu nhiên của dữ liệu nên chúng tôi lặp lại 3 lần việc tính toán kết quả như bảng 3, sau đó chúng tôi lấy trung bình kết quả của 3 lần với **RMSE = 0.900996**.

Bảng 4 thống kê kết quả các phương pháp được sử dụng để so sánh.

Bảng 4: RMSE của các phương pháp gợi ý

Phương pháp	RMSE
Global Average	1.052609
USER Average	0.967062
ITEM Average	0.960684
USER-KNNCF	0.900996

Dưới đây là biểu đồ biểu diễn lỗi RMSE của các phương pháp đã đề cập.



Hình 2: So sánh RMSE của các phương pháp gợi ý

Biểu đồ cho ta thấy lỗi RMSE của phương pháp K láng giềng trong lọc cộng tác tốt hơn dự đoán bằng trung bình toàn cục cũng như trung bình trên sản phẩm và trung bình trên người dùng.

Chúng tôi sẽ sử dụng phương pháp K láng giềng trong lọc cộng tác (User-KNNCF) để xây

dựng hệ thống gợi ý sản phẩm trong bán hàng trực tuyến.

5 TÍCH HỢP GIẢI THUẬT LỌC CỘNG TÁC VÀO HỆ THỐNG BÁN HÀNG TRỰC TUYẾN

5.1 Đặc tả hệ thống

Hệ thống giới thiệu và bán sản phẩm nước hoa trực tuyến cho phép khách hàng bất kỳ có thể tìm kiếm và xem các sản phẩm nước hoa của tất cả các thương hiệu nổi tiếng mà người dùng ưa thích. Hệ thống sẽ hiển thị các sản phẩm được ưa thích nhất và các sản phẩm bán chạy nhất.

Hệ thống có chức năng cho khách hàng chấm điểm, đánh giá sản phẩm và có thể đặt hàng thông qua hệ thống này. Đặc biệt hệ thống sẽ gợi ý các sản phẩm cho khách hàng trong quá trình chọn sản phẩm sử dụng kỹ thuật lọc cộng tác và hiển thị các sản phẩm tương tự với sản phẩm mà khách hàng đang xem sử dụng các thuộc tính tương tự về mùi hương và tính cách của sản phẩm. Đây là giải thuật có độ tin cậy cao vì đã được kiểm nghiệm qua nhiều công trình nghiên cứu nên việc gợi ý sử dụng kỹ thuật lọc cộng tác có thể sẽ tốt hơn các gợi ý thông thường như gợi ý sản phẩm bổ trợ sản phẩm vừa mua, gợi ý sản phẩm tương tự với sản phẩm mà khách hàng ưa thích, gợi ý sản phẩm theo giá chênh lệch với sản phẩm mà khách hàng đang xem...

Khách hàng muốn thực hiện các chức năng trên chỉ khi khách hàng là thành viên của hệ thống.

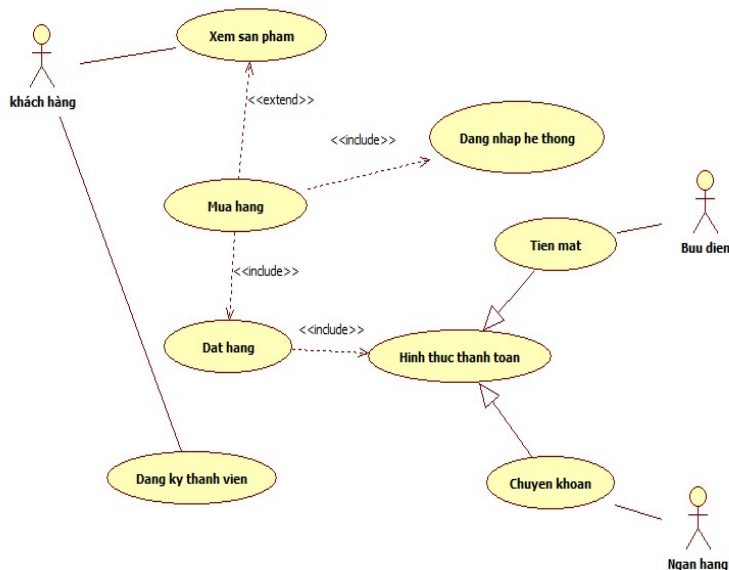
Muốn trở thành thành viên của hệ thống, khách hàng phải đăng ký tài khoản thông qua hệ thống. Sau khi đăng nhập vào hệ thống, nếu là khách hàng mới thì hệ thống sẽ dựa vào thông tin về tính cách và mùi hương ưa thích của khách hàng để tư vấn những sản phẩm theo thông tin vừa thu thập. Tuy nhiên, nếu khách hàng đã có chấm điểm cho sản phẩm, hệ thống sẽ gợi ý các sản phẩm theo giải thuật lọc cộng tác.

Ngoài ra, hệ thống còn cung cấp các công cụ quản trị như: quản trị khách hàng, quản trị thông tin về thương hiệu và sản phẩm, quản trị đơn đặt hàng, thống kê doanh số bán hàng, công cụ cho phép người quản trị có thể xuất thông tin đánh giá của người dùng trên sản phẩm để huấn luyện lại mô hình của giải thuật và chức năng kiểm tra hiệu quả của hệ thống gợi ý sản phẩm.

5.2 Sơ đồ các trường hợp sử dụng (Use case)

Tác nhân “Khách hàng” sử dụng hệ thống để đặt hàng. Các trường hợp sử dụng ở dạng tổng quát này là “xem sản phẩm”, “mua hàng”, “đặt hàng” và “đăng ký thành viên”. Trường hợp sử dụng “xem sản phẩm” có thể được sử dụng bởi khách hàng chỉ khi khách hàng chỉ muốn tìm và xem sản phẩm. Trường hợp sử dụng này cũng có thể được sử dụng như là một phần của trường hợp sử dụng “mua hàng”. Trường hợp sử dụng “đăng ký thành viên” cho phép khách hàng đăng ký trên hệ thống. Ngoài tác nhân “khách hàng” còn có các tác nhân khác được mô tả trong Hình 3.

Hình 3: Use case tổng quát của hệ thống bán hàng trực tuyến

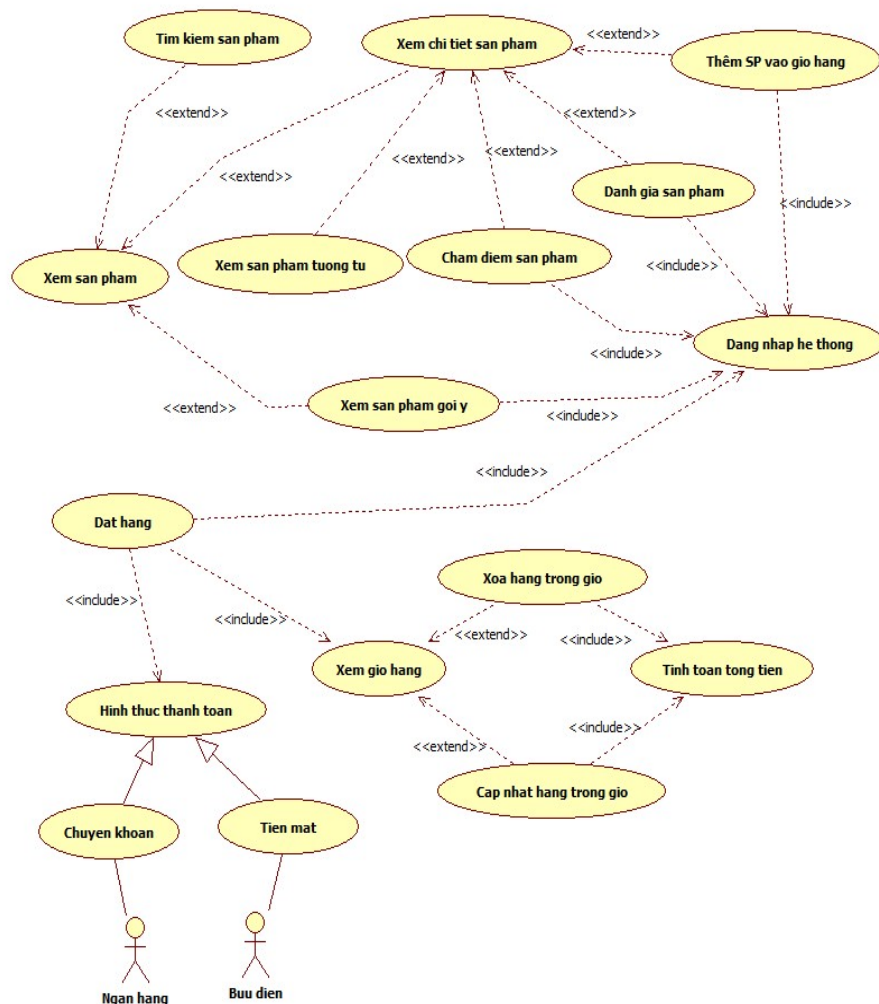


Trường hợp sử dụng “xem sản phẩm” được mở rộng thành một vài trường hợp sử dụng tùy chọn: khách hàng có thể tìm sản phẩm, xem chi tiết sản phẩm, xem những sản phẩm tương tự với sản phẩm, đánh giá và chấm điểm cho sản phẩm, xem những sản phẩm gợi ý cho khách hàng và thêm sản phẩm vào giỏ hàng. Tất cả các trường hợp sử dụng này là trường hợp sử dụng mở rộng (extend) bởi vì chúng cung cấp một số chức năng tùy chọn cho phép khách hàng tìm sản phẩm.

Trường hợp sử dụng “đăng nhập hệ thống” được bao gồm (include) trong trường hợp sử dụng

“đánh giá sản phẩm”, “chấm điểm sản phẩm”, “xem sản phẩm gọi ý” và “thêm sản phẩm vào giỏ hàng” bởi vì các thành phần này yêu cầu khách hàng phải chứng thực tài khoản.

Trường hợp sử dụng “đặt hàng” bao gồm một vài trường hợp sử dụng cần thiết như “xem, cập nhật số lượng hàng và xóa đơn hàng trong giỏ hàng”, “tính toán tổng tiền”. Khách hàng phải chứng thực tài khoản. Điều này có thể được thực hiện thông qua đăng nhập khách hàng (login page). Trường hợp sử dụng “đặt hàng” cũng bao gồm hình thức thanh toán bằng tiền mặt hay chuyển khoản.



Hình 4: Use case xem sản phẩm gợi ý, đặt hàng và thanh toán

Ngoài ra “hệ thống” thực hiện một số các trường hợp sử dụng thống kê và hiển thị danh sách

các đối tượng cần quản lý được mô tả như Hình 5.

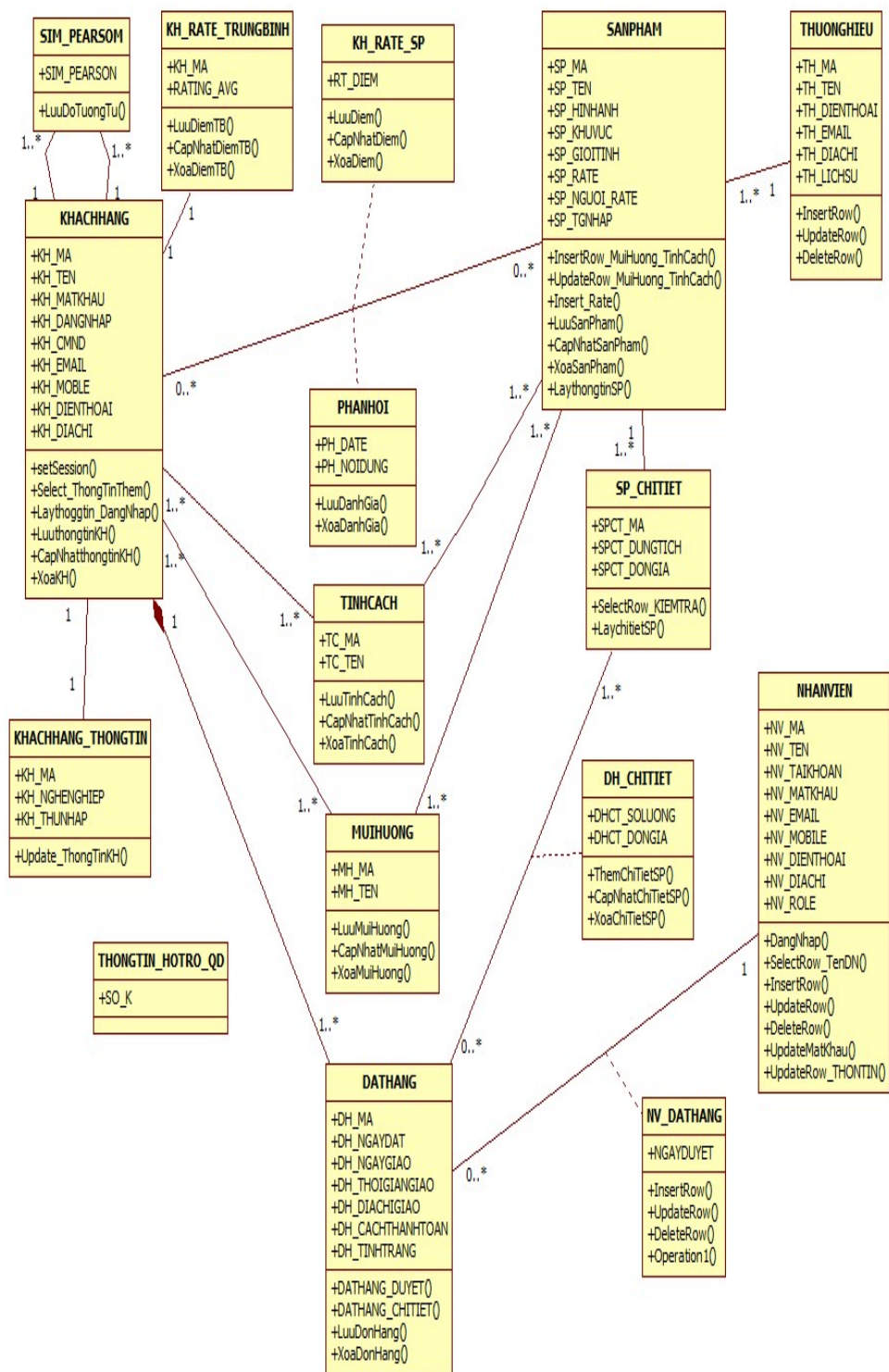


Hình 5: Use case quản trị hệ thống bán hàng trực tuyến

5.3 Sơ đồ lớp của hệ thống

Từ các trường hợp sử dụng được thiết kế ở mục

trên, chúng tôi thiết kế sơ đồ lớp cho hệ thống bán hàng trực tuyến.

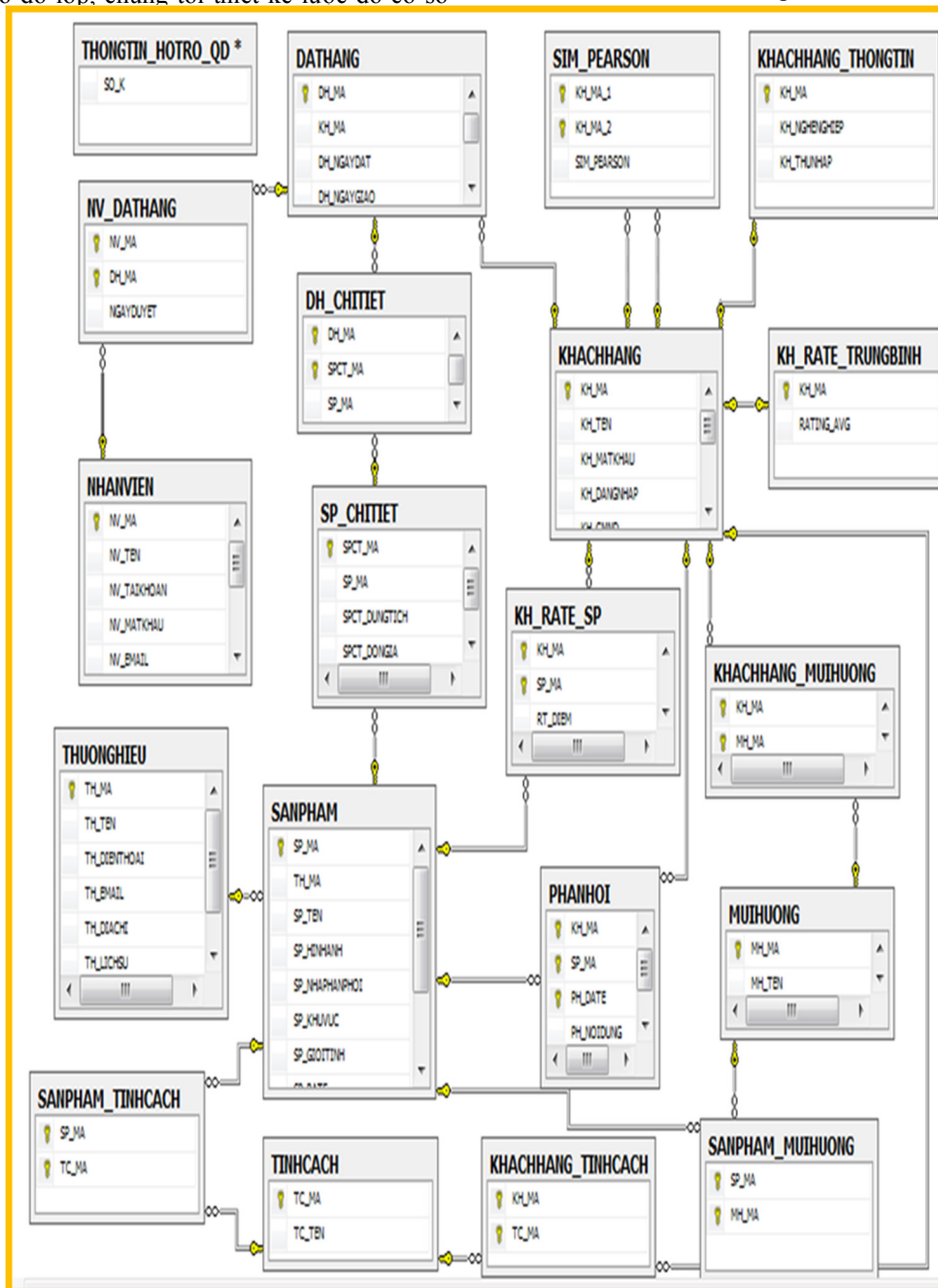


Hình 6: Sơ đồ lớp của hệ thống bán hàng

5.4 Lược đồ cơ sở dữ liệu

Từ sơ đồ lớp, chúng tôi thiết kế lược đồ cơ sở

dữ liệu trên hệ quản trị cơ sở dữ liệu SQL Server để lưu dữ liệu của hệ thống như Hình 7 sau:



Hình 7: Lược đồ cơ sở dữ liệu của hệ thống

5.5 Môi trường phát triển ứng dụng

Hệ thống bán sản phẩm nước hoa trực tuyến được phát triển trên môi trường Visual Studio.NET với .NetFramework 4.0 có sử dụng DevExpress 12.0. Hệ thống được cài đặt bằng ngôn ngữ ASP.NET, C# và SQL (Structure query language)

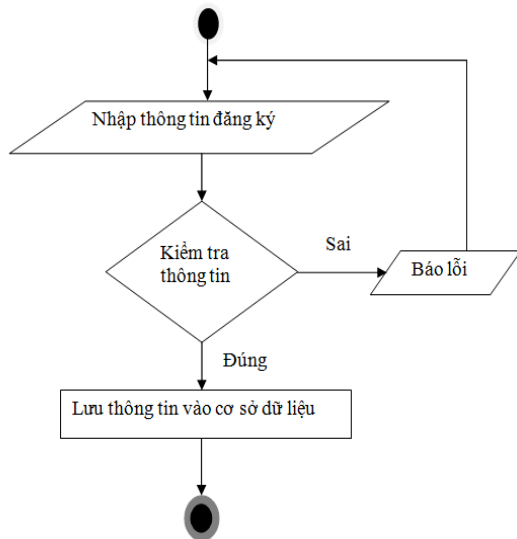
đồng thời thiết kế và lưu trữ dữ liệu trên hệ quản trị cơ sở dữ liệu SQL Server. Tất cả các module như: đăng ký thành viên, tính toán độ tương tự, gợi ý sản phẩm... đều được viết dưới dạng thủ tục lưu trữ (Stored Procedure), bẫy lỗi (Trigger) bằng ngôn ngữ SQL và được lưu trên hệ quản trị cơ sở dữ

liệu. Do đó, có thể dễ dàng tích hợp vào bất kỳ một hệ thống bán hàng trực tuyến khác.

Hệ thống có một ứng dụng cho phép học lại mô hình để xác định tham số K láng giềng sau một thời gian hoạt động.

5.6 Giải thuật xử lý chính trong hệ thống

5.6.1 Giải thuật đăng ký thành viên



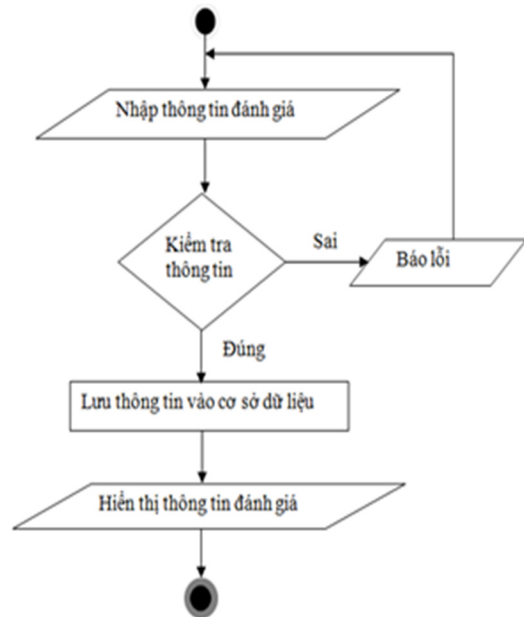
Hình 8 : Giải thuật đăng ký thành viên

Ý nghĩa:

- Mỗi thành viên muốn nhận thông tin gợi ý, chấm điểm, đánh giá sản phẩm hoặc đặt hàng đều phải là thành viên của hệ thống.
- Khách hàng nhập các thông tin được cung cấp trong biểu mẫu.
- Hệ thống sẽ kiểm tra thông tin mà khách hàng nhập. Đặc biệt khách hàng không được nhập trùng với tên đăng nhập đã có. Nếu khách hàng nhập trùng tên đăng nhập, hệ thống sẽ thông báo lỗi.
- Sau khi khách hàng đã đăng ký, khách hàng

có thể đăng nhập vào hệ thống để thực hiện các chức năng được liệt kê ở trên.

5.6.2 Giải thuật khách hàng đánh giá sản phẩm

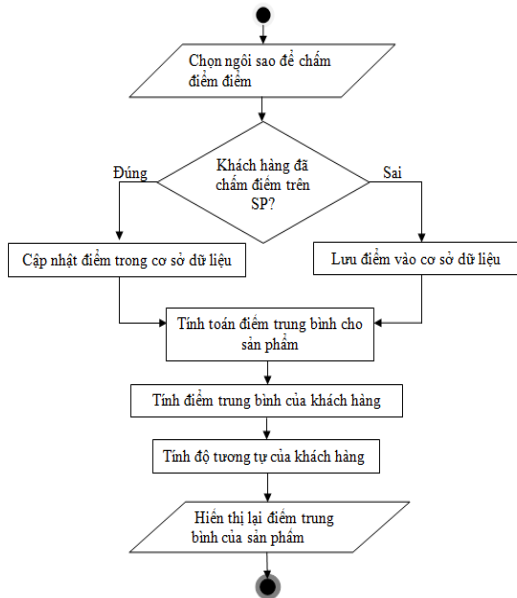


Hình 9: Giải thuật khách hàng đánh giá sản phẩm

Ý nghĩa:

- Để đánh giá cho sản phẩm, khách hàng phải đăng nhập thành công vào hệ thống. Nếu khách hàng chưa đăng nhập thì hệ thống đánh giá không hoạt động.
- Khách hàng nhập thông tin đánh giá sản phẩm.
- Hệ thống sẽ kiểm tra thông tin vừa nhập của khách hàng, số ký tự nhập.
- Nếu thỏa mãn, hệ thống sẽ lưu thông tin đánh giá vào cơ sở dữ liệu. Ngược lại, khách hàng phải nhập lại thông tin.
- Hiển thị đánh giá của khách hàng.

5.6.3 Giải thuật khách hàng chấm điểm sản phẩm

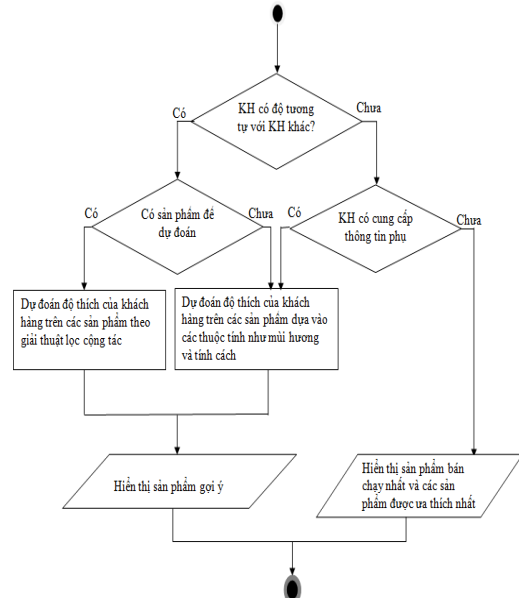


Hình 10: Khách hàng chấm điểm sản phẩm

Ý nghĩa:

- Đề chấm điểm cho sản phẩm, khách hàng phải đăng nhập thành công vào hệ thống. Nếu khách hàng chưa đăng nhập thì hệ thống chấm điểm không hoạt động.
- Khách hàng chọn ngôi sao để chấm điểm cho sản phẩm (điểm tăng dần tính từ trái sang phải với giá trị từ 1 đến 5).
- Hệ thống sẽ kiểm tra xem khách hàng này đã chấm điểm cho sản phẩm này chưa. Nếu chưa, hệ thống sẽ lưu điểm mà khách hàng vừa chấm vào cơ sở dữ liệu. Ngược lại, hệ thống sẽ cập nhật điểm mới mà khách hàng vừa chấm cho sản phẩm.
- Hệ thống tính toán điểm trung bình cho sản phẩm đó.
- Hệ thống tính điểm trung bình của khách hàng đã chấm điểm.
- Hệ thống tính độ tương tự của khách hàng theo Pearson.
- Hiện thị điểm trung bình của sản phẩm.

5.6.4 Giải thuật gợi ý sản phẩm cho khách hàng



Hình 11: Giải thuật xử lý gợi ý sản phẩm

Ý nghĩa:

- Giải thuật này dùng để hiện thị những sản phẩm gợi ý cho khách hàng.
- Khách hàng phải đăng nhập thành công vào hệ thống.
- Hệ thống sẽ kiểm tra xem khách hàng có độ tương tự với bất kỳ khách hàng nào khác không. Nếu có độ tương tự với khách hàng khác, hệ thống lại kiểm tra tiếp có sản phẩm được khách hàng có độ tương tự với khách hàng đang xét đánh giá không. Nếu có, hệ thống sẽ dự đoán độ thích của khách hàng trên các sản phẩm theo giải thuật lọc cộng tác. Ngược lại, dự đoán độ thích của khách hàng trên các sản phẩm dựa vào các thuộc tính như mùi hương và tính cách mà khách hàng đã cung cấp. Nếu không có độ tương tự với khách hàng khác, hệ thống kiểm tra có thông tin về mùi hương mà khách hàng thích và tính cách mà khách hàng cung cấp lúc đăng ký hay không. Nếu có, dự đoán độ thích của khách hàng trên các sản phẩm dựa vào các thuộc tính như mùi hương và tính

cách. Ngược lại, sẽ hiển thị các sản phẩm bán chạy nhất và các sản phẩm được ưa thích nhất cho khách hàng.

– Hiển thị các sản phẩm đã gợi ý cho khách hàng.

5.7 Các giao diện chính của hệ thống

5.7.1 Giao diện chính của hệ thống

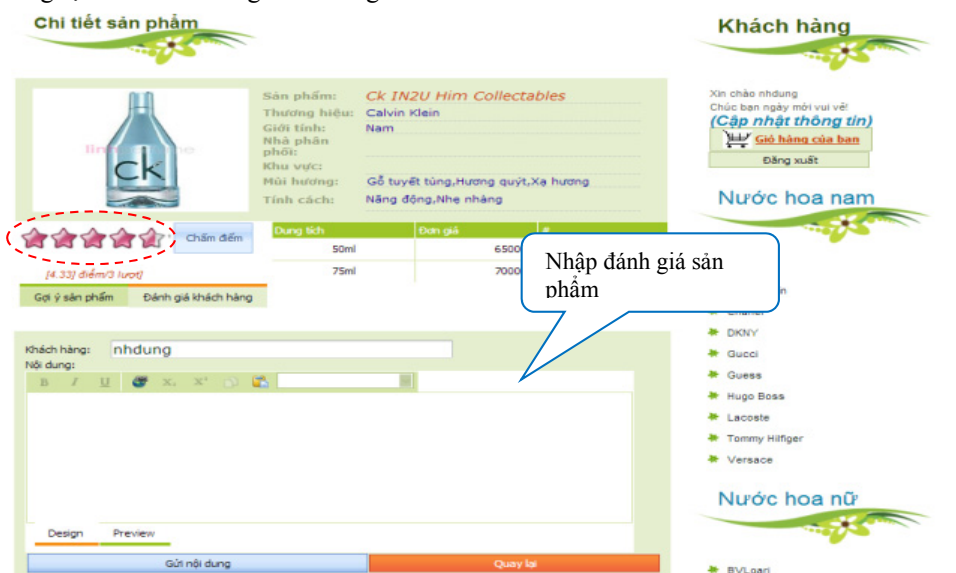
Hình 12: Giao diện chính của hệ thống bán nước hoa trực tuyến



5.7.2 Khách hàng chấm điểm và đánh giá sản phẩm

Khách hàng có thể chấm điểm cho sản phẩm. Dựa vào điểm mà khách hàng đã chấm hệ thống sẽ tính độ tương tự của khách hàng với những khách

hàng khác. Sau đó dự đoán các sản phẩm mà khách hàng có thể thích dựa trên giải thuật lọc cộng tác. Ngoài ra, khách hàng có thể gửi đánh giá (feedback) cho sản phẩm. Đây cũng là hình thức gợi ý cho khách hàng chọn sản phẩm như Hình 13.



Hình 13: Giao diện khách hàng chấm điểm và đánh giá sản phẩm

5.7.3 Gợi ý sản phẩm cho khách hàng

Sau khi khách hàng đã đánh giá sản phẩm, hệ

thống sẽ gợi ý một số sản phẩm mới cho khách hàng thông qua kỹ thuật lọc cộng tác.

Hình 14: Gợi ý sản phẩm mới cho khách hàng



5.7.4 Gợi ý sản phẩm đối với khách hàng mới (khắc phục vấn đề người dùng mới)

Sau khi khách hàng đăng ký thành viên nhưng

chưa đánh giá bất kỳ sản phẩm nào, hệ thống chưa có thông tin gì về quá khứ của khách hàng. Vì vậy, hệ thống sẽ yêu cầu khách hàng nhập thêm một số thông tin để gợi ý sản phẩm được tốt hơn.

Hình 15: Biểu mẫu cung cấp thêm thông tin để gợi ý sản phẩm



Sau khi khách hàng đã nhập thêm thông tin thì hệ thống sẽ sử dụng các thuộc tính như mùi hương

và tính cách mà khách hàng cung cấp đó để tư vấn các sản phẩm cho khách hàng

Hình 16: Gợi ý sản phẩm mới cho khách hàng mới



5.7.5 Đối với sản phẩm mới nhập (khắc phục vấn đề sản phẩm mới)

Hệ thống sẽ hiển thị các sản phẩm mới nhập

tại giao diện chính và có biểu tượng 'New' trên sản phẩm.

Hình 17: Giao diện hiển thị những sản phẩm mới



Ngoài ra, khi hiển thị chi tiết mỗi sản phẩm, website có một không gian để hiển thị các sản phẩm tương tự với sản phẩm mà khách hàng đang

chọn xem dựa vào giá trị tương tự về mùi hương và tính cách trên sản phẩm. Vì vậy, những sản phẩm mới cũng có thể được gợi ý cho người dùng.

Hình 18: Giao diện liệt kê sản phẩm tương tự

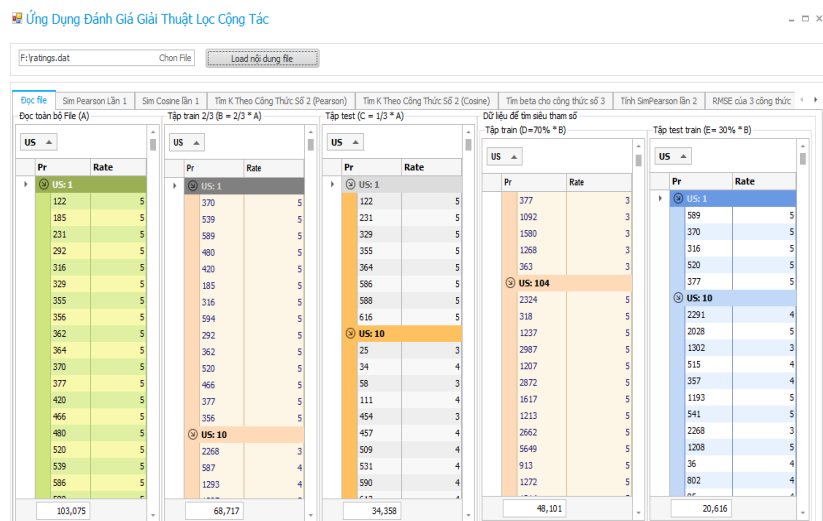


5.7.6 Giao diện huấn luyện lại mô hình

Sau khi có tập tin đánh giá của khách hàng trên

sản phẩm, chúng tôi có một ứng dụng đi kèm để huấn luyện lại mô hình sau một thời hoạt động.

Hình 19: Giao diện của ứng dụng huấn luyện lại mô hình



6 KẾT LUẬN

Trong bài viết này, chúng tôi đã trình bày mô hình láng giềng trong lọc cộng tác – mô hình tư vấn dựa trên độ tương tự trực tiếp giữa hai người dùng hoặc sản phẩm. Trong mô hình này, chúng tôi tính toán độ tương tự giữa hai người dùng, từ đó đưa ra dự đoán đánh giá của người dùng với sản phẩm mới. Đối với những người dùng mới thì chúng tôi sẽ lọc dựa trên một số thuộc tính thu thập từ người dùng để tư vấn những sản phẩm.

Bên cạnh đó, chúng tôi đã cài đặt, đánh giá giải thuật, xây dựng hệ thống bán hàng trực tuyến hoàn chỉnh có tích hợp kỹ thuật lọc cộng tác để gợi ý sản phẩm cho khách hàng. Qua đó, giúp người đọc có thể nắm được một quy trình xây dựng hệ thống gợi ý trong thực tế, đây là công việc vẫn chưa thấy đề cập đến trong các nghiên cứu liên quan.

Hướng nghiên cứu trong tương lai của chúng tôi là kiểm nghiệm lại thuật toán này dựa trên bộ dữ liệu của hệ thống bán hàng trực tuyến (sau một thời gian vận hành) và đưa ra một số cải tiến cho giải thuật.

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. Goldberg, D. Nichols, B. M. Oki, and D. Terry, "Using collaborative filtering to weave an information tapestry" Communications of ACM, vol. 35, no. 12, pp. 61–70, 1992.
3. J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl, "Evaluating collaborative filtering recommender systems," ACM Transactions on Information Systems, vol. 22, no. 1, pp. 5–53, January 2004.
4. J. L. Herlocker, J. A. Konstan, A. Borchers, and J. Riedl, "An algorithmic framework for performing collaborative filtering," in Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval, ser. SIGIR '99. New York, NY, USA: ACM, 1999, pp. 230–237. [Online].

Available:

<http://doi.acm.org/10.1145/312624.312682>.

5. L. Herlocker, J. A. Konstan, et al., An Algorithmic Framework for Performing Collaborative Filtering, Proceedings of the 22nd Annual International ACM SIGIR Conference, ACM Press, 1999, pp. 230–237.
6. N. Thai-Nghe, L. Drumond, A. Krohn-Grimberghe, and L. Schmidt-Thieme, "Recommender system for predicting student performance," in Proceedings of the ACM RecSys 2010 Workshop on Recommender Systems for Technology Enhanced Learning (RecSysTEL 2010), vol. 1. Elsevier's Procedia Computer Science, 2010, pp. 2811 – 2819.
7. 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.
8. P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl, "GroupLens: an open architecture for collaborative filtering of netnews," in Proceedings of the 1994 ACM conference on Computer supported cooperative work, ser. CSCW'94. New York, NY, USA: ACM, 1994, pp. 175–186.
9. P. Resnick, N. Iakovou, M. Sushak, P. Bergstrom, and J. Riedl. "GroupLens: An open architecture for collaborative filtering of netnews". In Proceedings of the 1994 Computer Supported Cooperative Work Conference, 1994.
10. U. Shardanand and P. Maes, "Social information filtering: algorithms for automating 'word of mouth'," in Proceeding of the SIGCHI Conference on Human Factors in Computing Systems, ser. CHI '95. New York, NY, USA: ACM Press/Addison-Wesley Publishing Co., 1995, pp. 210–217. [Online]. Available: <http://dx.doi.org/10.1145/223904.223931>.