

MÔ HÌNH TƯ VẤN LỌC CỘNG TÁC TÍCH HỢP DỰA TRÊN MA TRẬN TƯƠNG ĐỒNG SẢN PHẨM

ITEM-BASED COLLABORATIVE FILTERING RECOMMENDATION MODEL BASED ON SIMILARITY MATRIX OF ITEMS

Phan Quốc Nghĩa¹, Đặng Hoài Phương², Huỳnh Xuân Hiệp³

¹Trường Đại học Trà Vinh; nghiatvnt@tvu.edu.vn

²Trường Đại học Bách khoa, Đại học Đà Nẵng; dhphuong@dut.udn.vn

³Trường Đại học Cần Thơ; hxhiep@ctu.edu.vn

Tóm tắt - Bài báo đề xuất một phương pháp để cải thiện độ chính xác của mô hình tư vấn lọc cộng tác dựa trên sản phẩm (IBCF) bằng cách tích hợp ma trận tương đồng dựa trên các thuộc tính sản phẩm vào quá trình xây dựng mô hình. Trong mô hình tư vấn lọc cộng tác dựa trên sản phẩm truyền thống, kết quả tư vấn được xây dựng chỉ dựa trên ma trận xếp hạng của người dùng cho các sản phẩm. Kết quả tư vấn của mô hình đề xuất được xây dựng dựa trên hai ma trận tương đồng: (1) ma trận tương đồng dựa trên giá trị hạng của người dùng cho các sản phẩm; (2) ma trận tương đồng dựa trên thông tin mô tả các sản phẩm. Thông qua thực nghiệm trên tập dữ liệu MSWeb cho thấy rằng mô hình đề xuất cho kết quả chính xác hơn mô hình tư vấn lọc cộng tác truyền thống.

Từ khóa - độ đo tương đồng; tư vấn lọc cộng tác; ma trận xếp hạng; ma trận tương đồng; ma trận tích hợp.

Abstract - In this paper, we propose a method to improve the accuracy of item-based collaborative filtering recommendation model (IBCF) by integrating the similarity matrix based on the information of item attributes into the process of building recommendation model. In the traditional item-based collaborative filtering recommendation model, the recommendation results are built based only on the rating matrix of users for the items. In this integrated recommendation model, the recommendation results are based on two similarity matrices: the similarity matrix based on the rating value of users for items and the similarity matrix based on the information of item attributes. Through experiments on MSWeb dataset, it shows that the results of our recommendation model are more accurate than the results of traditional item-based collaborative filtering recommendation model.

Key words - similarity measures; item-based collaborative filtering recommendation system; rating matrix; similarity matrix; integrated similarity matrix.

1. Đặt vấn đề

Hệ tư vấn lọc cộng tác dựa trên sản phẩm là một mô hình hệ tư vấn được phát triển dựa trên hệ tư vấn lọc cộng tác [7], [14], [19]. Nó được giới thiệu lần đầu tiên trên tạp chí ACM vào năm 2001 [1]. Không lâu sau đó, hệ thống này được Amazon.com ứng dụng để giới thiệu các sản phẩm của họ đến người dùng [6]. Khác với hệ tư vấn lọc cộng tác dựa trên người dùng (User-based) sử dụng sự tương đồng giữa các người dùng để dự đoán sở thích của người dùng, hệ tư vấn lọc cộng tác dựa trên sản phẩm tìm ra các sản phẩm để giới thiệu đến người dùng bằng cách tìm ra sự tương đồng giữa các sản phẩm từ ma trận xếp hạng (Rating matrix) [15]. Việc xác định sự tương đồng giữa hai sản phẩm được suy luận từ kết quả xếp hạng của người dùng, chứ không dựa trên đặc tính của sản phẩm như trong mô hình tư vấn dựa trên nội dung (content-based) [2]. Nếu hai sản phẩm được xếp hạng cao cùng một người dùng thì hai sản phẩm này được xem là tương đồng và người dùng được hy vọng sẽ có cùng sở thích trên các sản phẩm tương đồng. Hệ tư vấn lọc cộng tác dựa trên sản phẩm là giải pháp hiệu quả cho các hệ thống tư vấn online [18], [13]. Bởi vì trong các hệ thống này, số lượng người dùng thường tăng rất nhanh so với số lượng sản phẩm. Do đó, việc tìm ra kết quả tư vấn dựa trên các người dùng sẽ phức tạp và mất nhiều thời gian hơn so với việc tìm ra kết quả tư vấn dựa trên các sản phẩm. Ngoài ra, mô hình này còn giảm được kích thước của ma trận tương đồng thay vì với mỗi sản phẩm được chứa trong ma trận tương đồng kích thước $n \times n$ (n là số sản phẩm) thì được giảm xuống với kích thước $n \times k$ (k số sản phẩm tương đồng nhất). Điều này đã

tạo nên một bước cải tiến đáng kể về không gian lưu trữ và thời gian chi phí cho các tính toán phức tạp của hệ thống này [1], [15].

Trong bài viết này, chúng tôi đề xuất một phương pháp để cải thiện độ chính xác của mô hình tư vấn lọc cộng tác dựa trên sản phẩm bằng cách tích hợp ma trận tương đồng dựa trên các thuộc tính sản phẩm vào quá trình xây dựng mô hình tư vấn. Trong đó, mô hình tư vấn được xây dựng dựa trên hai ma trận tương đồng: ma trận tương đồng được xây dựng dựa trên dữ liệu xếp hạng của người dùng cho các sản phẩm và ma trận tương đồng được xây dựng dựa trên các thuộc tính sản phẩm. Sau khi xây dựng mô hình, chúng tôi triển khai chạy thực nghiệm mô hình trên tập dữ liệu MSWeb [12], đồng thời so sánh kết quả với mô hình tư vấn lọc cộng tác dựa trên sản phẩm truyền thống.

2. Ma trận tương đồng của các sản phẩm dựa trên các thuộc tính

2.1. Khoảng cách giữa hai sản phẩm

Hiện tại, có nhiều cách tính khoảng cách giữa hai sản phẩm trong không gian vector n chiều. Để tính khoảng cách giữa 2 sản phẩm, chúng tôi áp dụng phương pháp tính khoảng cách Euclidean [17]. Đây là phương pháp tính khoảng cách được sử dụng nhiều trong các bài toán phân lớp dữ liệu do tính đơn giản và hiệu quả của nó. Giả sử ta cần tính khoảng cách giữa hai sản phẩm có dạng vector I chiều như sau: $I_1(x_1, x_2, \dots, x_I)$ và $I_2(y_1, y_2, \dots, y_I)$. Khoảng cách Euclidean giữa hai sản phẩm được xác định bởi công thức sau:

$$d_{Euclide}(I_1, I_2) = \left(\sum_{j=1}^l |x_j - y_j|^2 \right)^{\frac{1}{2}} \quad (1)$$

2.2. Độ đo tương đồng giữa hai sản phẩm

Dựa trên khoảng cách giữa hai sản phẩm, ta có công thức tính giá trị tương đồng giữa hai sản phẩm I_1 và I_2 như sau:

$$S(I_1, I_2) = 1 - d_{Euclide}(I_1, I_2) \quad (2)$$

2.3. Ma trận tương đồng của các sản phẩm

Ma trận tương đồng giữa các sản phẩm là một ma trận đối xứng với cấu trúc: dòng, cột của ma trận là các sản phẩm, các ô của ma trận (giao giữa dòng và cột) là giá trị tương đồng giữa hai sản phẩm trên dòng và cột tương ứng. Cho một tập các sản phẩm xác định $I = \{i_1, i_2, \dots, i_n\}$, mỗi sản phẩm được mô tả bởi một vector các thuộc tính l chiều $A = \{a_1, a_2, \dots, a_l\}$, ma trận tương đồng giữa các sản phẩm được xác định như sau:

$$Matrix_{sim}(I) = \begin{pmatrix} 0 & d_{12} & \dots & d_{1n} \\ d_{21} & 0 & \dots & d_{2n} \\ \vdots & \cdot & \ddots & \vdots \\ d_{n1} & d_{n2} & \dots & 0 \end{pmatrix}$$

Với d_{ij} là giá trị tương đồng giữa sản phẩm i và j được tính bằng công thức (2).

3. Xây dựng ma đồng tích hợp

Ma trận tích hợp là ma trận có cùng kích thước với ma trận tương đồng dựa trên giá trị xếp hạng. Nó được tính dựa trên hai ma trận: ma trận tương đồng dựa trên giá trị xếp hạng và ma trận tương đồng dựa trên thuộc tính sản phẩm theo trọng số quan trọng w cho trước bởi công thức (3):

$$S_I = S_R * w + S_A * (1 - w) \quad (3)$$

Trong đó:

- S_I : ma trận tích hợp;
- S_R : ma trận tương đồng dựa trên giá trị xếp hạng;
- S_A : ma trận tương đồng dựa trên thuộc tính sản phẩm;
- w : trọng số xác định mức độ quan trọng của ma trận tương đồng dựa trên thuộc tính sản phẩm.

4. Mô hình tư vấn lọc cộng tác tích hợp dựa trên sản phẩm

Mô hình tư vấn lọc cộng tác tích hợp dựa trên sản phẩm được định nghĩa như sau:

Gọi $U = \{u_1, u_2, \dots, u_m\}$ là tập m người dùng; $I = \{i_1, i_2, \dots, i_n\}$ là tập n sản phẩm, với mỗi sản phẩm i_k được mô tả bởi một tập gồm l thuộc tính $A = \{a_1, a_2, \dots, a_l\}$; $R = \{r_{j,k}\}$ là ma trận xếp hạng của người dùng cho các sản phẩm (ma trận xếp hạng có thể là ma trận dạng nhị phân hoặc ma trận dạng số thực) với mỗi dòng biểu diễn cho một người dùng u_j ($1 \leq j \leq m$), mỗi cột biểu diễn cho một sản phẩm i_k ($1 \leq k \leq n$), $r_{j,k}$ là giá trị xếp hạng của người dùng u_j cho sản phẩm i_k , N là số sản phẩm có giá trị tương đồng cao nhất và $u_a \in U$ là người dùng cần tư vấn.

Dựa trên dữ liệu đầu vào, mô hình thực thi qua các bước sau:

Bước 1: Xây dựng ma trận tương đồng giữa các sản phẩm dựa trên giá trị xếp hạng (S_R).

Bước 2: Xây dựng mô hình tư vấn dựa trên ma trận tương đồng (S_R).

Bước 3: Xây dựng ma trận tương đồng giữa các sản phẩm dựa trên các thuộc tính của sản phẩm (S_A).

Bước 4: Xây dựng ma trận tích hợp (S_I).

Bước 5: Thay ma trận tích hợp vào mô hình.

Bước 6: Xác định danh mục các sản phẩm tương đồng với các sản phẩm mà người dùng u_a đã mua.

Bước 7: Giới thiệu đến người dùng u_a N sản phẩm có giá trị tương đồng cao nhất.

5. Đánh giá mô hình tư vấn

Đánh giá độ chính xác của mô hình tư vấn là một khâu quan trọng trong qui trình xây dựng hệ tư vấn [8], [10]. Nó giúp cho người thiết kế mô hình lựa chọn mô hình, kiểm tra độ chính xác của mô hình trước khi đưa mô hình vào ứng dụng thực tế. Để đánh giá mô hình tư vấn lọc cộng tác dựa trên sản phẩm, người xây dựng hệ thống cần thực hiện qua 2 bước sau:

5.1. Chuẩn bị dữ liệu cho đánh giá

Tập dữ liệu thực nghiệm được chia làm hai tập: tập dữ liệu huấn luyện (Training set) và tập dữ liệu kiểm tra (Testing set) [15]. Hiện tại, có ba phương pháp để chia tập dữ liệu cho việc đánh giá mô hình tư vấn được sử dụng phổ biến: cắt tập dữ liệu thành hai phần theo tỷ lệ cho trước (Splitting); cắt tập dữ liệu ngẫu nhiên nhiều lần (Bootstrap sampling) và cắt tập dữ liệu thành k phần bằng nhau (K-fold cross-validation) [15]. Trong nghiên cứu này, dữ liệu đánh giá được chia thành 5 phần bằng nhau (5-fold) để phục vụ cho 5 lần đánh giá mô hình. Trong mỗi lần đánh giá, 4 phần dữ liệu được sử dụng để xây dựng mô hình và 1 phần dữ liệu được sử dụng để đánh giá. Kết quả đánh giá của mô hình là giá trị trung bình của 5 lần đánh giá.

5.2. Đánh giá mô hình tư vấn

Có hai phương pháp để đánh giá mô hình tư vấn: đánh giá dựa trên các xếp hạng (Evaluation the ratings) và đánh giá dựa trên các gợi ý (Evaluation the recommendations) [8], [10]. Trong bài viết này, chúng tôi chỉ trình bày phương pháp đánh giá dựa trên các gợi ý của mô hình. Bởi vì phương pháp này có thể áp dụng được cả ma trận xếp hạng nhị phân và ma trận xếp hạng dạng số thực. Đánh giá dựa trên các gợi ý là phương pháp đánh giá độ chính xác của mô hình bằng cách so sánh các gợi ý của mô hình đưa ra với các lựa chọn mua hay không mua của người dùng. Phương pháp này sử dụng ma trận hỗn độn 2x2 (Confusion matrix) để tính giá trị của ba chỉ số: độ chính xác (Precision), độ bao phủ (Recall) và trung bình điều hòa giữa độ chính xác và độ bao phủ (F-measure). Mô hình được đánh giá là tốt khi ba chỉ số trên có giá trị cao [8], [10].

Bảng 1. Ma trận hỗn độn

Lựa chọn của người dùng	Gợi ý của mô hình	
	Gợi ý	Không gợi ý
Mua	TP	FN
Không mua	FP	TN

Với:

TP: Những sản phẩm được mô hình gợi ý đã được mua.

FP: Những sản phẩm được mô hình gợi ý không được mua.

FN: Những sản phẩm không được mô hình gợi ý đã được mua.

TN: Những sản phẩm không được mô hình gợi ý không được mua.

$$\text{Precision} = \frac{\text{Số sản phẩm gợi ý chính xác}}{\text{Tổng số sản phẩm được gợi ý}} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{\text{Số sản phẩm gợi ý chính xác}}{\text{Tổng số sản phẩm được mua}} = \frac{TP}{TP + FN}$$

$$F - \text{measure} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

6. Thực nghiệm

6.1. Xử lý dữ liệu thực nghiệm

Mô hình được thực nghiệm trên tập dữ liệu MSWeb. Đây là tập dữ liệu về người dùng Microsoft truy cập các trang web trong thời gian một tuần trong tháng 2 năm 1998 được lấy mẫu và xử lý từ file log của địa chỉ www.microsoft.com. Tập dữ liệu này bao gồm 38.000 người dùng nặc danh truy cập truy cập trên 285 địa chỉ web gốc và được xử lý và tổ chức thành ma trận nhị phân với 32.710 hàng, 285 cột và 98.653 giá trị xếp hạng. Tuy nhiên, qua khảo sát chúng tôi thấy rằng có khá nhiều người dùng chỉ truy cập một vài trang web và khá nhiều trang web chỉ được truy cập bởi một vài người dùng.

Để tăng độ tin cậy của kết quả gợi ý của mô hình, chúng tôi tiến hành xây dựng tập dữ liệu cho mô hình theo điều kiện chỉ chọn những người dùng truy cập ít nhất 10 địa chỉ web và những trang web được truy cập ít nhất 50 người dùng.

Sau khi thực hiện các thao tác chọn lọc, chúng tôi đã có ma trận nhị phân cho thực nghiệm có kích thước 796 x 135. Ma trận dữ liệu thực nghiệm được chia làm hai tập con: Tập dữ liệu huấn luyện có kích thước 626 x 135 (chiếm 80%), Tập dữ liệu kiểm tra có kích thước 170 x 135 (chiếm 20%).

6.2. Công cụ thực nghiệm

Để triển khai thực nghiệm, chúng tôi sử dụng công cụ ARQAT được triển khai trên ngôn ngữ R. Đây là gói công cụ được nhóm nghiên cứu phát triển từ nền tảng của công cụ ARQAT phát triển trên Java [11].

Công cụ này gồm các chức năng: xử lý dữ liệu; chức năng tính ma trận tương đồng của các sản phẩm dựa trên ma trận xếp hạng; chức năng tính ma trận tương đồng giữa các sản phẩm dựa trên các thuộc tính; các chức năng xây dựng và đánh giá mô hình của hệ tư vấn [16].

6.3. Mô hình IBCF dựa trên ma trận tương đồng xếp hạng

Chúng tôi tiến hành xây dựng mô hình dựa trên ma trận

tương đồng dựa vào giá trị xếp hạng và kiểm tra mô hình trên tập dữ liệu kiểm tra với 156 người dùng. Kết quả tư vấn của mô hình được xuất ra theo định dạng ma trận với cấu trúc 6 x 156 (mỗi cột là một người dùng, mỗi ô là một trang web được chọn để giới thiệu cho người dùng ở cột tương ứng). Hình 1 trình bày kết quả tư vấn cho 6 người dùng đầu tiên, với mỗi người dùng chọn 6 trang web được xếp hạng cao nhất.

[1]	[2]
[1] "Free Downloads"	"isapi"
[2] "Microsoft.com Search"	"Free Downloads"
[3] "Windows Family of Oss"	"Windows Family of Oss"
[4] "Internet Explorer"	"Internet Explorer"
[5] "Internet Site Construction for Developers"	"Internet Site Construction for Developers"
[6] "Windows95 Support"	"Developer Workshop"
[3]	[4]
[1] "isapi"	"Free Downloads"
[2] "Microsoft.com Search"	"Internet Explorer"
[3] "Windows Family of Oss"	"Internet Site Construction for Developers"
[4] "Knowledge Base"	"Developer Workshop"
[5] "Developer Network"	"MS Office Info"
[6] "Windows NT Server"	"Developer Network"
[5]	[6]
[1] "Microsoft.com Search"	"Free Downloads"
[2] "Internet Site Construction for Developers"	"Windows Family of Oss"
[3] "Support Desktop"	"Knowledge Base"
[4] "Knowledge Base"	"Internet Explorer"
[5] "Developer Workshop"	"Internet Site Construction for Developers"
[6] "MS Office Info"	"Developer Workshop"

Hình 1. Kết quả IBCF dựa trên ma trận tương đồng xếp hạng cho 6 người dùng đầu tiên

6.4. Mô hình IBCF dựa trên ma trận tích hợp

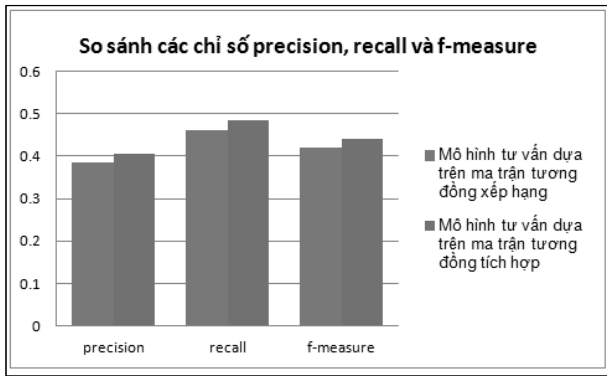
Sau khi ma trận tích hợp được thay vào mô hình, chúng tôi chạy lại mô hình một lần nữa với tập dữ liệu kiểm tra (156 người dùng). Tương tự lần chạy trước, kết quả tư vấn của mô hình được xuất ra theo định dạng ma trận với cấu trúc 6 x 156. Hình 2 trình bày kết quả tư vấn cho 6 người dùng đầu tiên, với mỗi người dùng chọn 6 sản phẩm được xếp hạng cao nhất.

[1]	[2]
[1] "Free Downloads"	"isapi"
[2] "Microsoft.com Search"	"Free Downloads"
[3] "Windows Family of Oss"	"Windows Family of Oss"
[4] "Internet Explorer"	"Internet Explorer"
[5] "MS Office"	"Internet Site Construction for Developers"
[6] "Windows95 Support"	"Windows95 Support"
[3]	[4]
[1] "isapi"	"Free Downloads"
[2] "Microsoft.com Search"	"Internet Explorer"
[3] "Windows Family of Oss"	"Internet Site Construction for Developers"
[4] "Knowledge Base"	"Developer Workshop"
[5] "ActiveX Technology Development"	"MS Office Info"
[6] "Web Site Builder's Gallery"	"Windows NT Workstation"
[5]	[6]
[1] "Microsoft.com Search"	"Free Downloads"
[2] "Support Desktop"	"Windows Family of Oss"
[3] "Knowledge Base"	"Knowledge Base"
[4] "Internet Site Construction for Developers"	"Internet Explorer"
[5] "MS Office Info"	"MS Office Info"
[6] "Developer Workshop"	"Windows95 Support"

Hình 2. Kết quả IBCF dựa trên ma trận tích hợp cho 6 người dùng đầu tiên

6.5. So sánh kết quả hai mô hình

Để so sánh độ chính xác của hai mô hình, chúng tôi tính các chỉ số TP, FP, FN, TN, Precision, Recall và F-measure của hai mô hình dựa trên dữ liệu được xây dựng bằng phương pháp K-fold (k=5). Kết quả tính toán các chỉ số trên của hai mô hình được trình bày trong Hình 3. Trong hình này, chúng tôi thấy rằng các chỉ số Precision, Recall và F-measure trên mô hình IBCF dựa trên ma trận tích hợp có giá trị cao hơn so với các giá trị này trên mô hình IBCF dựa trên ma trận tương đồng xếp hạng. Điều này cho thấy việc tích hợp ma trận tương đồng dựa trên thuộc tính sản phẩm vào mô hình IBCF đã làm tăng độ chính xác của kết quả tư vấn.



Hình 3. So sánh các chỉ số Precision, Recall và F-measure của hai mô hình

7. Kết luận

Theo các khảo sát thực tế, các hệ thống online đang có xu hướng gia tăng số lượng người dùng đáng kể. Điều này đã gây ra một vấn đề khó khăn cho các hệ tư vấn lọc cộng tác dựa trên người dùng do phải mất quá nhiều thời gian cho việc tìm ra các người dùng tương đồng. Chính vì thế, mô hình tư vấn lọc cộng tác dựa trên sản phẩm đang được nhiều nhà thiết kế hệ tư vấn lựa chọn để rút ngắn thời gian tính toán bởi vì số lượng các sản phẩm luôn ít hơn và ổn định hơn số lượng khách hàng. Trong bài báo này, chúng tôi đề xuất một phương pháp để cải thiện độ chính xác của mô hình tư vấn lọc cộng tác dựa trên sản phẩm bằng cách xây dựng ma trận tích hợp để thay thế cho ma trận tương đồng dựa trên giá trị xếp hạng trong mô hình IBCF truyền thống. Kết quả thực nghiệm trên tập dữ liệu MSWeb cho thấy, mô hình mà chúng tôi đề xuất có các chỉ số Precision, Recall và F-measure cao hơn mô hình tư vấn lọc cộng tác dựa trên sản phẩm truyền thống.

TÀI LIỆU THAM KHẢO

- [1] B. M. Sarwar, G. Karypis, J. A. Konstan, and J. Reidl, "Item-based collaborative filtering recommendation algorithms", *ACM WWW '01*, 2001, 10 (285–295).
- [2] Bobadilla, Ortega, Hernando and Gutiérrez, "Recommender systems survey", *Knowledge-Based Systems* 46 (2013), 2013, 23 (109–132).
- [3] Dietmar Jannach, Markus Zanker, Alexander Felfernig and Gerhard Friedrich, *Recommender Systems An Introduction*, Cambridge University Press, ISBN 978-0-521-49336-9, 2011.
- [4] F. Liu and H. J. Lee, "Use of social network information to enhance collaborative filtering performance", *Expert Systems with Applications* 37(7), 2010, 6 (4772–4778).
- [5] Francesco Ricci, Lior Rokach and Bracha Shapira, "Introduction to Recommender Systems Handbook", *Springer-Verlag and Business Media LLC*, 2011, 35(1–35).
- [6] G. Linden, B. Smith, and J. York, "Amazon.com recommendations: Item-to-item collaborative filtering", *IEEE Internet Computing*, vol. 7, no. 1, 2003, 14 (76–80).
- [7] Gabor Takacs et al, "Scalable collaborative filtering approaches for large recommender systems", *Journal of Machine Learning Research*, 2009, 33 (623–656).
- [8] Gunawardana A and Shani G, "A Survey of Accuracy Evaluation Metrics of Recommendation Tasks", *Journal of Machine Learning Research*, v10, 2009, 27 (2935–2962).
- [9] Guy Shani and Asela Gunawardana, "Evaluating Recommendation Systems", *Microsoft Research*, 2009, 42 (1–42).
- [10] Herlocker JL, Konstan JA, Terveen LG and Riedl JT, "Evaluating collaborative filtering recommender systems", *ACM Transactions on Information Systems*, 22(1), ISSN 1046-8188, 2004, 42 (5–53).
- [11] Hiep Xuan Huynh, Fabrice Guillet, Henri Briand, "ARQAT: An Exploratory Analysis Tool For Interestingness Measures", 2005, 10 (334–344).
- [12] Jack S. Breese, David Heckerman and Carl M. Kadie, "Anonymous web data from www.microsoft.com", *Microsoft Research*, Redmond WA. 98052-6399, USA, 1998, <https://kdd.ics.uci.edu/databases/msweb/msweb.html>.
- [13] Martin P. Robillard, Walid Maalej, Robert J. Walker and Thomas Zimmermann, *Recommendation Systems in Software Engineering*, Springer Heidelberg New York Dordrecht London, ISBN 978-3-642-45135-5 (eBook), 2014.
- [14] Michael D. Ekstrand, John T. Riedl and Joseph A. Konstan, "Collaborative Filtering Recommender Systems", *Foundations and Trends in Human-Computer Interaction* Vol. 4, No. 2 (2010), 2010, 92 (81–173).
- [15] Michael Hahsler, "recommenderlab: A Framework for Developing and Testing Recommendation Algorithms", *The Intelligent Data Analysis Lab at SMU*, <http://lyle.smu.edu/IDA/recommenderlab/>, 2011.
- [16] Michael Hahsler, "Lab for Developing and Testing Recommender Algorithms", Copyright (C) Michael Hahsler (PCA and SVD implementation) (C) Saurabh Bathnagar), <http://R-Forge.R-project.org/projects/recommenderlab/>, 2015.
- [17] Michel Marie Deza and Elena Deza, *Encyclopedia of Distances*, Springer, 2014.
- [18] Prem Meville and Vikas Sindhwani, "Recommender Systems. Encyclopedia of Machine Learning", *Springer-Verlag*, 2010, 9 (829–838).
- [19] Xiaoyuan Su and Taghi M. Khoshgoftaar, "A Survey of Collaborative Filtering Techniques", *Advances in Artificial Intelligence archive*, Volume 2009, Article No. 4, 2009, 20 (1–20).

(BBT nhận bài: 01/11/2016, hoàn tất thủ tục phản biện: 11/12/2016)