PHƯƠNG PHÁP LỌC CỘNG TÁC SỬ DỤNG TỐI ƯU BẦY ĐÀN

THE COLLABORATIVE FILTERING METHOD USING PARTICLE SWARM OPTIMIZATION

Nguyễn Thị Hoàng Phương[1], Nguyễn Văn Hiệu[2]

[1]Trường Đại học Phạm Văn Đồng; nthoangphuong90@gmail.com
[2]Trường Đại học Bách khoa - Đại học Đà Nẵng; nvhieuqt@dut.udn.vn

Tóm tắt - Bài báo đề xuất một phương pháp để cải thiện hệ thống khuyến nghị truyền thống - lọc cộng tác dựa trên phân cụm cộng tác kết hợp với trọng số cho các người dùng và sản phẩm. Trong phương pháp tư vấn lọc cộng tác truyền thống, kết quả tư vấn được xây dựng chỉ dựa trên độ tương tự các điểm dữ liệu gần nhau nhất để dự đoán các giá trị khuyết trong ma trận đánh giá. Kết quả tư vấn của phương pháp đề xuất được xây dựng dựa trên độ tương tự các điểm dữ liệu trong cùng cụm kết hợp trọng số thể hiện mức độ quan trọng đối với từng điểm dữ liệu để dự đoán các giá trị khuyết trong ma trận đánh giá. Thông qua thực nghiệm trên tập dữ liệu MovieLens 100k cho thấy rằng phương pháp đề xuất cho kết quả dự đoán tốt hơn so phương pháp tư vấn truyền thống.

Từ khóa - Lý thuyết bầy đàn; tư vấn lọc cộng tác; ma trận xếp hạng; ma trận tương đồng; ma trận tương đồng kết hợp

Abstract - In this paper, we propose a method in order to improve the traditional recommender system – a feature weighting method for both item-based collaborative filtering and user-based collaborative filtering recommender system. In traditional collaborative filtering, the recommendation results are just based on the similar nearest neighbor measure to predict unknown values in evaluation matrix. In this proposed method, the recommendation result is built by the combination of similarity features in the same cluster and weighting which show the extent of importance of each feature to predict the unknow values in evaluaion matrix. Through experiments on MovieLens 100k dataset, it shows that the results of our recommender method is better than those by the traditional method.

Key words - Particle Swarm Optimization; collaborative filtering recommender system; rating matrix; similarity matrix; integrated similarity matrix

1. Đặt vấn đề

Hệ khuyến nghị các sản phẩm dựa vào sự tương đồng giữa phẩm hoặc người dùng được phát triển bởi [1], 5[7]. Sản phẩm được gợi ý cho người dùng dựa trên những người dùng có cùng hành vi hay những sản phẩm tương tự. Tuy nhiên, các nghiên cứu trước đây chưa đề cập đến mức độ quan trọng giữa các người dùng hay mức độ quan trọng giữa các sản phẩm, dẫn đến hệ thống khuyến nghị giả định sẽ dự đoán không hoàn toàn chính xác. Bài toán đặt ra vấn đề là làm cách nào để gợi ý được sản phẩm thích hợp (sản phẩm chưa được người dùng đánh giá xếp hạng) đến với người dùng, dựa trên các xếp hạng mà người dùng đã đánh giá các sản phẩm trước đó. Bài toán được chia làm hai hướng giải quyết: phân cụm người dùng, phân cụm sản phẩm sử dụng lý thuyết bầy đàn và sự kết hợp giữa chúng. Sau khi đề xuất phương pháp mới, nhóm tác giả triển khai chạy thực nghiệm phương pháp trên tập dữ liệu MovieLens 100k (https://goo.gl/BzHgtq) được công bố năm 1998 bởi tổ chức GroupLens (https://grouplens.org/), đồng thời so sánh kết quả với phương pháp tư vấn truyền thống.

2. Nghiên cứu tổng quan

2.1. Phương pháp lọc cộng tác sử dụng mô hình láng giềng

Cho tập hợp U = {u_1, u_2,. . . , u_N} biểu diễn cho tập người dùng và tập hợp các sản phẩm I = {i_1, i_2,. . . , i_M}. Xếp hạng của người dùng cho các sản phẩm được lưu trữ trong ma trận xếp hạng R ở dạng tường minh. Tuân thủ đúng quy trình, có ba bước cần thiết để xây dựng hệ thống gợi ý: thu thập dữ liệu để tạo hồ sơ người dùng; thiết lập tập láng giềng; dự đoán và khuyến nghị. Sau khi dự đoán các đánh giá, hệ thống sẽ xác định top-N sản phẩm tiêu biểu với giá trị dự đoán cao nhất và gợi ý cho người dùng. Giá trị dự đoán $\hat{r}_{u,i}$cho sản phẩm i của người dùng u được tính:

$$\hat{r}_{u,i} = \bar{r}_u + \frac{\sum_{u' \in K_u} sim(u,u').(r_{u',i} - \bar{r}_{u'})}{\sum_{u' \in K_u} |sim(u,u')|}$$

Mã giả minh hoạ cho phương pháp làng giềng dựa vào người dùng và dựa vào sản phẩm được biểu diễn như sau:

1: **procedure** USERKNN-CF ($\bar{\boldsymbol{r}}_{\boldsymbol{u}}$,*r*,$D^{train}$)	1: **procedure** ITEMKNN-CF ($\bar{\boldsymbol{r}}_{\boldsymbol{u}}$,*r*,$D^{train}$)
2: **For** *u*=1 to *N* **do**	2: **for** *i*=1 to *M* **do**
3: Tính $Sim_{uu'}$	3: Tính $Sim_{ii'}$
4: end for	4: **end for**
5: Sort $Sim_{uu'}$	5: Sort $Sim_{ii'}$
6: for *k*=1 to *K* **do**	6: **for** *i*=1 to *K* **do**
7: *Ku* ← *k*	7: *Ki* ← *k*
8: end for	8: **end for**
9: for i = 1 to M **do**	9: **for** *u* = 1 to *N* **do**
10: Tính $\hat{\boldsymbol{r}}_{\boldsymbol{u,i}}$	10: Tính $\hat{\boldsymbol{r}}_{\boldsymbol{u,i}}$
11: end for	11: **end for**
12: end procedure	12: **end procedure**

Hình 1. *Giải thuật láng giềng trên người dùng và sản phẩm*

2.2. Phương pháp lọc cộng tác sử dụng phân cụm Spectral

Bản chất của phương pháp này là ứng dụng kỹ thuật phân cụm Spectral vào trong lọc cộng tác dựa trên cảngười dùng và sản phẩm, trong đó các xếp hạng chưa biết được suy ra từ các xếp hạng tường minh của nhóm người dùng hoặc sản phẩm tương tự.

Trong kỹ thuật phân cụm Spectral:

Đầu vào: một tập hợp n điểm (có thể là n người hoặc n sản phẩm) kí hiệu X= $\{x_1, x_2, x_3, \dots, x_n\}$ và k cụm.

Đầu ra: k cụm $C_1, C_2, \dots, C_k$.

Bước 1: Phân cụm người dùng và sản phẩm theo kỹ thuật Spectral

Tính ma trận mối liên hệ S với các phần tử theo định nghĩa như sau:

$$s_{ij} = exp\left(-\frac{\|\overrightarrow{x_i} - \overrightarrow{x_j}\|^2}{2\times\sigma^2}\right) \text{ nếu i≠j và } s_{ii} = 1$$

với i, j = 1, …, n

Trong đó, s_{ij} là độ tương đồng của đối tượng i, j; $\overrightarrow{x_i}, \overrightarrow{x_j}$ là các vector tương ứng với hàng thứ i, j trong ma trận R, đại diện cho đối tượng i, j; σ là tham số điều chỉnh độ lớn của tập láng giềng. Nếu σ nhỏ sẽ thu được một cấu hình địa phương tốt hơn đối với tập láng giềng. Ở đây mẫu số được tính tương ứng theo công thức sau:

$$2 \times \sigma^2 = n$$

Tính ma trận đường chéo chính D, trong đó các phần tử d_i được tính:

$$d_i = \sum_{j=1}^{n} s_{ij}$$

Tính ma trận chuẩn hóa Laplacian L tương ứng:

$$L = D^{-\frac{1}{2}}(D - S)D^{-\frac{1}{2}}$$

Tính k giá trị vector đầu tiên $v_1, v_2, \dots, v_k$ của bài toán tổng quát:

$$Lv = \lambda Dv$$

Xây dựng ma trận với $V \in R^{nk}$ chứa các vector $v_1, v_2, \dots, v_k$ tương ứng với các cột của ma trận.

Gọi $y_i \in R^k$ là các vector hàng thứ i của V. Dùng thuật toán k-means để phân cụm các điểm $(y_i)_{i=1,\dots,n}$ trong R^k thành các cụm $C_1, C_2, \dots, C_k$.

Gán các điểm ban đầu $(x_i)_{i=1,\dots,n}$ vào cụm C_j nếu nó tương ứng với vector $(y_i)_{i=1,\dots,n}$ đã được gán trước đó.

Bước 2: Tính giá trị chưa biết $\mathbf{r_{ui}^U}$ và $\mathbf{r_{ui}^I}$ dựa trên người dùng và dựa trên sản phẩmbằng công thức:

$$r_{ui}^U = k_c \sum_l sim^U(u,l) \times r_{li}$$

$$r_{ui}^I = k_g \sum_{jj} sim^I(i,j) \times r_{ul}$$

Trong đó: k_c và k_g các giá trị chuẩn hóa, được tính: $k_c = \frac{1}{\sum_l |sim^U(u,l)|}$ và $k_g = \frac{1}{\sum_j |sim^I(j,k)|}$; $sim^U(u,l)$ và $sim^I(i,j)$ là độ tương tự giữa 2 người dùng và giữa 2 sản phẩm tương ứng.

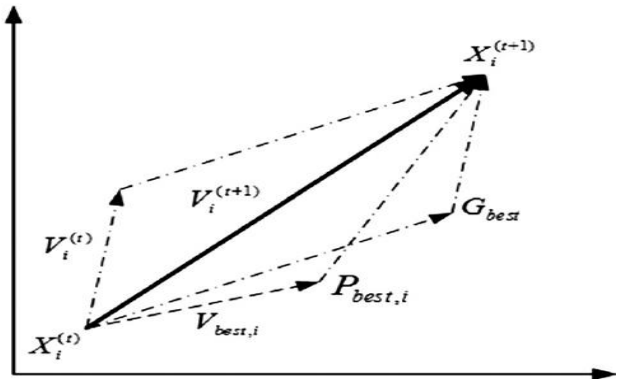
Bước 3: Dự đoán các giá trị đánh giá chưa biết $\mathbf{r_{ui}}$ của người dùng ứng với sản phẩm i được tổng hợp thông qua r_{ui}^U và r_{ui}^I bằng công thức:

$$r_{ui} = \propto r_{ui}^I + (1 - \propto) r_{ui}^U, \propto \in [0,1].$$

Phương pháp lọc cộng tác sử dụng kỹ thuật phân cụm Spectral đã thành công trong việc giải quyết được vấn đề về dữ liệu thưa, người dùng mới. Tuy nhiên, chưa giải quyết được mức độ quan trọng của các sản phẩm và của các người dùng.

2.3. Lý thuyết tối ưu bầy đàn

Phương pháp tối ưu hóa bày đàn là một dạng của các thuật toán tiến hóa quần thể, được giới thiệu lần đầu vào năm 1995 bởi James Kennedy và Russell C. Eberhart. Phương pháp được khởi tạo bằng một nhóm cá thể ngẫu nhiên và sau đó tìm nghiệm tối ưu bằng cách cập nhật các thế hệ. Trong mỗi thế hệ, mỗi cá thể được cập nhật theo hai vị trí tốt nhất. Giá trị thứ nhất là vị trí tốt nhất từng đạt được tới thời điểm hiện tại, gọi là tối ưu cục bộ P_{best}. Giá trị thứ hai là vị trí tốt nhất trong tất cả quần thể từ đầu cho tới thời điểm hiện tại, gọi là tối ưu toàn cục G_{best}. Nói cách khác, mỗi cá thể trong quần thể cập nhật vị trí của nó theo vị trí tốt nhất của nó và của cả quần thể tính tới thời điểm hiện tại.



Hình 2. *Sơ đồ tìm kiếm bằng lý thuyết bầy đàn*

Trong đó:

X_i^k: vị trí cá thể thứ i trong thế hệ thứ k;

V_i^k: vận tốc cá thể thứ i trong thế hệ thứ k;

X_i^{k+1}: vị trí cá thể thứ i trong thế hệ thứ k + 1;

V_i^{k+1}: vận tốc cá thể thứ i trong thế hệ thứ k + 1;

P_{best}: vị trí tốt nhất của cá thể thứ i;

G_{best} : vị trí tốt nhất trong quần thể thứ i.

Vận tốc và vị trí của cá thể trong quần thể được cập nhật theo công thức:

$$V_i^{k+1} = \omega * V_i^k + c_1 * r_1 * \left(P_{best}^k - V_i^k\right) + c_2 * r_2 * (G_{best}^k - V_i^k)$$

$$X_i^{k+1} = X_i^k + V_i^{k+1}$$

Trong đó:

ω : là hệ số quán tính, giảm tuyến tính từ 1 đến 0 tùy thuộc vào số lần lập xác định trước.

c_1, c_2 : Các hệ số gia tốc, nhận giá trị từ [1,173; 2,5]

r_1, r_2: các giá trị ngẫu số nhận giá trị [0, 1]

3. Phương pháp đề xuất sử dụng lý thuyết bầy đàn

Để khắc phục vấn đề các người dùng và các sản phẩm có mức quan trọng ngang bằng nhau, chúng ta sử dụng lý thuyết tối ưu bầy đàn để ước tính trọng số cho người dùng và sản phẩm. Trọng số này được dùng để cải tiến công thức phân cụm và mô hình dự đoán.

Cho $\{w_1^I, w_2^I, \dots, w_m^I;\ w_1^U, w_2^U, \dots, w_n^U\}$ là tập phương án, thể hiện trọng số của các cá thể trong lý thuyết bầy đàn với $w_j^I, j = 1..m$ và $w_i^U, i = 1..n$, nhận giá trị trong đoạn [0,1], làm đại diện cho trọng số của các sản phẩm và người dùng. Dữ liệu ban đầu là một tập hợp các cá thể được khởi tạo ngẫu nhiên. Cấu hình tham số của lý thuyết bầy đàn được cung cấp theo điều kiện dữ liệu thực tế.

Trong mô hình dự đoán các trọng số *w* chính là các cá

thể của thuật toán tối ưu bầy đàn. Các trọng số được sử dụng để cập nhật độ tương tự theo phương pháp cosine:

$$wsim(\overrightarrow{x_a}, \overrightarrow{x_b}) = \frac{\sum_i w_i x_{a,i} w_i x_{b,i}}{\sqrt{\sum_i w_i{}^2 x^2{}_{a,i}} \sqrt{\sum_i w_i{}^2 x^2{}_{b,i}}}$$

3.1. Cấu hình cho bầy đàn

Cấu hình cá thể

$$\boldsymbol{w} = [[w_1^U, w_2^U, \dots, w_n^U], [w_1^i, w_2^i, \dots, w_m^i]]$$

Với n số người dùng, m số sản phẩm, miền giá trị trọng số $D = [0, 1]$.

Kích thước quần thể(số cá thể): 20.

Số bước lặp tối đa (số thế hệ quần thể): 100.

Hệ số gia tốc cục bộ (c_1): 2.

Hệ số gia tốc toàn cục (c_2): 2.

Hệ số quán tính (w): 0,9.

3.2. Xây dựng hàm thích nghi

Xây dựng hàm thích nghi bằng cách đánh giá trung bình cộng sai số tuyệt đối giữ kết quả dự đoán và kết quả thực trong tập dữ liệu test.

Bước 1: Phân cụm theo người dùngvà sản phẩm bằng kỹ thuật spectral và kết hợp trọng số.

Xây dựng ma trận tương đồng giữa các người dùng (các sản phẩm) trên cơ sở ma trận đánh giá và trọng số tương quan sản phẩm (người dùng) theo công thức:

$$\boldsymbol{S}_{ij} = Exp\left(-\frac{\left\|\overrightarrow{\boldsymbol{wx}_i} - \overrightarrow{\boldsymbol{wx}_j}\right\|^2}{2 \times \sigma^2}\right)$$

Sau khi tính độ tương đồng, tiến hành phân cụm spectral như đã đề cập ở bước 1, Mục 2.2, nhận được các cụm tương ứng với người dùng và sản phẩm.

Bước 2: Thực hiện đánh giá các giá trị chưa biết trong mỗi cụm.

$$x_{ij}^U = \frac{\sum_l \boldsymbol{\omega}^I \times \boldsymbol{sim}^U(i,l) \times x_{lj}}{\sum_l \boldsymbol{\omega}^I \times \boldsymbol{sim}^U(i,l)}$$

Với $\boldsymbol{sim}^U(i,l)$ là độ tương đồng giữa user i và user l trong cùng một cụm.

$$x_{ij}^I = \frac{\sum_l \boldsymbol{\omega}^U \times \boldsymbol{sim}^I(i,l) \times x_{lj}}{\sum_l \boldsymbol{\omega}^U \times \boldsymbol{sim}^I(i,l)}$$

Với $\boldsymbol{sim}^U(i,l)$ là độ tương đồng giữa user i và user l trong cùng một cụm.

Ta có giá trị đánh giá kết hợp user-item base:

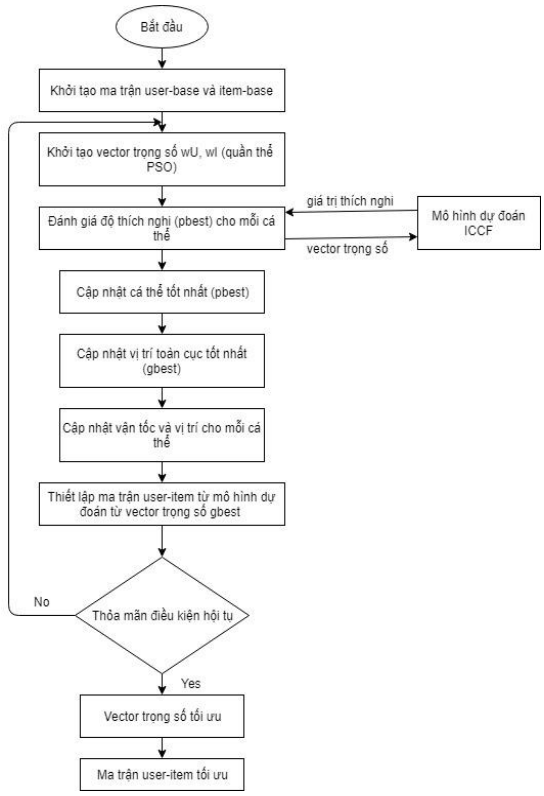
$$x'_{ij} = \alpha \times x_{ij}^U + (1 - \alpha) \times x_{ij}^I$$

Bước 3: Ước lượng giá trị mục tiêu cho mỗi cá thể trong tối ưu bầy đàn

$$MAE = \frac{\sum_{n\_test} \left|x'_{ij} - x_{ij}\right|}{n\_test}$$

Trong đó n_test là tổng số rating cần đánh giá (hay kích thước của tập dữ liệu test), x'_{ij} giá trị dự đoán trong mô hình đánh giá, x_{ij} giá trị xếp hạng trong tập dữ liệu test.

3.3. Tối ưu vector trọng số



Hình 3. *Lưu đồ tối ưu vector trọng số*

4. Đánh giá phương pháp tư vấn

Đánh giá độ chính xác của phương pháp tư vấn là một khâu quan trọng trong quy trình xây dựng hệ tư vấn [[3]. Nó giúp cho người thiết kế lựa chọn phương pháp, kiểm tra độ chính xác trước khi đưa phương pháp vào ứng dụng thực tế. Trong nghiên cứu này, nhóm tác giả sử dụng độ đo sai số tuyệt đối trung bình (Mean Absolute Error - MAE) giữa kết quả đánh giá từ hệ thống và kết quả đánh thực từ người dùng trong tập dữ liệu kiểm thử.

4.1. Chuẩn bị dữ liệu cho đánh giá

Tập dữ liệu thực nghiệm được chia làm hai tập: tập dữ liệu huấn luyện và tập dữ liệu kiểm tra [6]. Hiện tại, có ba phương pháp để chia tập dữ liệu cho việc đánh giá phương pháp 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) [6]. Trong nghiên cứu này, dữ liệu đánh giá được lưu trong tập*u.data* và được chia thành hai tập con theo hai cách ngẫu nhiên khác nhau, một cho huấn luyện (*ua.base, ub.base*), một cho kiểm thử (*ua.test, ub.test*).

4.2. Đánh giá phương pháp tư vấn

Có hai phương pháp để đánh giá: đánh giá dựa trên các xếp hạng (Evaluation the ratings) và đánh giá dựa trên các gợi ý (Evaluation the recommendations) [3] 3, [4]. Trong bài viết này, nhóm tác giả 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.

5. Thực nghiệm

5.1. Xử lý dữ liệu thực nghiệm

Mô hình được thực nghiệm trên hai tập dữ liệu con ua.base và ua.test và hoàn toàn có thể thực nghiệm lại với cách chia dữ liệu còn lại ub.base và ub.test (hoặc thực nghiệm trên các tập dữ liệu lớn hơn MovieLens 10M, MovieLens 20M).

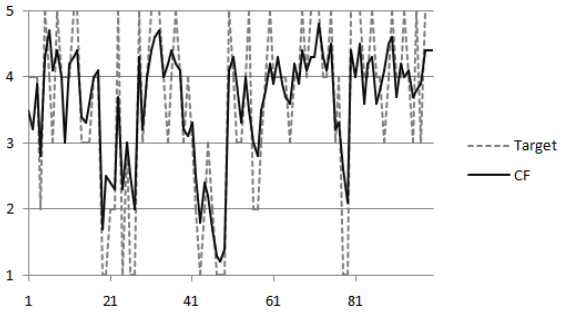
Ma trận dữ liệu thực nghiệm được chia làm hai tập con: Tập dữ liệu huấn luyện có kích thước 90570 giá trị xếp hạng (chiếm 90%), Tập dữ liệu kiểm tra có kích thước 9430 giá trị xếp hạng (chiếm 10%).

5.2. Công cụ thực nghiệm

Để triển khai thực nghiệm, nhóm tác giả sử dụng các thư viện hỗ trợ tính toán như scipy, sklearn, matplotlib, numpy được triển khai trên ngôn ngữ Python 3.

5.3. Phương pháp lọc cộng tác sử dụng mô hình láng giềng

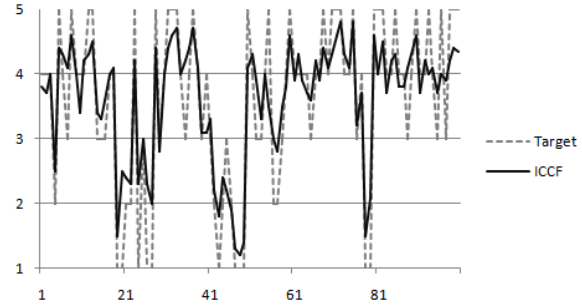
Tiến hành xây dựng phương pháp dựa trên độ đo cosincủa ma trận đánh giávà kiểm tra phương pháp trên tập dữ liệu kiểm tra với 9430 đánh giá. Kết quả tư vấn của phương pháp được xuất ra theo định dạng ma trận trong mỗi dòng gồm {id_user, id_item, rating}. Kết quả tư vấn 100 đánh giá đầu tiên từ hệ thống so với kết quả trong tập kiểm thử.



Hình 4. *Kết quả dự đoán mô hình CF*

5.4. Phương pháp lọc cộng tác sử dụng phân cụm Spectral

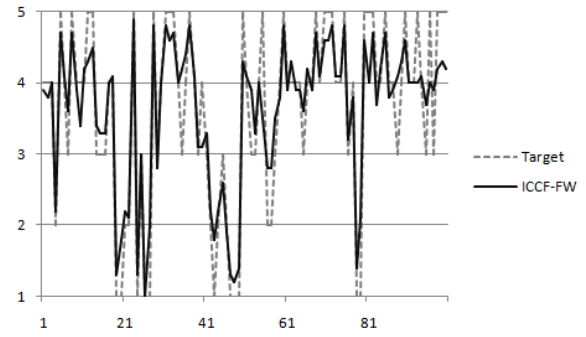
Nhóm tác giả tiến hành xây dựng phương pháp dựa trên phân cụm người dùng bằng kỹ thuật spectral, sau đó tiến hành dự đoán dựa trên độ đo cosin trong mỗi cụm. Kết quả tư vấn cho 100 đánh giá đầu tiên từ hệ thống so với kết quả trong tập dữ liệu test.



Hình 5. *Kết quả dự đoán mô hình ICCF*

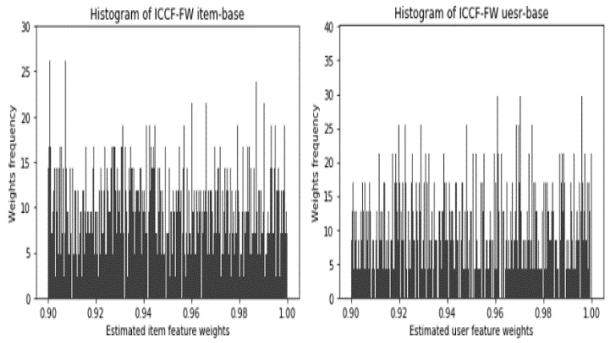
5.5. Phương pháp lọc cộng tác sử dụng phân cụm kết hợp với lý thuyết bầy đàn

Sử dụng lý thuyết bầy đàn để xác định trọng số cho người dùng và các sản phẩm. Các trọng số này được sử dụng để xây dựng ma trận tương quan bằng độ đo cosin. Tương tự với dữ liệu ở Mục 5.3 và 5.4, kết quả tư vấn của mô hình được xuất ra theo định dạng ma trận mỗi dòng bao gồm {id_user, id_item, rating}. Kết quả tư vấn cho 100 đánh giá đầu tiêntrong mô hình so với trong tập dữ liệu test.



Hình 6. *Kết quả dự đoán mô hình ICCF-FW*

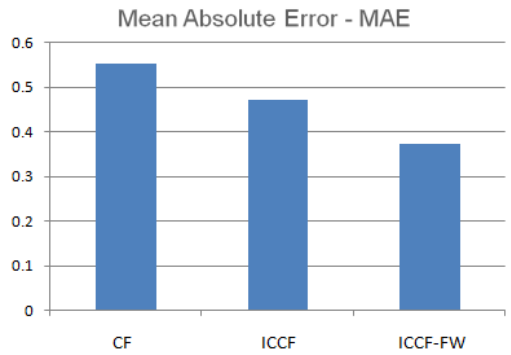
Kết quả vector trọng số người dùng và sản phẩm được thể hiện dưới dạng biểu đồ histogram vector trọng số.



Hình 7. *Histogram trọng số item - user*

5.6. So sánh kết quả ba phương pháp

Để so sánh độ chính xác của 3 phương pháp, nhóm tác giả tính độ sai lệch tuyệt đối giữa các mô hình CF, ICCF và ICCF-FW. Kết quả cho thấy, mô hình ICCF-FW có độ sai lệch dự đoán thất nhất so với hai mô hình còn lại, cho thấy sự hiệu quả của cải tiến này.



Hình 8. *Biểu đồ đánh giá sai số của các mô hình dự đoán*

6. Kết luận

Theo các khảo sát thực tế, các hệ thống online đang có xu hướng quảng cáo các sản phẩm đến người dùng một cách chính xác nhất, giúp làm tăng doanh thu bán hàng, vì vậy hệ thống khuyến nghị - recommender system là một giải pháp giúp cho việc quảng cáo sản phẩm đến người dùng một cách chính xác và hiệu quả nhất so với các phương pháp quảng cáo truyền thống.

Trong bài báo này, đề xuất một phương pháp để cải thiện độ chính xác của phương pháp tư vấn lọc cộng bằng cách giả thuyết rằng mỗi người dùng hoặc sản phẩm có độ ưu tiên khác nhau và được phân thành các cụm theo phương pháp Spectral. Tiến hành chọn lọc bộ trọng số cho kết quả dự đoán tốt nhất dựa theo thuật toán tối ưu hóa bày đàn.

Kết quả thực nghiệm trên tập dữ liệu MovieLens 100k cho thấy phương pháp lập lọc phân cụm cộng tác kết hợp trọng số (ICCF-FW) mà nhóm tác giả đề xuất có độ chính xáccao hơn phương pháp tư vấn lọc cộng tác truyền thống.

TÀI LIỆU THAM KHẢO

[1] F. Isinkaye, Y. Folajimi, and B. Ojokoh, "Recommendation systems: Principles, methods and evaluation", (in en), Egyptian Informatics Journal, vol. 16, no. 3, pp. 261-273, 2015.

[2] Gabor Takacs et al, "Scalable collaborative filtering approaches for large recommender systems". *Journal of Machine Learning Research*, 2009, 33 (623-656).

[3] Gunawardana A and Shani G, "A Survey of Accuracy Evaluation Metrics of Recommendation Tasks", *Journal of Machine Learning Research*, v10, 2009, 27 (2935–2962).

[4] 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).

[5] 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).

[6] 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.

[7] 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: 20/3/2019, hoàn tất thủ tục phản biện: 20/4/2019)