

Intelligent Learning System based on Personalized Recommendation Technology

Phạm Đức Thế

Khoa Học Và Kỹ Thuật Thông Tin

Đại Học Công Nghệ Thông Tin

TPHCM, Việt Nam

195222536@gm.uit.edu.vn

Trần Thành Luân

Khoa Học Và Kỹ Thuật Thông Tin

Đại Học Công Nghệ Thông Tin

TPHCM, Việt Nam

19521810@gm.uit.edu.vn

Tóm tắt nội dung—Với sự phát triển không ngừng của networks, web-based e-learning đang thay đổi cách mọi người tiếp thu kiến thức. Ngày càng có nhiều người học mong muốn tiếp thu thêm kiến thức thông qua các phương tiện thông minh và được cá nhân hóa. Dựa trên thuật toán khuyến nghị nội dung và khuyến nghị lọc cộng tác, bài báo này đề xuất một thuật toán khuyến nghị kết hợp có thể cải thiện hiệu quả của thuật toán khuyến nghị truyền thống. Nghiên cứu được trình bày giới thiệu toàn bộ quy trình của mô hình sở thích user và mô hình tài nguyên giảng dạy, mô hình này cũng thiết kế và triển khai nguyên mẫu hệ thống tài nguyên giảng dạy mạng được cá nhân hóa. Cuối cùng, so với thuật toán khuyến nghị truyền thống, thuật toán khuyến nghị kết hợp cải tiến có nhiều ưu điểm hơn trong hệ thống khuyến nghị tài nguyên giáo dục thông minh được cá nhân hóa.

Từ khóa—Smart education, Learning resource, Collaborative filtering, SVM

I. INTRODUCTION

Với sự phát triển không ngừng của công nghệ networks, web-based e-learning [1], [2] đang thay đổi cách con người tiếp thu kiến thức; ngày càng nhiều người học mong muốn tiếp thu thêm kiến thức thông qua cách thức cá nhân hóa và thông minh hơn. Trong môi trường e-learning, với sự mở rộng nhanh chóng của các nguồn thông tin và tài nguyên giảng dạy, tình trạng “*quá tải thông tin (information overload)*”, “*mất tài nguyên (resources lost)*” và các vấn đề khác lần lượt xuất hiện. Làm thế nào để đưa ra nguồn thông tin tài nguyên phù hợp nhất cho học sinh khối lượng thông tin khổng lồ là một bài toán quan trọng cần giải quyết như một phần không thể thiếu của công nghệ thông tin; *tài nguyên giảng dạy (teaching resources)* đóng một vai trò quan trọng trong giáo dục. *Hệ thống tài nguyên giảng dạy trực tuyến (network teaching resources system)* hiện có không thể đáp ứng nhu cầu cá nhân hóa của user. Do đó, để đạt được mục tiêu dịch vụ cá nhân hóa đáp ứng nhu cầu của user, công nghệ khuyến nghị được cá nhân hóa được ứng dụng vào hệ thống tài nguyên giảng dạy trực tuyến.

- (1) Hiện tại, user khó có thể đáp ứng nhu cầu đặc biệt của họ khi sử dụng vô số tài nguyên giảng dạy. Để giải quyết vấn đề này, khuyến nghị được cá nhân hóa được đưa vào hệ thống tài nguyên giảng dạy trực tuyến, cung cấp dịch vụ cá nhân hóa cho những người học có *nền tảng (background)* và *sở thích (interest)* khác nhau.

- (2) *Khuyến nghị được cá nhân hóa (Personalized recommendation)* [3] là không thể nếu không xác định sở thích của user. *Mô hình sở thích của user (A user interest model)* được đề xuất bằng cách kết hợp theo dõi user rõ ràng và ngầm định. Đầu tiên, thông tin cơ bản của user được thu thập từ dữ liệu đăng ký của họ. Tiếp theo, mô hình sở thích của user được thiết lập bằng cách trích xuất sự quan tâm của user từ hành vi hệ thống của họ. Cuối cùng, mô hình được cập nhật với những thay đổi trong hành vi của user.
- (3) Các công nghệ khuyến nghị được cá nhân hóa có liên quan được mô tả. Xem xét các đặc điểm của tài nguyên giảng dạy truyền thống và hệ thống tài nguyên giảng dạy network-based, một thuật toán mới kết hợp các phương pháp khuyến nghị *dựa trên nội dung (content-based)* và *dựa trên lọc cộng tác (collaborative filtering-based)* được đề xuất để khuyến nghị tài nguyên cho user. Một nền tảng giáo dục thông minh được triển khai hỗ trợ việc tùy chỉnh các tài nguyên giảng dạy trực tuyến. Tính khả thi và hiệu quả của thuật toán đề xuất được chứng minh.

II. RELATED THEORIES

A. Personalized recommendation of teaching resources

Những tiến bộ nhanh chóng trong công nghệ Internet và sự gia tăng theo cấp số nhân của lượng dữ liệu có sẵn từ Internet làm nổi bật nhu cầu về dịch vụ được cá nhân hóa. Trong bối cảnh này, tùy biến dịch vụ đã trở thành một vấn đề chính của nghiên cứu về xử lý thông tin thông minh và các kỹ thuật mạng. Mặc dù với nhiều năm phát triển, khuyến nghị dịch vụ được cá nhân hóa cho đến nay vẫn chưa hoàn thiện lắm, nhưng nó đã thâm nhập vào cuộc sống hàng ngày của chúng ta. Hiện nay, dịch vụ khuyến nghị được cá nhân hóa đã được sử dụng rộng rãi trong nhiều ngành khác nhau, bao gồm cả *thương mại điện tử (e-commerce)* và *công cụ tìm kiếm (search engine)*. Ngoài ra, nó đã được kết hợp vào các trang web giáo dục để cung cấp cho user những tài nguyên mà họ muốn.

Dựa theo đặc điểm của học sinh, học tập cá nhân hóa tạo ra một tập hợp các quá trình học tập được cá nhân hóa để học sinh nâng cao trình độ kiến thức và cho phép họ đạt được mục đích của việc học. Mục tiêu cơ bản của việc học tập cá nhân hóa là phát triển các học sinh một cách riêng biệt và hỗ trợ việc học tập được cá nhân hóa. Học tập cá nhân hóa tập

trung vào việc điều chỉnh trạng thái học tập của chính học sinh. Không cần sự tham gia của giáo viên, học sinh vẫn có thể học theo đặc điểm và nhu cầu của bản thân. Trong trường hợp này, việc xây dựng hệ thống học tập cá nhân hóa cần quan tâm hơn đến những vấn đề tồn tại trong quá trình học tập và hướng dẫn học sinh nâng cao hiệu quả học tập.

Học tập cá nhân hoá (personalized learning) là việc thực hiện các hoạt động giáo dục theo đặc điểm nhân cách của học sinh, xem xét đầy đủ tính chủ động của học sinh, thúc đẩy sự phát triển nhân cách của học sinh trên cơ sở phát huy tính toàn diện, tự do và phối hợp phát triển. Trên thực tế, học tập cá nhân hóa là học tập mang tính khám phá, thực hành và sáng tạo.

Về bản chất, hệ thống khuyến nghị được cá nhân hóa liên quan đến các trang web khác nhau hoặc các hệ thống ứng dụng khác (ví dụ: truy xuất thông tin) thu thập sở thích của user, phân tích thông tin user, xây dựng mô hình sở thích của user, cập nhật mô hình trong thời gian thực và cung cấp cho user nội dung đề cập những gì họ cần.

So với hệ thống tài nguyên giảng dạy trực tuyến truyền thống, hệ thống tài nguyên cá nhân hóa có hai ưu điểm sau.

(1) Đơn giản hóa việc truy xuất thông tin

Đối với những người học có mục đích rõ ràng, có thể dễ dàng xác định các nguồn thông tin mong muốn của họ bằng cách sử dụng hệ thống truy xuất [4], [5]. Nhưng đối với những người duyệt trang web một cách ngẫu nhiên, đó là một thách thức để tìm các nguồn mà họ quan tâm từ cơ sở dữ liệu thông tin đang tăng theo cấp số nhân. Việc kết hợp các khuyến nghị được cá nhân hóa vào hệ thống tài nguyên giảng dạy trực tuyến giúp có thể khuyến nghị các tài nguyên mong muốn cho user dựa trên hành vi đăng ký và hệ thống của họ, chuyển user từ trình duyệt thụ động sang người học chủ động.

(2) Thúc đẩy người học truy xuất thông tin

Việc truy cập nhanh chóng và thuận tiện vào các nguồn tài liệu thú vị sẽ làm tăng ham muốn học tập và thúc đẩy họ truy cập hệ thống thường xuyên hơn. Hơn nữa, nhiều bản ghi về hành vi của user sẽ được tích lũy trong hệ thống để cải thiện độ chính xác của khuyến nghị.

B. User interest profile

Hồ sơ sở thích của user (User interest profile) là một mô tả có thể tính toán được về thông tin sở thích của user. Nó xây dựng một mô hình để ghi lại và quản lý sở thích của user, mô tả yêu cầu về sở thích tiềm năng của user và ghi lại hành vi của user.

Mục đích chính của hồ sơ sở thích của user là dự đoán ý định của user trong một môi trường nhất định và cung cấp hoạt động trợ giúp cho user để user có thể tìm thấy các tài nguyên cần thiết một cách nhanh chóng và chính xác từ các nguồn thông tin khổng lồ.

Chìa khóa của khuyến nghị được cá nhân hóa là thiết lập hồ sơ sở thích của user. Chỉ bằng cách mô tả chính xác sở thích của user, chúng ta có thể cung cấp tài nguyên giảng dạy được cá nhân hóa và dịch vụ khuyến nghị theo sở thích của user. Hồ sơ sở thích của user trước tiên cần thu hút sự quan tâm

của user, có thể được thực hiện thông qua thông tin đăng ký của user, tải xuống, thu thập, đánh giá cũng như các hành vi khác để thu thập thông tin hành vi liên quan đến sở thích của user. Sau đó, phương pháp thích hợp được áp dụng để thiết lập mô hình sở thích của user. Cuối cùng, hồ sơ sở thích của user được cập nhật với sự thay đổi trong sở thích của user.

1) *Thu hút sự quan tâm của user (Obtaining user's interest)*: Quá trình theo dõi sở thích của user liên quan đến việc thu hút sự quan tâm của user. Quá trình này thường bao gồm hai khía cạnh: theo dõi rõ ràng và theo dõi ngầm.

Theo dõi rõ ràng (Explicit tracking) đề cập đến việc user điền vào biểu mẫu bằng cách nhập thông tin cá nhân hoặc trả lời các câu hỏi do hệ thống đưa ra, tham gia trực tiếp vào quá trình lập mô hình (ví dụ: thông tin cá nhân và đánh giá tài nguyên mà user đã điền khi đăng ký).

Theo dõi ngầm (Implicit tracking) không yêu cầu user cung cấp thông tin. Tất cả việc theo dõi được thực hiện tự động bởi hệ thống. User duyệt web, nhấp chuột, đánh dấu trang, kéo thanh cuộn và các bản ghi hành vi khác có thể cho biết sở thích tiềm năng của user. Nghiên cứu chỉ ra rằng duyệt trang, đánh dấu trang và kéo thanh cuộn có thể tiết lộ sở thích của user một cách hiệu quả, trong khi các hành động đơn giản (chẳng hạn như nhấp chuột) không thể thực hiện hiệu quả.

2) *Trình bày hồ sơ sở thích của user (Representation of user interest profile)*: Việc biểu diễn mô hình sở thích của user cần phản ánh thông tin thực của user và có thể tính toán được, nhưng nó cũng hạn chế sự lựa chọn phương pháp lập mô hình của user ở một mức độ nào đó. Có nhiều cách trình bày dựa trên các yêu cầu khác nhau của hệ thống khuyến nghị được cá nhân hóa. Một vài ví dụ như vậy là biểu diễn dựa trên mô hình không gian vectơ, biểu diễn dựa trên ma trận đánh giá, v.v.

(1) Biểu diễn dựa trên mô hình không gian vectơ (Representation based on vector space model)

Mô hình không gian vectơ (Vector Space Model - VSM)

[6], [7] là một phương pháp biểu diễn phổ biến của mô hình sở thích của user. VSM là biểu diễn không gian của tài liệu văn bản và được sử dụng phổ biến nhất trong các phương pháp biểu diễn. Trong mô hình này, một văn bản được coi là một tập các *item đặc trưng (feature item)* và văn bản được biểu diễn dưới dạng một vectơ n -chiều (n -dimensional) của không gian vectơ. Mỗi chiều tương ứng với một item đặt trưng trong toàn bộ tập văn bản. Ưu điểm của mô hình không gian vectơ là văn bản được biểu diễn dưới dạng vectơ bằng các thuật ngữ đặc trưng và trọng số để việc tính toán mối tương quan giữa các đoạn văn bản có thể được chuyển thành phép toán của mỗi tương quan giữa các vectơ. Mô hình sở thích của user sử dụng phương pháp này được biểu thị dưới dạng vectơ đặc trưng n -chiều. Trong phương pháp này, mô hình sở thích của user được biểu diễn dưới dạng vectơ đặc trưng n -chiều $\{(k_1, w_1), (k_2, w_2), \dots, (k_n, w_n)\}$. Mỗi chiều của eigenvector n -chiều bao gồm một từ khóa và một trọng số tương ứng. Trọng số có thể là giá trị thực hoặc giá trị Boolean, cho biết liệu user có quan tâm đến một tài nguyên bao gồm cả mức độ quan tâm hay không. Mô

hình không gian vectơ sử dụng tổ hợp các từ đặc trưng sở thích của user để mô tả sở thích của user và thể hiện tầm quan trọng của từng từ đặc trưng trong mô hình sở thích của user.

(2) *Biểu diễn dựa trên ma trận đánh giá (Representation based on evaluation matrix)*

Phương pháp biểu diễn ma trận đánh giá user-item là một ma trận $R_{m \times n}$ để thể hiện mô hình sở thích của user, trong đó m là số lượng user trong hệ thống và n là số lượng item. Mỗi phần tử r_{ij} trong ma trận đại diện cho đánh giá của user i đối với item j và nhận một giá trị nguyên trong một phạm vi thực (chẳng hạn như 1–5 và điểm càng lớn thì mức độ ưa thích của user đối với item đó càng cao; giá trị null cho biết rằng user không đánh giá item). Hầu hết các hệ thống dựa trên ma trận đánh giá là hệ thống khuyến nghị được cá nhân hóa dựa trên thuật toán lọc cộng tác.

III. SYSTEM DESIGN

Hệ thống được đề xuất để khuyến nghị tùy chỉnh các nguồn tài nguyên giảng dạy phục vụ học sinh, giáo viên và các nhân viên khác, cung cấp cho họ các tài nguyên đa phương tiện mà họ cần.

A. The construction of user interest model

Xác định sở thích của user là một quá trình thu thập phản hồi về mỗi quan tâm của user. Quá trình này là rõ ràng hoặc ngầm. Để xác định chính xác sở thích của user, hệ thống được đề xuất thu thập dữ liệu tĩnh về sở thích của user thông qua phản hồi rõ ràng và dữ liệu động về sở thích của user thông qua phản hồi ngầm. Mô hình sở thích của user được thể hiện trong Hình 1. Trong khi đăng ký, user mới nhập thông tin cơ bản và sở thích của họ theo cách thủ công, được lưu trữ trong bảng cơ sở dữ liệu thông tin user. Mô hình sở thích của user được khởi tạo bằng cách sử dụng dữ liệu đăng ký của user.

Hệ thống thu thập và theo dõi các đặc điểm của user đã đăng ký thông qua phản hồi ngầm. Cụ thể, các từ đặc trưng có thể đại diện cho các tài nguyên mà user quan tâm được xác định dựa trên lượt tải xuống. Những từ này có thể được

coi là nguồn quan tâm của user để khởi tạo mô hình sở thích ban đầu của user.

Đối với hệ thống tài nguyên giảng dạy, sự quan tâm của user đối với một số môn học là ổn định và user cũng có thể quan tâm đến các môn học khác do nhu cầu bổ sung. Do đó, hệ thống được đề xuất tiếp tục cập nhật mô hình sở thích của user dựa trên việc tải xuống, đánh giá tài nguyên và các hành vi khác của user. Do đó, độ chính xác của mô hình sở thích của user được cải thiện.

B. Personalized recommendation algorithm

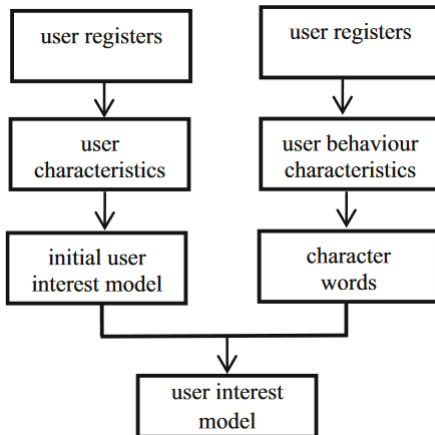
So sánh các thuật toán khuyến nghị khác nhau [8]–[10] chỉ ra rằng phương pháp lọc cộng tác item-based phù hợp hơn với hệ thống được đề xuất. Cốt lõi của hệ thống tài nguyên giảng dạy trực tuyến tùy chỉnh là mô-đun khuyến nghị, nó quyết định phần lớn hiệu suất của hệ thống. Mô-đun khuyến nghị phụ thuộc vào việc lựa chọn thuật toán khuyến nghị thích hợp, thuật toán này dự kiến sẽ được tích hợp vào ứng dụng, giảm các vấn đề có thể xảy ra, cải thiện chất lượng khuyến nghị và giảm bớt độ phức tạp của hệ thống. *Lọc cộng tác (Collaborative filtering)* là kỹ thuật khuyến nghị được nghiên cứu và sử dụng rộng rãi nhất. Nó rất hiệu quả và dựa trên sở thích của những user hàng xóm. Nghĩa là, mức độ tương tự của user được xác định dựa trên mức độ mà user khác yêu thích tài nguyên. Điểm đánh giá của user đối với một tài nguyên được xác định dựa trên mức độ mà những user tương tự thích tài nguyên đó. Với những dữ liệu này, hệ thống có thể đưa ra khuyến nghị tùy chỉnh một cách chính xác. Phương pháp khuyến nghị lọc cộng tác có thể được phân thành ba loại: user-based, item-based và model-based.

Để giải quyết *vấn đề thưa thớt (sparsity problem)* của phương pháp lọc cộng tác item-based, chúng tôi kết hợp phương pháp này với khuyến nghị *dựa trên nội dung (content-based)*. Tức là, mức độ quan tâm của user đối với tài nguyên không được đánh giá trước tiên sẽ được tính toán thông qua khuyến nghị dựa trên nội dung. Kết quả tính toán được định nghĩa là điểm đánh giá dự đoán của user đối với tài nguyên không được đánh giá, tạo thành ma trận đánh giá user–resource. Cuối cùng, ma trận đánh giá được sử dụng để tính toán mức độ tương tự item-based, tạo ra top-N các khuyến nghị hàng đầu.

(1) Input user–resource evaluation matrix

Ma trận đánh giá user–resource đầu vào $R(m, n)$ được suy ra từ đánh giá của user về tài nguyên và sở thích của user do hệ thống tạo ra đối với tài nguyên, trong đó hàng biểu thị m user, cột biểu thị n resource và phần tử $R_{i,j}$ biểu thị điểm của user i cho resource j . User phân bổ thứ hạng thích hợp cho tài nguyên và điều này xác định sự đóng góp của tài nguyên cho user. Thứ hạng càng cao, user càng quan tâm đến tài nguyên.

$$R(m, n) = \begin{bmatrix} R_{1,1} & R_{1,2} & \cdots & R_{1,n} \\ R_{2,1} & R_{2,2} & \cdots & R_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ R_{m,1} & R_{m,2} & \cdots & R_{m,n} \end{bmatrix} \quad (1)$$



Hình 1: Mô hình sở thích của user

Danh sách điểm đánh giá tài nguyên và danh sách điểm dự đoán cần được cập nhật bất cứ khi nào user đánh giá một tài nguyên mới hoặc thay đổi đánh giá một tài nguyên cũ. Đầu tiên, danh sách các điểm đánh giá tài nguyên được truy xuất để kiểm tra xem tài nguyên đã được đánh giá hay chưa. Nếu vậy, bản đánh giá mới sẽ thay thế bản đánh giá cũ. Nếu không, nó sẽ được chèn vào dưới dạng đánh giá của user về tài nguyên. Sau đó, đánh giá mới được sử dụng để tính lại sự quan tâm của user đối với các tài nguyên chưa được đánh giá. Bằng cách này, danh sách các điểm đánh giá được cập nhật để đảm bảo tính chính xác của ma trận đánh giá user-resource.

(2) Compute the nearest neighbor

Hệ số tương quan Pearson được sử dụng để tính toán mối tương quan giữa các item i và j .

$$Sim(i, j) = \frac{\sum_{u \in I_{ij}} (R_{u,i} - \bar{R}_i)(R_{u,j} - \bar{R}_j)}{\sqrt{\sum_{u \in I_{ij}} (R_{u,i} - \bar{R}_i)^2} \sqrt{\sum_{u \in I_{ij}} (R_{u,j} - \bar{R}_j)^2}} \quad (2)$$

Trong đó I_{ij} là tập hợp những user đã đánh giá các item i và j , $R_{u,i}$ là điểm đánh giá của user u cho item i và \bar{R}_i và \bar{R}_j là điểm trung bình của các item i và j . Sau đó, một giá trị thích hợp của k được chọn để trích xuất k item giống nhau nhất dưới dạng tập các hàng xóm gần nhất [11]–[13] của i .

(3) Make recommendation

Dự đoán điểm của user mục tiêu u cho item $P_{u,j}$ là:

$$P_{u,j} = \frac{\sum_{j=1}^k Sim(i, j) \times R_{u,j}}{\sum_{j=1}^k Sim(i, j)} \quad (3)$$

Trong đó k là tập hợp các tài nguyên trong danh sách hàng xóm gần nhất giống nhất với item i , $Sim(i, j)$ là sự giống nhau giữa các item i và j , $R_{u,j}$ là điểm đánh giá của thuật toán content-based dự đoán bằng cách sử dụng điểm đánh giá hiện có của u cho j .

Sau khi tính toán dự đoán điểm đánh giá của u cho các item khác nhau, top- N item hàng đầu có điểm cao nhất được xác định là nhóm khuyến nghị top- N hàng đầu.

IV. ANALYSIS AND DESIGN OF RECOMMENDATION STRATEGY

Thuật toán cải tiến của khuyến nghị được cá nhân hóa dựa trên khuyến nghị *kết hợp* (mixed) được xây dựng từ thuật toán khuyến nghị lọc cộng tác truyền thống và khuyến nghị dựa trên nội dung. Thuật toán này giới thiệu mô hình sở thích hiện có của user, sở thích tiềm năng của user, sở thích tổng hợp, v.v. Tiếp theo, việc triển khai thuật toán khuyến nghị được cá nhân hóa cải tiến dựa trên thuật toán khuyến nghị kết hợp sẽ được mô tả chi tiết.

(1) Establishing the Existing User Interest Model (EUIM)

Đối với bất kỳ user nào trong hệ thống, các từ khóa (f_1, f_2, \dots, f_k) của thông tin tài nguyên sở thích của user F , trong đó k đại diện cho k từ khóa của tài nguyên,

được tính toán bằng vectơ hóa văn bản và thu được trọng số vectơ F_1 . Công thức toán học có thể được biểu thị như sau:

$$EM = (w_{11}, w_{12} \dots w_{1j} \dots w_{1k}) \quad (4)$$

Trong đó w_{1j} là trọng số của từ khóa f_j trong EUIM.

(2) Building Potential User Interest Model (PUIM)

Đối với user trong hệ thống, các tài nguyên sở thích có liên quan trong tập hợp lân cận có mức độ tương quan cao được đẩy đến user mục tiêu bằng thuật toán khuyến nghị lọc cộng tác và sau đó vectơ trọng số của các từ khóa tài nguyên sẽ được thu lại. Công thức toán học có thể được biểu thị như sau:

$$PM = (w_{21}, w_{22} \dots w_{2j} \dots w_{2k}) \quad (5)$$

Trong đó w_{2j} là các từ khóa f_j của F trong PUIM.

(3) Building Fusion User Interest Model (FUIM)

Đối với user trong hệ thống, EUIM và PUIM được tính toán để tạo thành một vectơ trọng số mới và cuối cùng thiết lập FUIM. Biểu thức toán học của mô hình như sau:

$$FM = (w_{31}, w_{32} \dots w_{3j} \dots w_{3k}) \quad (6)$$

Trong đó w_{3j} là trọng số f_j của từ khóa F trong FUIM.

A. Construction method of user's existing interest model

Chúng ta lấy vectơ hóa văn bản của tập tài nguyên giáo dục đã cho $D = \{d_1, d_2 \dots d_i \dots d_n\}$ và từ khóa tài nguyên $f = \{f_1, f_2 \dots f_i \dots f_k\}$. Ma trận từ khóa và tài nguyên d_i trong tập tài nguyên giáo dục kết hợp với nhau để tạo thành mô hình không gian vectơ. $d_i = (w_{i1}, w_{i2} \dots w_{ij} \dots w_{ik})$, với w_{ij} là trọng số của từ khóa f_j trong tài nguyên của d_i . Trong trường hợp $w_{ij} = 0$, f_j không tồn tại trong tài nguyên d_i . Công thức toán học của ma trận trọng số tài nguyên như sau:

$$DM = \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1k} \\ w_{21} & w_{22} & \dots & w_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \dots & w_{nk} \end{bmatrix} \quad (7)$$

Các phương pháp phổ biến của biểu thức trọng số là: biểu diễn tần suất, biểu diễn TF-IDF và biểu diễn Boolean. Bài báo này chủ yếu áp dụng phương pháp TF-IDF.

B. Construction method of user's potential interest model

Mô hình sở thích tiềm năng của user khác với mô hình sở thích hiện tại của user. Nó không thể được truy xuất thông qua các bình luận hoặc lịch sử của user. Do số lượng lớn các tài nguyên giáo dục, danh sách được khuyến nghị không chỉ chứa các tài nguyên có liên quan trong mô hình sở thích của user mà còn chứa các tài nguyên sở thích tiềm năng của user. Trong bài báo này, một phương pháp lọc cộng tác được đề xuất để khuyến nghị sở thích của các nhóm user tương tự cho user mục tiêu, phương pháp này có thể được sử dụng để đại diện cho sở thích tiềm năng của user mục tiêu.

Trong lọc cộng tác truyền thống và phép tính độ tương đồng kết hợp, user chỉ sử dụng user để tính điểm tài nguyên, điều này không thể giải thích đầy đủ mức độ tương tự của user.

Ví dụ: trong hệ thống khuyến nghị tài nguyên giáo dục, mô tả hoặc loại file có thể có cho cùng một tài nguyên là khác nhau. Một số user quan tâm đến tài nguyên, nhưng số tài nguyên không giống nhau. Trong thuật toán lọc cộng tác, những user duyệt qua các tài nguyên khác nhau khó có thể được phân loại là những user tương tự. Theo quan điểm của vấn đề trên, bài báo này sử dụng một phương pháp tính toán mức độ hình thức kết hợp hành vi và nội dung. Sự giống nhau của user được mở rộng thành hai phần: score similarity ($sim_{grade}(u, v)$) và content similarity ($sim_{content}(u, v)$).

Giả sử tập xếp hạng tài nguyên của user u như sau:

$$\begin{aligned} D_u &= \{d_{u1}, d_{u2}, \dots, d_{ui}, \dots, d_{um}\} \\ EM_u &= \{w1_{u1}, w1_{u2}, \dots, w1_{uj}, w1_{uk}\} \end{aligned}$$

Xếp hạng tài nguyên được thiết lập cho user v như sau:

$$\begin{aligned} D_v &= \{d_{v1}, d_{v2}, \dots, d_{vi}, \dots, d_{vm}\} \\ EM_v &= \{w1_{v1}, w1_{v2}, \dots, w1_{vj}, w1_{vk}\} \end{aligned}$$

Score similarity của user u và v như sau:

$$sim_{grade}(u, v) = \frac{\sum_{i \in D_u \cap D_v} \frac{1}{\log(1+|U(i)|)}}{\sqrt{|D_u| |D_v|}} \quad (8)$$

Trong đó $U(i)$ là tập user đã nhận xét về tài nguyên d_i .

Công thức tính content similarity của user u và v như sau:

$$sim_{content}(u, v) = \frac{EM_u \cdot EM_v}{|EM_u| \cdot |EM_v|} \quad (9)$$

Công thức kết hợp của bài báo này được trình bày như sau:

$$sim(u, v) = \beta sim_{grade}(u, v) + (1 - \beta) sim_{content}(u, v) \quad (10)$$

Hệ số β là trọng số được xác định bằng thực nghiệm, là tham số phần trăm tương tự và phạm vi giá trị của nó là 0–1.

Khi $\beta = 0$, phép tính độ tương tự chỉ xem xét dữ liệu content feature và khi $\beta = 1$, phép tính độ tương tự chỉ xem xét dữ liệu score feature. Score similarity và content similarity của user u và v được tính toán, sau đó, hệ số có trọng số β được sử dụng để kết hợp hai similarity nhằm thu được similarity kết hợp của user. Có được sự tương đồng giữa user mục tiêu và tất cả những user khác, và user tương tự nhất và h user được chọn làm nhóm user tương tự; và nhóm user tương tự được khuyến nghị cho user mục tiêu bằng cách lọc cộng tác. Có được mô hình sở thích tiềm năng của user mục tiêu.

Đặt nhóm user tương tự là $U_u = \{v_1, v_2, \dots, v_i, \dots, v_k\}$ của user u , sự tương đồng giữa user u và bất kỳ user v_i nào là $sim(u, v_i)$. Mô hình sở thích hiện tại của user v_i là $EM_{v_i} = \{w1_{v_i1}, w1_{v_i2}, \dots, w1_{v_ij}, \dots, w1_{v_ik}\}$. Tính trọng số của feature term f_i của mô hình sở thích tiềm năng của user bằng công thức sau:

$$w2_{uj} = \sum_{v_i \in U_u} \frac{sim(u, v_i)}{\sum_{v_i \in U_u} sim(u, v_i)} \cdot w1_{v_ij} \quad (11)$$

C. A method of constructing mixed user interest model

Sau khi có được mô hình sở thích hiện tại (EUM) và mô hình sở thích tiềm năng (PUIM) của user mục tiêu, trọng số của các feature word của hai mô hình được hợp nhất và có thể thu được sở thích kết hợp của user mục tiêu. Sau đó, tính toán sự tương đồng giữa vectơ trọng số của các main feature word của candidate educational resource và FUIM, so sánh kết quả tính được với ngưỡng đã đặt để thu được các kết quả khuyến nghị cuối cùng.

Đặt EUM của user u : $EM_u = \{w1_{u1}, w1_{u2}, \dots, w1_{uj}, \dots, w1_{uk}\}$, đặt PUIM của user u : $PM_u = \{w2_{u1}, w2_{u2}, \dots, w2_{uj}, \dots, w2_{uk}\}$, đặt FUIM của user u như sau: $FM_u = \{w3_{u1}, w3_{u2}, \dots, w3_{uj}, \dots, w3_{uk}\}$, candidate resource $d = \{wd_1, wd_2, \dots, wd_j, \dots, wd_m\}$.

$$w3_{uj} = \max(w1_{uj}, w2_{uj}) \quad (12)$$

Trong Công thức 12, hàm \max đại diện cho việc lựa chọn một giá trị lớn của $w1_{uj}$ và $w2_{uj}$ được khuyến nghị cho user.

V. EXPERIMENTAL RESULT AND DISCUSSION

A. Experimental data source

Dữ liệu sử dụng trong thực nghiệm này được cung cấp bởi nền tảng học tập trên web (<http://evaluate.guoshi.com/publishg/>). Chúng tôi sử dụng 1-month log file once again và trích xuất dữ liệu về hành vi truy cập của các học giả, với số lượng học giả là 540, số lượng dữ liệu trực tuyến 2780 và gần 160,000 records. Để đánh giá kết quả tốt hơn, 80% dữ liệu được lấy làm tập dữ liệu training và dữ liệu tương ứng còn lại làm tập dữ liệu testing.

B. System evaluation metrics

Để đánh giá tính khả thi và hiệu quả của thuật toán khuyến nghị được sử dụng trong hệ thống tài nguyên giảng dạy trực tuyến, phương pháp khuyến nghị kết hợp được so sánh với mô hình item-based truyền thống và mô hình lọc cộng tác. Các thước đo đánh giá bao gồm: *recall*, *precision* và *F measure*.

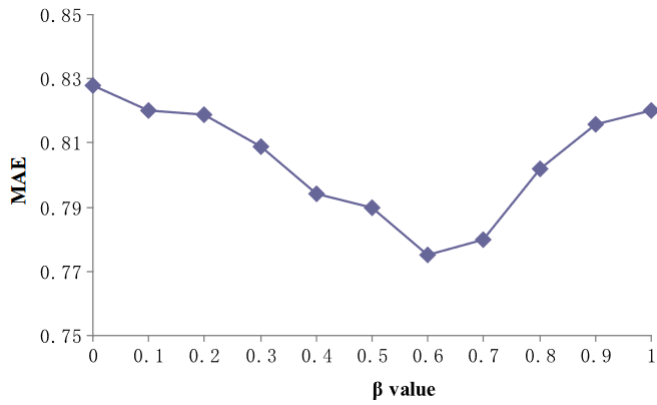
Sai số tuyệt đối trung bình (Mean Absolute Error – MAE) đề cập đến giá trị tuyệt đối trung bình của sự khác biệt giữa score thực tế và score dự đoán của tài nguyên.

$$MAE = \frac{\sum_{i=1}^n |q_i - p_i|}{n} \quad (13)$$

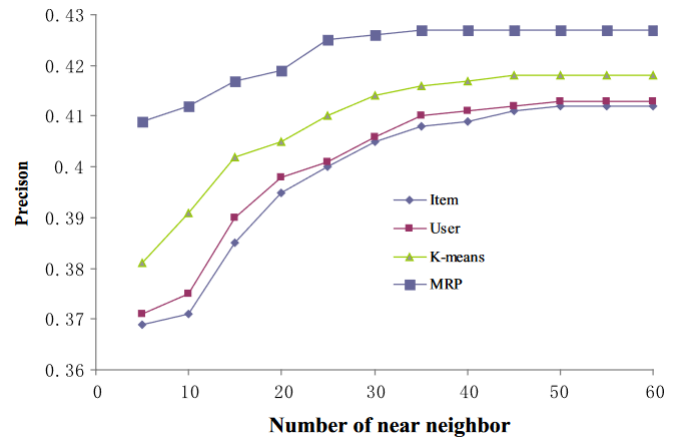
Trong đó p_i là score đánh giá user-predicted, q_i là score user thực tế và tập hợp là $\{q_1, q_2, q_i, \dots, q_n\}$.

C. Experimental results and analysis

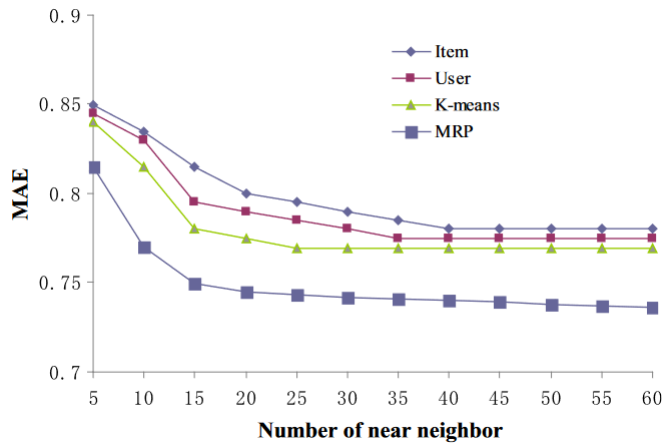
Trong công thức 12, tham số β là một tham số tỷ lệ similarity dựa trên hàm trọng số thời gian và kích cỡ tập hợp của nó sẽ ảnh hưởng trực tiếp đến hiệu quả của khuyến nghị. Do đó, lặp lại thực nghiệm là cần thiết để xác định giá trị tốt nhất, đảm bảo rằng kết quả khuyến nghị là tối ưu và khoảng giá trị của β là [0–1]. Trong thực nghiệm này, các giá trị của



Hình 2: Kết quả thực nghiệm tham số β



Hình 4: Kết quả thực nghiệm precision



Hình 3: Kết quả thực nghiệm MAE

β được đặt từ 0 đến 1 và giá trị tăng trưởng là 0.1. Kết quả thực nghiệm được thể hiện trong Hình 2.

Như hình trên có thể thấy, khi $\beta = 0.6$, chúng ta có thể nhận được giá trị sai số khuyến nghị MAE nhỏ nhất và độ chính xác của khuyến nghị cao nhất. Khi tập hợp là 0 và 1, nó thể hiện rằng phép tính độ tương đồng chỉ xem xét dữ liệu content feature và rating feature. Nhưng hiệu suất khuyến nghị không phải là tốt nhất, vì vậy ta lấy $\beta = 0.6$ trong thực nghiệm sau.

Dữ liệu thu được trong phần này được áp dụng tương ứng cho các thuật toán lọc cộng tác phổ biến, thuật toán lọc cộng tác item-based, thuật toán lọc cộng tác dựa trên K-means và thuật toán Material Requirements Planning (MRP) được đề xuất trong bài báo này và phân tích kết quả thực nghiệm. Tính hiệu quả của thuật toán mới được xác minh bằng cách so sánh một số thuật toán lọc cộng tác phổ biến. Giá trị MAE của các phương pháp khác nhau thu được bằng cách chọn các số lân cận gần nhất khác nhau. Kết quả thực nghiệm được thể hiện trong Hình 3.

Từ Hình 3, chúng ta có thể thấy rằng thuật toán lọc cộng tác dựa trên khuyến nghị kết hợp tốt hơn các thuật toán truyền thống về hiệu suất MAE. Khi số lượng lân cận gần nhất là khoảng 30, giá trị của MAE sẽ không thay đổi. Người ta chứng minh rằng thuật toán sẽ đạt đến trạng thái tốt nhất khi số lân

cận gần nhất là khoảng 30. Nó được chứng minh bằng cách so sánh rằng thuật toán đề xuất mới đóng một vai trò nhất định trong việc cải thiện độ chính xác của các khuyến nghị.

Để xác minh độ chính xác của thuật toán, các giá trị precision của thuật toán này được so sánh với các thuật toán trên. Kết quả của thực nghiệm được thể hiện trong Hình 4.

Kết quả thực nghiệm cho thấy độ chính xác của thuật toán được đề xuất cao hơn so với thuật toán lọc cộng tác truyền thống và tính hợp lệ của thuật toán đã được xác minh.

VI. CONCLUSION

Với sự phát triển nhanh chóng của Internet và khoa học công nghệ, ngày càng có nhiều người tiếp thu kiến thức và kỹ năng thông qua mạng. “Information overload” đã trở thành một yếu tố quan trọng hạn chế sự phát triển của hình thức học trực tuyến. Trong bài báo này, tác giả đã nghiên cứu một công nghệ khuyến nghị được cá nhân hóa phổ biến và giới thiệu nguyên tắc, quy trình và chiến lược của phương pháp một cách chi tiết. Trên cơ sở nghiên cứu sâu rộng các tài liệu về công nghệ khuyến nghị, tác giả đã nghiên cứu các vấn đề tồn tại trong công nghệ khuyến nghị lọc cộng tác theo tài nguyên giảng dạy và đặc điểm của user, từ đó đưa ra một số giải pháp.

Dịch vụ khuyến nghị được cá nhân hóa tài nguyên giảng dạy có triển vọng lớn. Mặc dù bài báo này đã thực hiện một số nghiên cứu về khuyến nghị cá nhân hóa, nhưng hạn chế về thời gian và điều kiện, nên vẫn còn một số vấn đề chưa được giải quyết. Công việc chính sau đây bao gồm: thuật toán khuyến nghị chỉ cải thiện ở khía cạnh *khởi động lạnh* (cold start) và các vấn đề mới của dự án vẫn chưa được giải quyết. Hệ thống chỉ có thể được khuyến nghị cho user thông qua các tài nguyên mới nhất, với mức độ cá nhân hóa thấp và cần nghiên cứu thêm trong tương lai.

TÀI LIỆU

- [1] Bond, S.E., Crowther, S.P., Adhikari, S., Chubaty, A.J., Yu, P., Borchard, J.P., Boutlis, C.S., Yeo, W.W. and Miyakis, S., 2017. Design and implementation of a novel web-based e-learning tool for education of health professionals on the antibiotic vancomycin. Journal of medical Internet research, 19(3), p.e6971.

- [2] Kassak, O., Kompan, M. and Bielikova, M., 2016. Student behavior in a web-based educational system: Exit intent prediction. *Engineering Applications of Artificial Intelligence*, 51, pp.136-149.
- [3] Zuo, Y., Gong, M., Zeng, J., Ma, L. and Jiao, L., 2015. Personalized recommendation based on evolutionary multi-objective optimization [research frontier]. *IEEE Computational Intelligence Magazine*, 10(1), pp.52-62.
- [4] Kalloubi, F., 2016. Microblog semantic context retrieval system based on linked open data and graph-based theory. *Expert Systems with Applications*, 53, pp.138-148.
- [5] Wang, H., Zhang, Q. and Yuan, J., 2017. Semantically enhanced medical information retrieval system: a tensor factorization based approach. *Ieee Access*, 5, pp.7584-7593.
- [6] Xie, L., Li, G., Xiao, M. and Peng, L., 2016. Novel classification method for remote sensing images based on information entropy discretization algorithm and vector space model. *Computers & Geosciences*, 89, pp.252-259.
- [7] Zhang, Y., Lo, D., Xia, X., Scanniello, G., Le, T.D.B. and Sun, J., 2018. Fusing multi-abstraction vector space models for concern localization. *Empirical Software Engineering*, 23(4), pp.2279-2322.
- [8] Xu, Z., Luo, X., Zhang, S., Wei, X., Mei, L. and Hu, C., 2014. Mining temporal explicit and implicit semantic relations between entities using web search engines. *Future Generation Computer Systems*, 37, pp.468-477.
- [9] Chen, H., Li, Z. and Hu, W., 2016. An improved collaborative recommendation algorithm based on optimized user similarity. *The Journal of Supercomputing*, 72(7), pp.2565-2578.
- [10] Huang, S., Zhang, J., Wang, L. and Hua, X.S., 2015. Social friend recommendation based on multiple network correlation. *IEEE transactions on multimedia*, 18(2), pp.287-299.
- [11] Samanthula, B.K., Elmehdwi, Y. and Jiang, W., 2014. K-nearest neighbor classification over semantically secure encrypted relational data. *IEEE transactions on Knowledge and data engineering*, 27(5), pp.1261-1273.
- [12] Xu, Z., Liu, Y., Mei, L., Hu, C. and Chen, L., 2014. Generating temporal semantic context of concepts using web search engines. *Journal of Network and Computer Applications*, 43, pp.42-55.
- [13] Ye, J. and Ding, Y., 2018. Controllable keyword search scheme supporting multiple users. *Future Generation Computer Systems*, 81, pp.433-442.