



ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN
KHOA KHOA HỌC & KỸ THUẬT THÔNG TIN

Intelligent Learning System based on Personalized Recommendation Technology

SVTH: Team 10

- Phạm Đức Thể
- Trần Thành Luân

GVHD:

- Ths. Nguyễn Văn Kiệt
- CN. Huỳnh Văn Tín

NỘI DUNG

1

Giới thiệu

2

Cơ sở lý thuyết

3

Thiết kế hệ thống

4

Phân tích & Thiết kế
chiến lược khuyến nghị

5

Kết quả thực nghiệm

6

Kết luận

1

GIỚI THIỆU



- ❖ Trong môi trường e-learning, **tài nguyên giảng dạy** đóng một vai trò quan trọng trong giáo dục. Với sự mở rộng nhanh chóng của các nguồn thông tin và tài nguyên giảng dạy dẫn đến:
 - Quá tải thông tin
 - Mất tài nguyên
 - Các vấn đề khác lần lượt xuất hiện.
- **Hệ thống tài nguyên giảng dạy trực tuyến (network teaching resources system)** hiện thời không thể đáp ứng nhu cầu cá nhân hóa của user.

1 GIỚI THIỆU



- Mô hình sở thích của user (A user interest model) được đề xuất bằng cách kết hợp theo dõi user công khai và ngầm định.
- 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 cách cá nhân hóa.

2

CƠ SỞ LÝ THUYẾT

❖ Khuyến nghị về tài nguyên giảng dạy được cá nhân hóa

- Cá nhân hoá việc học tập (personalized learning) là việc thực hiện các hoạt động giáo dục theo những đặc điểm cá nhân của học sinh.
- Đơn giản hóa việc truy xuất thông tin mong muốn.
- Thúc đẩy user truy xuất thông tin.



CƠ SỞ LÝ THUYẾT

❖ Hồ sơ sở thích của user (User interest profile)

- Thu hút sự quan tâm của user (Obtaining user's interest)
 - Theo dõi công khai 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.
 - Theo dõi ngầm không yêu cầu user cung cấp thông tin. Tất cả việc theo dõi được thực hiện tự động bởi hệ thống.

2 CƠ SỞ LÝ THUYẾT

❖ Hồ sơ sở thích của user (User interest profile)

▪ Trình bày hồ sơ sở thích của user (Representation of user interest profile):

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

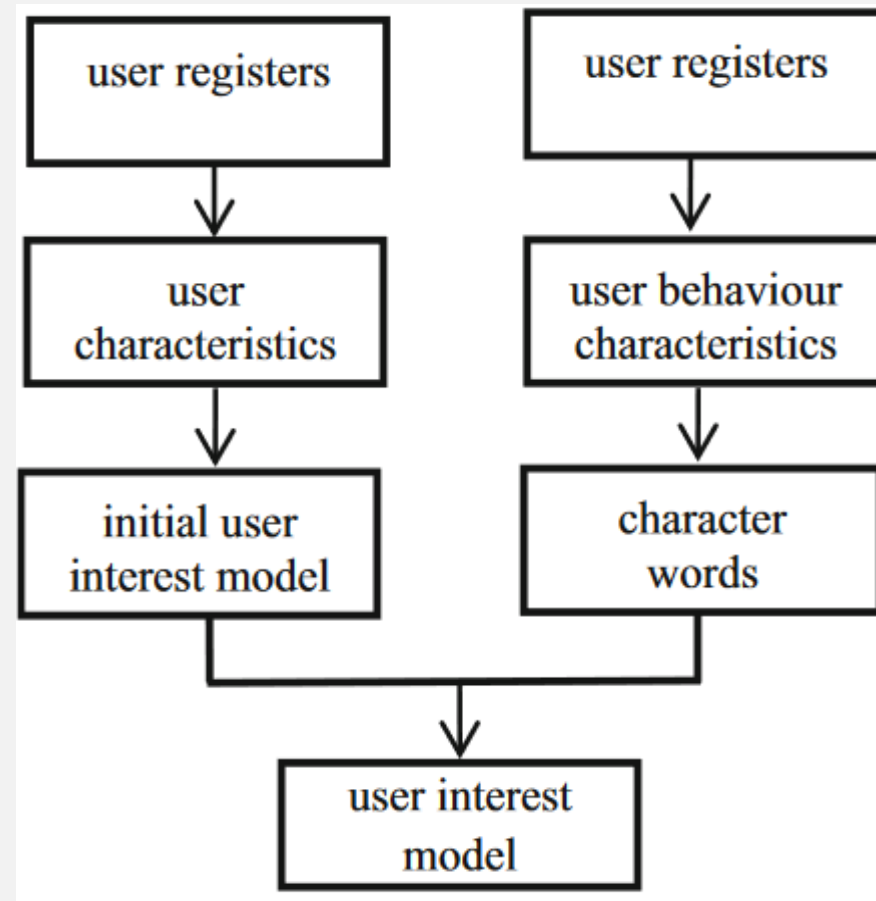
$$\{(k_1, w_1), (k_2, w_2), \dots, (k_n, w_n)\}$$

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

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

3 THIẾT KẾ HỆ THỐNG

❖ Xây dựng mô hình sở thích của user



3 THIẾT KẾ HỆ THỐNG

❖ Thuật toán khuyến nghị được cá nhân hóa

- Đầu vào ma trận đánh giá user–resource

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

- Tính toán các lân cận gần nhất

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

- Đưa ra khuyến nghị

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

4 PHÂN TÍCH & THIẾT KẾ CHIẾN LƯỢC KHUYẾN NGHỊ

- ❖ Phương pháp xây dựng mô hình sở thích hiện tại của user
- ❖ Phương pháp xây dựng mô hình sở thích tiềm năng của user
- ❖ Phương pháp xây dựng mô hình sở thích kết hợp của user

4 PHÂN TÍCH & THIẾT KẾ CHIẾN LƯỢC KHUYẾN NGHỊ

❖ Phương pháp xây dựng mô hình sở thích hiện tại của user

- Existing User Interest Model (EUIM).
- Vector hóa văn bản của tập tài nguyên giáo dục đã cho $D = \{d_1, d_2, \dots, d_i, \dots, d_n\}$.
- Từ khóa tài nguyên $f = \{f_1, f_2, \dots, f_i, \dots, f_k\}$.
- $d_i = (w_{i1}, w_{i2}, \dots, w_{ij}, \dots, w_{ik})$, 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 .

4 PHÂN TÍCH & THIẾT KẾ CHIẾN LƯỢC KHUYẾN NGHỊ

❖ Phương pháp xây dựng mô hình sở thích hiện tại của user

- Công thức toán học của ma trận trọng số tài nguyên như sau:

$$DM = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1k} \\ w_{21} & w_{22} & \cdots & w_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nk} \end{bmatrix}$$

- Các phương pháp phổ biến của biểu thức trọng số là:
 - Biểu diễn tần suất.
 - **Biểu diễn TF-IDF.**
 - Biểu diễn Boolean.

4 PHÂN TÍCH & THIẾT KẾ CHIẾN LƯỢC KHUYẾN NGHỊ

❖ Phương pháp xây dựng mô hình sở thích tiềm năng của user

- Potential User Interest Model (PUIM)
- 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.
- Một phương pháp lọc cộng tác được đề xuất để khuyến nghị sở thích của các nhóm user tương tự cho user mục tiêu.

4

PHÂN TÍCH & THIẾT KẾ CHIẾN LƯỢC KHUYẾN NGHỊ

❖ Phương pháp xây dựng mô hình sở thích tiềm năng của user

- 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.
- 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 (user similarity) được mở rộng thành hai phần:
 - Score similarity ($sim_{grade}(u, v)$).
 - Content similarity ($sim_{content}(u, v)$).

4 PHÂN TÍCH & THIẾT KẾ CHIẾN LƯỢC KHUYẾN NGHỊ

❖ Phương pháp xây dựng mô hình sở thích tiềm năng của user

- Giả sử:
 - Resources rating set for user u :

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

- Resources rating set for user v :

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

4 PHÂN TÍCH & THIẾT KẾ CHIẾN LƯỢC KHUYẾN NGHỊ

❖ Phương pháp xây dựng mô hình sở thích tiềm năng của user

- **Score similarity** của user u và v như sau:

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

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

- **Content similarity** của user u và v như sau:

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

4 PHÂN TÍCH & THIẾT KẾ CHIẾN LƯỢC KHUYẾN NGHỊ

❖ Phương pháp xây dựng mô hình sở thích tiềm năng của user

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

$$\text{sim}(u, v) = \beta \text{sim}_{\text{grade}}(u, v) + (1 - \beta) \text{sim}_{\text{content}}(u, v)$$

- Hệ số β là trọng số được xác định bằng thực nghiệm, là **tham số phần trăm tương tự (similarity percentage parameter)**.
- Khi $\beta = 0$, phép tính độ tương tự chỉ xem xét dữ liệu **content feature**.
- Khi $\beta = 1$, phép tính độ tương tự chỉ xem xét dữ liệu **score feature**.

4 PHÂN TÍCH & THIẾT KẾ CHIẾN LƯỢC KHUYẾN NGHỊ

❖ Phương pháp xây dựng mô hình sở thích tiềm năng của user

- Có được sự tương đồng giữa user mục tiêu và tất cả những user khác, và h user tương tự nhất được chọn làm nhóm user tương tự.
- 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.

4 PHÂN TÍCH & THIẾT KẾ CHIẾN LƯỢC KHUYẾN NGHỊ

❖ Phương pháp xây dựng mô hình sở thích tiềm năng của user

- Đặt nhóm user tương tự của user u là $U_u = \{v_1, v_2, \dots, v_i, \dots, v_k\}$, sự tương đồng giữa user u và bất kỳ user v_i nào là $\text{sim}(u, v_i)$. Mô hình sở thích hiện tại của user v_i là $EM_{v_i} = \{w1_{vi1}, w1_{vi2}, \dots, w1_{vij}, \dots, w1_{vik}\}$.
- Tính trọng số của **feature term** f_i của mô hình sở thích tiềm năng của user bằng công thức sau:

$$w2_{uj} = \sum_{v_i \in U_u} \frac{\text{sim}(u, v_i)}{\sum_{v_i \in U_u} \text{sim}(u, v_i)} \cdot w1_{vij}$$

4 PHÂN TÍCH & THIẾT KẾ CHIẾN LƯỢC KHUYẾN NGHỊ

❖ Phương pháp xây dựng mô hình sở thích kết hợp của user

- Fusion User Interest Model (FUIM)
- 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.
- Tính toán sự tương đồng giữa vector 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.

4 PHÂN TÍCH & THIẾT KẾ CHIẾN LƯỢC KHUYẾN NGHỊ

❖ Phương pháp xây dựng mô hình sở thích kết hợp của user

- Đặt EUIM 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 : $FM_u = \{w3_{u1}, w3_{u2}, \dots, w3_{uj}, \dots, w3_{uk}\}$, candidate resource $d = \{wd_1, wd_2, \dots, wd_j, \dots, wd_m\}$.

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

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



KẾT QUẢ THỰC NGHIỆM

- ❖ Dữ liệu thực nghiệm
- ❖ Độ đo đánh giá hệ thống
- ❖ Kết quả & Phân tích thực nghiệm

5

KẾT QUẢ THỰC NGHIỆM

❖ Dữ liệu thực nghiệm

- Dữ liệu sử dụng trong thực nghiệm này được cung cấp bởi nền tảng học tập trên web (<http://evaluate.guoshi.com/publishg/>).
- Tác giả sử dụng log file được cập nhật hàng tháng.
- Số lượng học giả là 540, số lượng dữ liệu trực tuyến 2780 và gần 160,000 record.
- 80% dữ liệu được lấy làm tập dữ liệu training và dữ liệu tương ứng còn lại làm tập dữ liệu testing.



KẾT QUẢ THỰC NGHIỆM

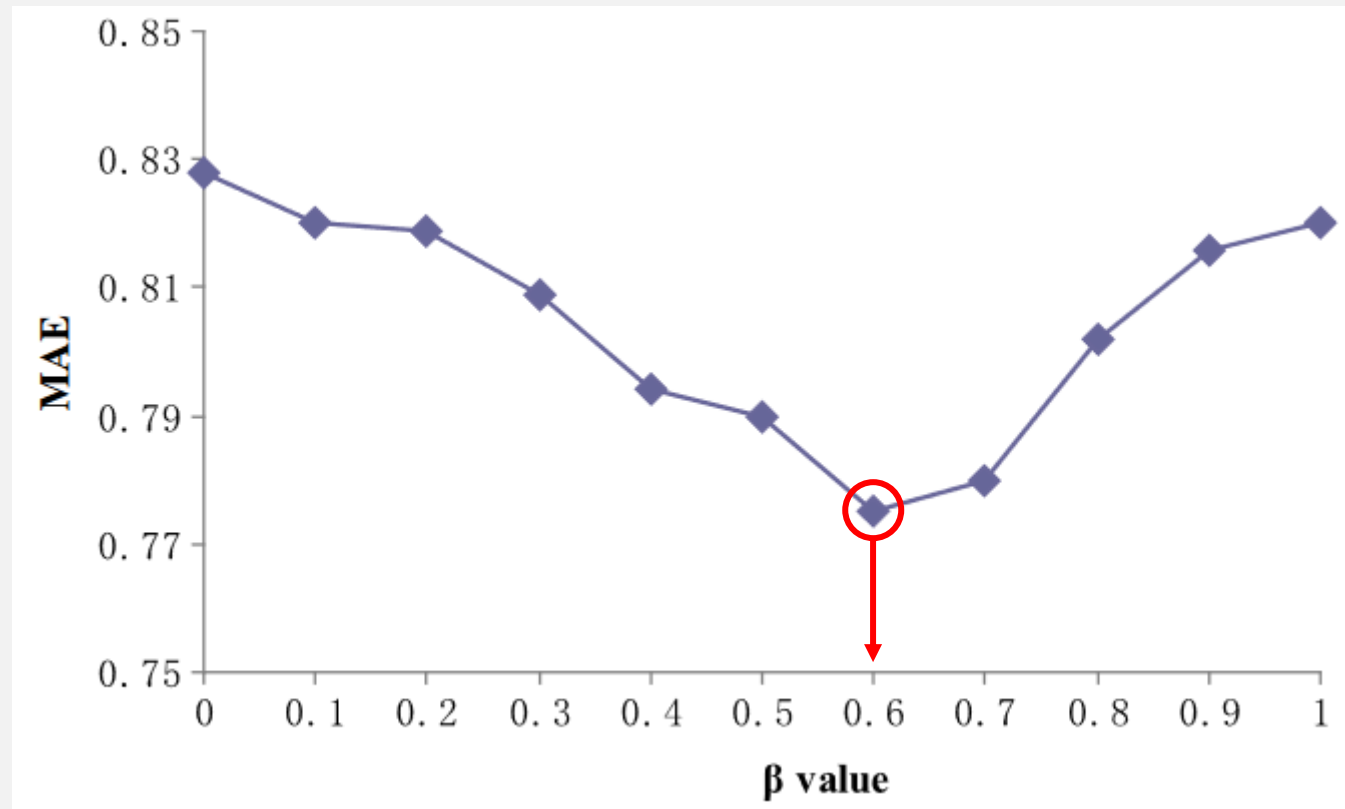
❖ Độ đo đánh giá hệ thống

- Recall
- Precision
- F measure
- Mean Absolute Error (MAE)

5

KẾT QUẢ THỰC NGHIỆM

❖ Kết quả & Phân tích thực nghiệm

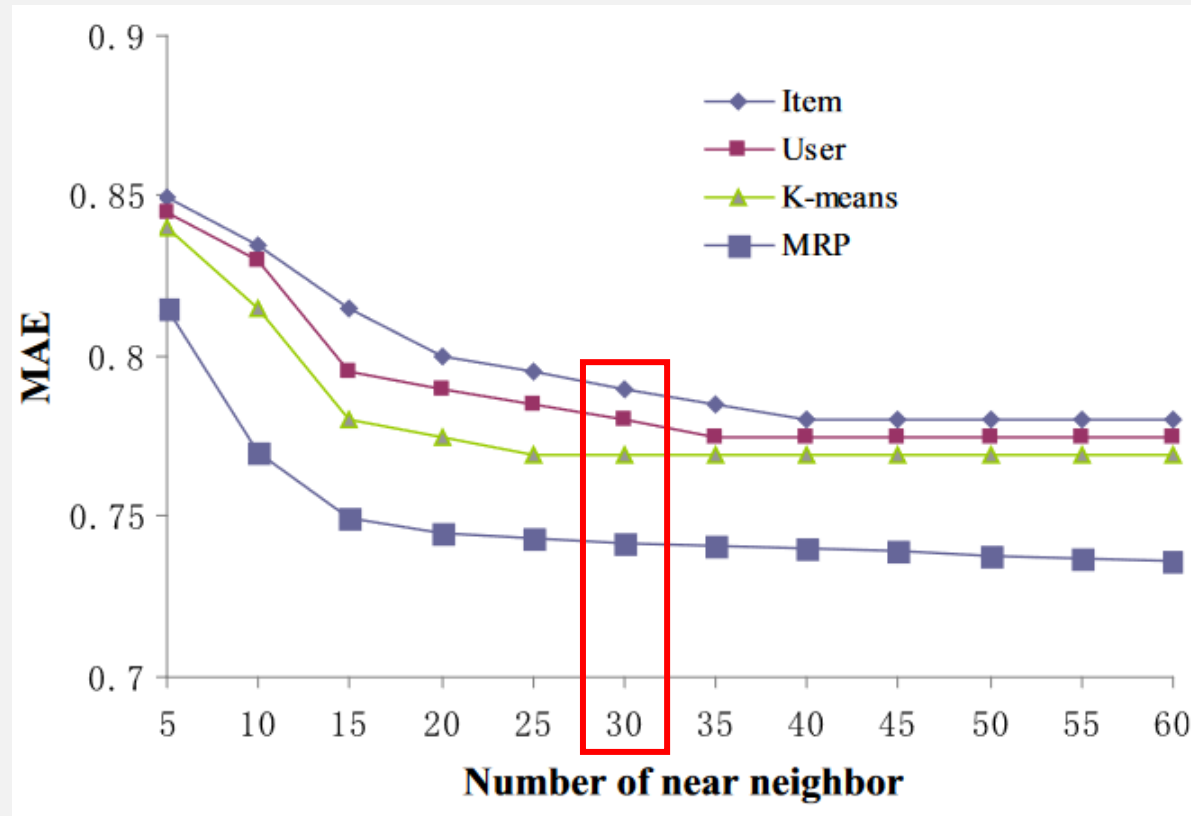


Kết quả thực nghiệm tham số β

5

KẾT QUẢ THỰC NGHIỆM

❖ Kết quả & Phân tích thực nghiệm

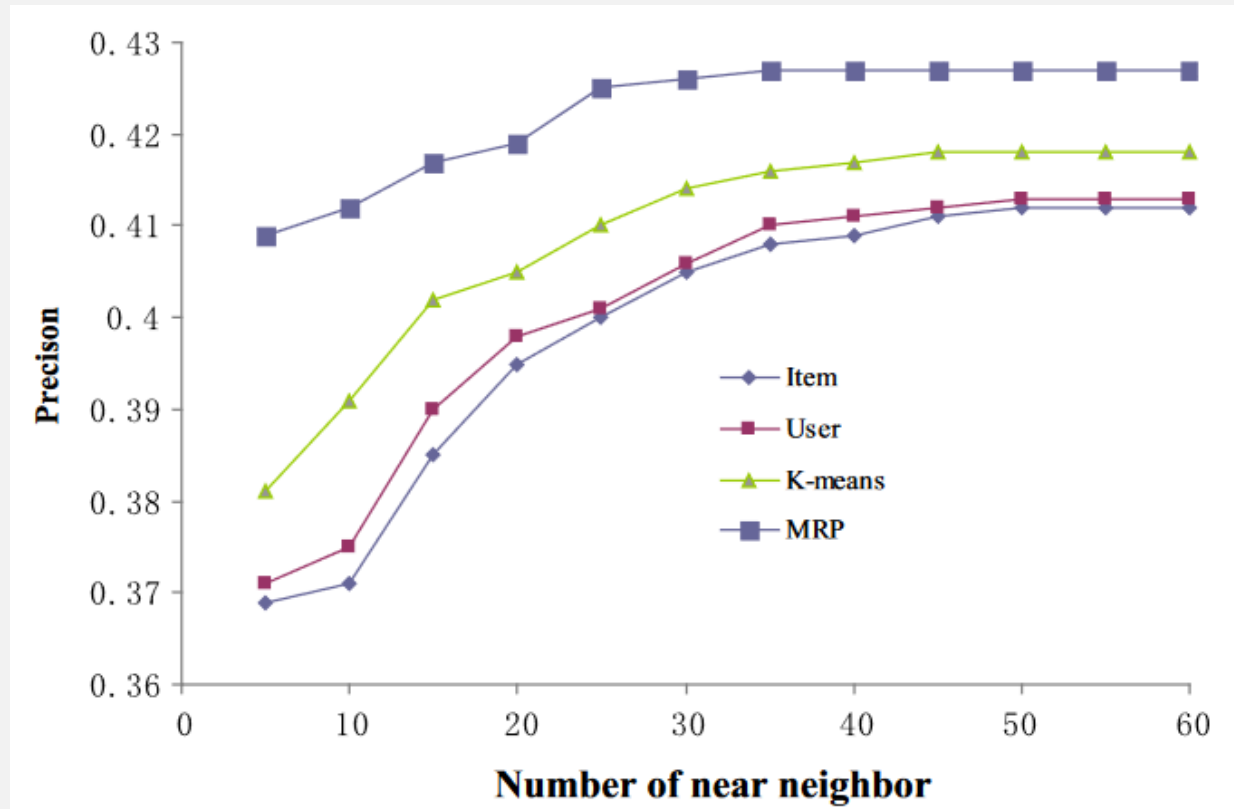


Kết quả thực nghiệm MAE

5

KẾT QUẢ THỰC NGHIỆM

❖ Kết quả & Phân tích thực nghiệm



Kết quả thực nghiệm Precision

6 KẾT LUẬN

❖ Kết luận

- 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.
- Mô hình khuyến nghị 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 đã giải quyết được vấn đề khuyến nghị tài nguyên học tập cho user một cách cá nhân hóa.
- Thuật toán khuyến nghị đã cải thiện ở khía cạnh cold start.

6 KẾT LUẬN

❖ Khó khăn

- Các vấn đề mới của dự án vẫn chưa được giải quyết
 - Hạn chế về thời gian, điều kiện.
 - Chỉ có thể được khuyến nghị cho user thông qua các tài nguyên mới nhất với mức độ cá nhân hóa thấp.
- Cần nghiên cứu thêm trong tương lai.



QUESTION & ANSWER

