



Đánh giá các phương pháp xây dựng cộng đồng người học trên hệ thống tư vấn học tập trong môi trường đào tạo trực tuyến

BÙI XUÂN HUY^{a,*}, NGUYỄN AN TẾ^a, TRẦN THỊ SONG MINH^b

^a Trường Đại học Kinh tế TP. Hồ Chí Minh

^b Trường Đại học Kinh tế Quốc dân

THÔNG TIN	TÓM TẮT
Ngày nhận: 31/01/2023 Ngày nhận lại: 08/04/2023 Duyệt đăng: 21/04/2023 Mã phân loại JEL: C61; C63; C67. Từ khóa: Đào tạo trực tuyến; Hệ thống tư vấn; Hồ sơ người học; Cộng đồng người học. Keywords: E-Learning; Recommend System; Learner profile; Learner community.	Đào tạo trực tuyến đang phát triển mạnh mẽ nhưng cũng khiến người học dễ bị lạc lõng, mất phương hướng trong môi trường học tập ảo rộng lớn vì không được tiếp xúc với giảng viên và bạn đồng học cụ thể. Từ đó dẫn đến tình trạng người học bị suy giảm động lực học tập. Vì vậy, nhu cầu phát triển các hệ thống tư vấn để giải quyết vấn đề quá tải thông tin, giúp cho người học dễ dàng truy cập chính xác, kịp thời những tài nguyên, dịch vụ mình quan tâm. Nguyên tắc của hệ tư vấn học tập là dựa trên cộng đồng những người học có đặc trưng tương tự và đạt kết quả học tập tốt để đưa ra tư vấn phù hợp cho một người học cụ thể trong các vấn đề cần ra quyết định như chọn cách thức học tập, tài nguyên học tập, nhóm học tập, đăng ký môn học. Mức độ chính xác của việc xây dựng cộng đồng người học sẽ ảnh hưởng đến chất lượng tư vấn. Nội dung của bài báo này sẽ trình bày về các phương pháp xây dựng cộng đồng người học trên hệ thống tư vấn học tập trong môi trường đào tạo trực tuyến cũng như vấn đề đánh giá hiệu quả của các phương pháp. Abstract Online training is growing strongly but it also makes learners easily lost and disoriented in a large virtual learning environment because there are no instructors and no specific classmates. As a result, the learner's motivation to learn is reduced. Therefore, there is a need to develop a recommender systems to solve the problem of information overload

* Tác giả liên hệ.

Email: huybx@ueh.edu.vn (Bùi Xuân Huy), tena@ueh.edu.vn (Nguyễn An Tế), minhst@neu.edu.vn (Trần Thị Song Minh).

Trích dẫn bài viết: Bùi Xuân Huy, Nguyễn An Tế, & Trần Thị Song Minh. (2023). Đánh giá các phương pháp xây dựng cộng đồng người học trên hệ thống tư vấn học tập trong môi trường đào tạo trực tuyến. *Tạp chí Nghiên cứu Kinh tế và Kinh doanh Châu Á*, 34(5), 16–26.

making it easy for learners to accurately and timely access the resources and services they are interested in. The principle of the learning recommender system is based on a community of learners with similar characteristics and good learning outcomes to give appropriate advice to a particular learner on decision-making issues such as: choosing learning methods, learning resources, study groups, and registering for subjects. The degree of precision in building a community of learners will affect the quality of advice. The content of this paper will present the methods of building a learner community on the learning recommend system in the online training environment as well as the issue of evaluating the effectiveness of the methods.

1. Mở đầu

Sự phát triển mạnh mẽ của công nghệ thông tin đã và đang đem lại những chuyển hóa to lớn trong ngành giáo dục ở mọi cấp độ. Việc giảng dạy và học tập không còn bị giới hạn trong những hình thức truyền thống mà còn dần dần tích hợp và chuyển sang nền tảng kỹ thuật số dựa trên công nghệ, kết nối Internet và các phương tiện truyền thông. Trong bối cảnh đó, xây dựng một hệ thống đào tạo trực tuyến hiệu quả là ứng dụng công nghệ để đem đến cho người học cơ hội học tập bất kể tuổi tác, giới tính, không gian và thời gian nhằm nâng cao chất lượng giáo dục (Rodrigues và cộng sự, 2019). Đây không chỉ là một nhu cầu mà còn là một thách thức đối với các cơ sở đào tạo. Mặt khác, đại dịch COVID-19 bùng phát dẫn đến việc tương tác trực tiếp trong hoạt động giảng dạy và học tập càng bị giới hạn do giãn cách hoặc cách ly xã hội. Dịch bệnh khiến gần 63 triệu trường học trên khắp thế giới, trong đó có Việt Nam phải chuyển đổi từ phương thức đào tạo truyền thống sang các khóa học trực tuyến (Valverde-Berrocoso và cộng sự, 2020). Đây là một “cú hích” khiến cho nhu cầu và thách thức nói trên ngày càng cấp thiết và sẽ không chấm dứt kể cả khi đại dịch kết thúc.

Một trong những điểm khác biệt rất quan trọng giữa đào tạo trực tuyến và đào tạo truyền thống là người học dễ bị lạc lõng, mất phương hướng trong môi trường học tập ảo rộng lớn vì không có giảng viên, không có bạn đồng học cụ thể, không biết nên học tập theo cách thức như thế nào, học nhóm với những ai, gặp khó khăn trong việc đánh giá mức độ cần thiết và thứ tự ưu tiên của các tài liệu học tập... Từ đó, dẫn đến tình trạng người học bị cô lập, thụ động và suy giảm động lực học tập. Vì vậy, cùng với sự phát triển mạnh mẽ của hình thức đào tạo trực tuyến là nhu cầu phát triển các hệ thống tư vấn để giải quyết vấn đề quá tải thông tin, giúp cho người học dễ dàng truy cập chính xác, kịp thời những tài nguyên, dịch vụ mình quan tâm hoặc xây dựng lộ trình học tập, xác định cách thức học tập phù hợp.

Các nghiên cứu về hệ thống tư vấn trong đào tạo trực tuyến có thể chia thành hai phương pháp chính: (1) Thích nghi cá nhân – xem xét tương tác giữa cá nhân với hệ thống, và (2) thích nghi cộng đồng – xem xét tương tác giữa các cá nhân với nhau. Đối với hướng thứ nhất, các nghiên cứu tập trung chủ yếu vào phương pháp tư vấn dựa trên nội dung (Content-Based Filtering). Người học được tư vấn các nội dung được những người học khác đánh giá cao hoặc có đặc điểm tương tự với nội dung họ đã chọn trong quá khứ. Ali và cộng sự (2021) giới thiệu hệ thống tư vấn khóa học dựa trên thuật toán phân cụm người học, áp dụng cho các khóa học trực tuyến đại chúng mở (Massive Open Online Courses – MOOC). Nghiên cứu của Wu và Wu (2020) đã phân tích những yếu tố nào có thể được sử

dụng để xác định lộ trình học tập và phát hiện ra sự khác biệt giữa những người học với các đặc điểm và nền tảng khác nhau. Wu và Wu (2020) đã phát triển một hệ thống tư vấn sử dụng tập dữ liệu điểm trong quá khứ có thể đưa ra các gợi ý cho người học về việc lựa chọn môn học. Kết quả nghiên cứu của Wu và Wu (2020) cho thấy tính khả thi của cách tiếp cận mới trong việc áp dụng dữ liệu lớn và công nghệ AI vào phân tích học tập và lựa chọn khóa học. Morsomme và Alferez (2019) đề xuất một mô hình tư vấn khóa học nhằm khuyến khích sinh viên đưa ra lựa chọn phù hợp với lộ trình học tập của họ. Đối với hướng thứ hai, các nghiên cứu tập trung chủ yếu vào hệ thống tư vấn dựa trên sự cộng tác (Collaborative Filtering). Phương pháp này định nghĩa cộng đồng của một người học, bao gồm những người có cùng đặc trưng với người học đó như: Thói quen học tập, điểm số... Một số nghiên cứu theo hướng này như nghiên cứu của Geng (2022) đề xuất một hệ thống lọc cộng tác cải tiến đưa ra các tư vấn dựa trên việc phân tích hành vi của người học. Joy và cộng sự (2021) đã triển khai giải pháp áp dụng bản thể học (Ontology) để giải quyết vấn đề khởi động nguội trong các hệ thống tư vấn. Thử nghiệm cho thấy rằng phần lớn người học hài lòng với các đề xuất được cá nhân hóa do giải pháp này cung cấp. Bên cạnh đó, nghiên cứu của Yassine và cộng sự (2021) đề xuất một hệ thống tư vấn thông minh dựa trên thuật toán học không giám sát và các thông tin nhân khẩu học của người học. Và những tư vấn được đưa ra dựa trên các ý kiến phản hồi hay đánh giá từ cộng đồng này. Các nghiên cứu hầu hết tập trung vào việc phát triển và đánh giá các thuật toán tư vấn. Ở Việt Nam, trong những năm gần đây, có rất nhiều nghiên cứu về lĩnh vực đào tạo trực tuyến, nhưng hầu hết đều không liên quan đến hệ thống tư vấn. Bên cạnh đó, các trường đại học đang xây dựng hệ thống đào tạo trực tuyến theo hướng quản lý học tập và Blended Learning – là mô hình giảng dạy kết hợp hình thức trực tuyến và trực tiếp. Trường Đại học Kinh tế Quốc dân là một trong số rất ít trường đã xây dựng hệ thống đào tạo trực tuyến hoàn toàn. Hệ thống đào tạo trực tuyến của trường bắt đầu hoạt động từ năm 2018 với số lượng người dùng đến hơn 13.000 nhưng cũng chưa nghiên cứu hệ thống tư vấn học tập.

Cả hai phương pháp tư vấn dựa trên nội dung và tư vấn dựa trên sự cộng tác đều quan tâm đến việc quản trị hồ sơ đặc trưng của người học bao gồm các bước: Biểu diễn, khởi tạo, cập nhật và khai thác. Hồ sơ đặc trưng của người học ghi nhận những đặc trưng của mỗi người như: Yếu tố nhân khẩu học, thói quen, nhu cầu học tập, điểm số... Việc đưa ra tư vấn hoặc sắp xếp người học vào cộng đồng phù hợp đều dựa trên hồ sơ đặc trưng của người học. Mức độ chính xác của việc xây dựng cộng đồng người học sẽ ảnh hưởng đến chất lượng tư vấn. Những nội dung về quản trị hồ sơ đặc trưng của người học và mô hình hệ thống tư vấn học tập đã được nhóm tác giả trình bày trong các nghiên cứu trước đây (Bùi Xuân Huy và cộng sự, 2020, 2021). Trong bài báo này, nhóm tác giả sẽ trình bày về các phương pháp xây dựng cộng đồng người học trên hệ thống tư vấn học tập trong môi trường đào tạo trực tuyến cũng như vấn đề đánh giá hiệu quả của các phương pháp.

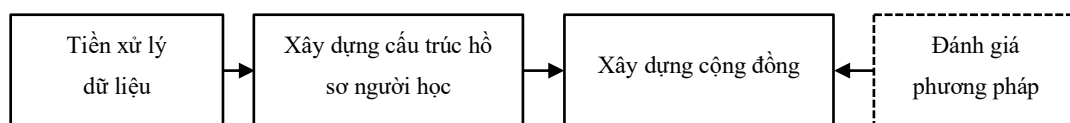
Cấu trúc bài báo gồm 6 phần chính: Phần một giới thiệu bối cảnh nghiên cứu và mục tiêu nghiên cứu; phần hai trình bày về mô tả dữ liệu và quy trình thực nghiệm; phần ba trình bày về xây dựng cấu trúc hồ sơ đặc trưng của người học; phần bốn trình bày về các phương pháp xây dựng cộng đồng người học; phần năm trình bày về kết quả thực nghiệm; phần sáu là kết luận và hướng phát triển tiếp theo.

2. Mô tả dữ liệu và quy trình thực nghiệm

Nhóm tác giả sử dụng dữ liệu của hệ thống đào tạo trực tuyến tại Trường Đại học Kinh tế Quốc dân để tiến hành thực nghiệm xây dựng cấu trúc hồ sơ đặc trưng của người học. Hệ thống đào tạo trực

tuyển của Trường Đại học Kinh tế Quốc dân đi vào hoạt động từ năm 2018, và đến năm 2022 có 6 chuyên ngành là: Kế toán, Luật, Luật Kinh tế, Quản trị Kinh doanh, Tài chính Doanh nghiệp, và Ngân hàng với trên 13.000 người học. Dữ liệu thu thập được bao gồm điểm cuối khóa và dữ liệu nhân khẩu học của tất cả người học lấy từ phân hệ quản lý giáo vụ, và dữ liệu phân theo từng lớp môn học lấy từ hệ thống đào tạo trực tuyến bao gồm tập tin lưu trữ các hoạt động (Log Files) và tập tin lưu trữ điểm quá trình (Grade Report File).

Quy trình thực nghiệm của nhóm tác giả bao gồm các giai đoạn: Tiền xử lý dữ liệu, xây dựng cấu trúc hồ sơ người học, xây dựng cộng đồng, và đánh giá phương pháp xây dựng cộng đồng.



Hình 1. Quy trình thực nghiệm

3. Xây dựng cấu trúc hồ sơ đặc trưng của người học

Do người học được yêu cầu chủ động nhập các thông tin về nhân khẩu học lên hệ thống nên các thông tin này hầu hết bị bỏ trống hoặc chứa dữ liệu vô nghĩa, ngoại trừ các thông tin về ngày sinh, quê quán, giới tính, ngành học. Ngoài ra, hệ thống đào tạo trực tuyến chỉ lưu trữ thời điểm bắt đầu truy cập nên không tính được các đặc trưng dùng để xác định thói quen học tập như: Tổng thời gian dành cho môn học, dành cho từng chương hoặc chủ đề cụ thể của môn học; số lần vào học; thời điểm tham gia học tập; thời gian trung bình cho mỗi lần vào học; tổng thời gian cho mỗi hoạt động học tập như thảo luận online, tham gia diễn đàn... Kết quả khảo sát cấu trúc tập tin lưu trữ điểm quá trình cho thấy mỗi lớp môn học có các hoạt động như sau: Bài kiểm tra, bài luyện tập, bài kiểm tra không tính điểm (có thể được giảng viên đánh giá hoặc không), hoạt động trên diễn đàn (có thể được giảng viên đánh giá hoặc không), bài tự luận, điểm chuyên cần, điểm quá trình. Vì số lượng bài kiểm tra, bài luyện tập và các bài kiểm tra không tính điểm và bài tự luận cũng như số diễn đàn của các lớp môn học không giống nhau nên nhóm tác giả không sử dụng con số tuyệt đối để đánh giá mức độ hoàn thành của người học mà đề xuất sử dụng tỷ lệ hoàn thành. Ví dụ: Lớp môn học có 10 bài kiểm tra, người học hoàn thành 5 bài thì tỷ lệ hoàn thành là 50%. Tỷ lệ hoàn thành bài kiểm tra, bài tự luận, bài luyện tập cho biết mức độ chuyên cần của người học qua việc thực hiện các hoạt động bắt buộc/có tính điểm trong lớp học. Ngoài ra, người học thường có thói quen chỉ ưu tiên thực hiện các hoạt động bắt buộc/có tính điểm. Do đó, số lần thực hiện các bài kiểm tra không tính điểm, tỷ lệ các bài kiểm tra không tính điểm được giảng viên đánh giá, số lần tương tác trên diễn đàn và tỷ lệ các bài tham gia diễn đàn được giảng viên đánh giá phản ánh mức độ tự giác đầu tư cho môn học. Hiển nhiên một người học chăm chỉ làm các bài kiểm tra không tính điểm và tích cực tương tác trên diễn đàn sẽ có kết quả học tập khả quan hơn một người học chỉ chú trọng làm bài kiểm tra bắt buộc/có tính điểm hoặc chỉ vào diễn đàn và xem.

Căn cứ vào phân loại của Brusilovsky và Milán (2007) và kết quả phân tích dữ liệu thực tế của hệ thống đào tạo trực tuyến tại trường Đại học Kinh tế Quốc dân, nhóm tác giả đề xuất cấu trúc hồ sơ đặc trưng của người học trong hệ thống tư vấn học tập trong môi trường đào tạo trực tuyến gồm ba

nhóm: (1) Đặc trưng về thông tin nhân khẩu học, (2) đặc trưng về kiến thức, và (3) đặc trưng về thói quen học tập.

Bảng 1.

Cấu trúc hồ sơ đặc trưng của người học

Các nhóm đặc trưng	Diễn giải
Các đặc trưng về thông tin nhân khẩu học	- Họ tên - Độ tuổi - Quê quán - Giới tính - Ngành học
Các đặc trưng về kiến thức	- Các môn học đã hoàn thành. - Các lớp đã tham gia. - Kết quả học tập của người học như: Điểm chuyên cần, điểm kiểm tra (giữa kỳ), điểm kết thúc học phần.
Các đặc trưng về thói quen học tập	- Số lần tương tác trên các diễn đàn. - Số lần thực hiện các bài kiểm tra không tính điểm. - Tỷ lệ các bài kiểm tra không tính điểm được giảng viên đánh giá. - Tỷ lệ các bài tham gia diễn đàn được giảng viên đánh giá. - Tỷ lệ hoàn thành các bài kiểm tra. - Tỷ lệ hoàn thành các bài tự luận. - Tỷ lệ hoàn thành các bài luyện tập.

Hoạt động quản trị hồ sơ đặc trưng của người học bao gồm: (1) Biểu diễn, (2) khởi tạo, (3) cập nhật, và (4) khai thác đã được nhóm tác giả trình bày trong nghiên cứu trước (Bùi Xuân Huy và cộng sự, 2021).

4. Các phương pháp xây dựng cộng đồng người học trên hệ thống tư vấn học tập trong môi trường đào tạo trực tuyến

Cộng đồng người học được hình thành từ những người học có đặc trưng tương tự nhau. Cộng đồng cùng với hồ sơ đặc trưng của người học sẽ hợp thành hai cơ sở quan trọng cho các hình thức tư vấn học tập. Về nguyên tắc, mỗi đặc trưng trong hồ sơ đặc trưng của người học đều có thể được dùng làm tiêu chí thành lập cộng đồng và chính là đặc trưng của cộng đồng. Hiển nhiên, nhiều đặc trưng khác nhau cũng có thể được kết hợp tạo thành tiêu chí thành lập cộng đồng. Đối với những tiêu chí định tính như đặc trưng nhân khẩu học và kiến thức thì việc thành lập cộng đồng có thể dựa trên sự so sánh trực tiếp các giá trị đặc trưng trong hồ sơ của người học. Ví dụ: Xét tiêu chí kết quả học tập môn Cơ sở dữ liệu, có thể xây dựng được cộng đồng những người đạt điểm A, điểm B, điểm C... môn học này. Đối với các tiêu chí định tính như vậy, nhóm tác giả đề xuất sử dụng các phương pháp

trích lọc dữ liệu bằng các câu lệnh truy vấn trên cơ sở dữ liệu để thành lập cộng đồng người học. Đối với các tiêu chí định lượng, ví dụ: Thành lập cộng đồng những người học có thói quen học tập tương tự nhau (tỷ lệ hoàn thành các bài tự luận từ 20% đến 30%) không thể dùng phương pháp so sánh trực tiếp các giá trị, nên nhóm tác giả đề xuất dùng phương pháp phân cụm dữ liệu để thành lập cộng đồng người học. Phương pháp trích lọc dữ liệu bằng các câu lệnh truy vấn đối với tiêu chí định tính là công việc đơn giản không cần đánh giá độ chính xác. Do đó, trong bài báo này, nhóm tác giả chỉ trình bày các phương pháp phân cụm dữ liệu để thành lập cộng đồng người học.

Có rất nhiều thuật toán phân cụm với cách tiếp cận khác nhau, tuy nhiên, theo van der Aalst (2011), các thuật toán phân cụm phổ biến thường được sử dụng K-means và thuật toán AHC (Agglomerative Hierarchical Clustering).

4.1. Thuật toán K-means

Ý tưởng của thuật toán K-means là phân nhóm tập dữ liệu người học vào K cụm (với K là số được xác định trước) sao cho tổng bình phương khoảng cách giữa các người học trong một cụm đến trọng tâm (Centroid) của cụm đó là nhỏ nhất.

$$SSE = \sum_{i=1}^k \sum_{x \in C_i} d(x, c_i)^2 \quad (1)$$

Trong đó, trọng tâm của mỗi cụm được tính như sau:

$$C = \frac{1}{n} \sum_{i=1}^n x_i \quad (2)$$

với n: Số người học; và x_i : Thuộc tính của người học thứ i.

Khoảng cách giữa một người học trong cụm đến trọng tâm của cụm đó được tính bằng một trong các hàm khoảng cách Manhattan, Euclidean hoặc Cosine...

$$\text{Hàm khoảng cách Manhattan: } d(x, y) = \sum_{i=1}^n |x_i - y_i| \quad (3)$$

$$\text{Hàm khoảng cách Euclidean: } d(x, y) = (\sum_{i=1}^n (x_i - y_i)^2)^{1/2} \quad (4)$$

$$\text{Hàm khoảng cách Cosine: } d(x, y) = \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n y_i^2}} \quad (5)$$

Thuật toán K-Means được thực hiện như sau:

- *Bước 1*: Chọn ngẫu nhiên k trọng tâm từ tập dữ liệu người học
- *Bước 2*: Với mỗi người học x, đưa x vào cụm có trọng tâm gần với x nhất. Nếu phân hoạch không đổi so với lần lặp trước thì dừng.
- *Bước 3*: Xác định lại k trọng tâm và quay lại Bước 2.

4.2. Thuật toán AHC

Thuật toán AHC được thực hiện như sau:

- *Bước 1*: Khởi tạo n cụm từ tập dữ liệu người học, mỗi cụm chứa 1 phần tử.

- *Bước 2*: Dựa trên công thức tính khoảng cách giữa hai cụm (Single-Link) gộp chung hai cụm gần nhau nhất thành một cụm. Công thức Single-Link tìm khoảng cách ngắn nhất giữa hai người học bất kỳ của hai cụm.

$$D(C_i, C_j) = \min_{x_i \in C_i, x_j \in C_j} d(x_i, x_j) \quad (6)$$

- *Bước 3*: Lặp lại Bước 2 cho đến khi gộp n phần tử vào một cụm duy nhất hoặc khi có đủ số cụm mong muốn.

4.3. Phương pháp đánh giá chất lượng của thuật toán phân cụm trên tập dữ liệu người học

Hiện nay, chất lượng của thuật toán phân cụm được dựa trên ba cách tiếp cận: (1) Đánh giá trong (Internal Evaluation), (2) đánh giá ngoài (External Evaluation), và (3) đánh giá tương quan (Relative Evaluation). Trong đó, “đánh giá trong” là cách tiếp cận dựa trên chính dữ liệu được phân cụm với tiêu chí xác định thuật toán phân cụm tốt là thuật toán ra các cụm mà các phần tử trong cụm có độ tương tự càng lớn càng tốt, và độ tương tự giữa các phần tử khác cụm càng nhỏ càng tốt; “đánh giá ngoài” là cách tiếp cận dựa vào tập dữ liệu mẫu đã được phân cụm từ trước đó; cuối cùng, “đánh giá tương quan” là cách tiếp cận dựa vào việc so sánh kết quả phân cụm với các kết quả phân cụm khác được sinh ra bởi cùng thuật toán nhưng với các giá trị tham số khác nhau. Nhóm tác giả sử dụng cách tiếp cận đánh giá trong để đánh giá chất lượng của thuật toán phân cụm dựa trên tập dữ liệu người học bởi vì chưa có tập dữ liệu chuẩn nào trong lĩnh vực tư vấn học tập trên môi trường đào tạo trực tuyến đã được phân cụm trước đây. Các độ đo phổ biến của cách tiếp cận “đánh giá trong” là độ đo bóng (Silhouette).

$$\text{Độ đo bóng: } s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (7)$$

Trong đó: $a(i)$: Khoảng cách trung bình từ người học i đến tất cả các người học cùng cụm với i ;

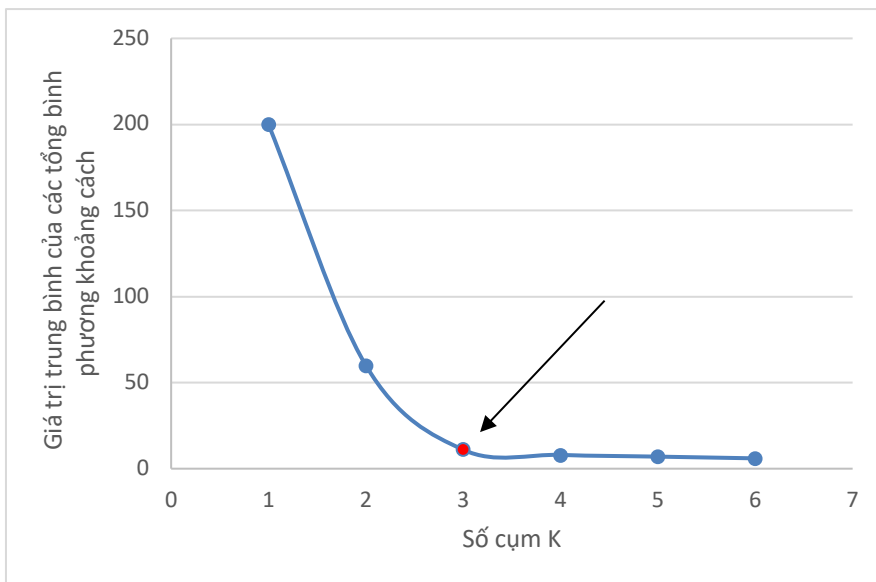
$b(i)$: Khoảng cách trung bình ngắn nhất từ người học i đến bất kỳ cụm nào không chứa i ;

$s(i)$ nằm trong đoạn $[-1; 1]$; $s(i)$ càng lớn thì kết quả phân cụm cho người học i càng chính xác.

Kaufman và Rousseeuw (2009) đã giới thiệu cách tính hệ số độ đo bóng để đánh giá kết quả phân cụm như sau: $SC = \max_k \bar{s}(k)$, với $\bar{s}(k)$ là giá trị trung bình của các $s(i)$; và k là số cụm. SC càng lớn thì kết quả phân cụm càng tốt.

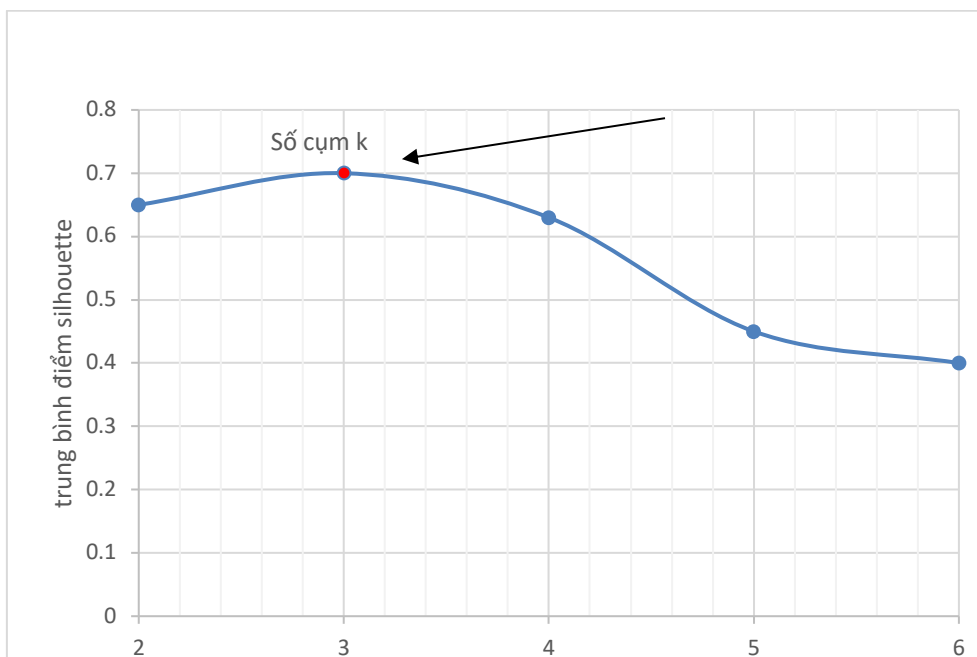
4.4. Phương pháp xác định số cụm K tối ưu cho phương pháp K -means

Số cụm K tối ưu có thể được xác định theo phương pháp khuỷu tay (Elbow Method) là số cụm đáp ứng được điều kiện giá trị trung bình của các tổng bình phương các sai số (Sum of Squares Errors – SSE) bắt đầu suy giảm tuyến tính (Zhu & Li, 2023).



Hình 2. Phương pháp khuỷu tay xác định số cụm K tốt nhất

Ngoài ra, số cụm K tối ưu cũng có thể được xác định bằng phương pháp phân tích chỉ số Silhouette (Silhouette Analysis) (JothiPrabha và cộng sự, 2023). Trong phương pháp này, số cụm K được chọn sao cho giá trị SC là lớn nhất.



Hình 3. Phương pháp Silhouette Analysis xác định số cụm K tốt nhất

5. Kết quả thực nghiệm

Các dữ liệu định lượng trong hồ sơ đặc trưng của mỗi người học mô tả kiến thức và thói quen học tập của người học đó theo từng lớp môn học (Bảng 1) nên thực hiện phân cụm người học trên toàn bộ dữ liệu của hệ thống mà không phân theo từng môn học cụ thể là không có ý nghĩa. Vì vậy, nhóm tác giả tiến hành phân cụm dữ liệu người học theo từng môn học. Bảng 2 cho biết kết quả so sánh điểm Silhouette của hai phương pháp K-means và AHC sau khi phân cụm người học theo cùng một môn học với số lượng người học là 706 người.

Bảng 2.

So sánh điểm Silhouette của thuật toán K-means và AHC trên tập dữ liệu của một môn học trong một học kỳ

Số cụm	2	3	4	5	6	7	8	9	10
K-means	0,448	0,365	0,399	0,421	0,418	0,379	0,399	0,407	0,396
AHC	0,615	0,601	0,575	0,720	0,431	0,365	0,365	0,517	0,517

Kết quả so sánh cho thấy thuật toán AHC cho kết quả đáng tin cậy hơn. Theo phương pháp Silhouette Analysis, thì số cụm tốt nhất khi sử dụng phương pháp AHC là 5 cụm, trong khi sử dụng phương pháp K-means số cụm tốt nhất là 2 cụm. Kiểm chứng bằng phương pháp khuy tay cho thấy số cụm tốt nhất khi dùng thuật toán K-means là 4.

Bảng 3.

Kết quả của phương pháp khuy tay để xác định số K

Số cụm	2	3	4	5	6	7	8	9
SSE Score	4,600	3,700	2,900	2,500	2,250	2,000	1,800	1,700

Tuy nhiên, phương pháp khuy tay chọn số K tốt nhất theo nguyên lý là điểm mà tại đó đồ thị bắt đầu suy giảm tuyến tính. Nhóm tác giả cho rằng, việc xác định theo nguyên lý như vậy thiên về cảm tính của người phân tích, trong khi xác định số K theo phương pháp Silhouette Analysis hợp lý hơn vì dễ dàng tính được vị trí mà tại đó điểm Silhouette là lớn nhất. Từ kết quả thực nghiệm cho thấy, sử dụng thuật toán AHC là phù hợp đối với vấn đề phân cụm người học theo các tiêu chí định lượng.

6. Kết luận

Nhóm tác giả đã đề xuất cấu trúc hồ sơ đặc trưng của người học trong môi trường đào tạo trực tuyến và phương pháp xây dựng cộng đồng người học theo các tiêu chí định tính và định lượng, bên cạnh đó, cũng chỉ ra cả phương pháp đánh giá hiệu quả của các thuật toán được sử dụng. Từ kết quả này, giai đoạn nghiên cứu tiếp theo là xây dựng các phân hệ tư vấn học tập trong môi trường đào tạo trực tuyến dựa trên hồ sơ đặc trưng và các cộng đồng của người học.

Lời cảm ơn

Nghiên cứu này được tài trợ bởi trường Đại học Kinh tế TP. Hồ Chí Minh, trong khuôn khổ đề tài nghiên cứu khoa học cấp trường, Mã số: CS-2021-52.

Tài liệu tham khảo

- Ali, H. A., Mohamed, C., Abdelhamid, B., & El Alami, T. (2021). *A course recommendation system for MOOCs based on online learning*. Paper presented at the 2021 XI International Conference on Virtual Campus (JICV), Salamanca, Spain.
- Brusilovsky, P., & Millán, E. (2007). User models for adaptive hypermedia and adaptive educational systems. In Brusilovsky, P., Kobsa, A., Nejdl, W. (eds), *The Adaptive Web. Lecture Notes in Computer Science* (pp. 3–53). Berlin, Heidelberg: Springer.
- Bùi Xuân Huy, Nguyễn An Tế, & Trần Thị Song Minh. (2020). Mô hình tư vấn học tập trong đào tạo trực tuyến dựa trên cộng đồng người học đa tiêu chí. *Tạp chí Nghiên cứu Kinh tế và Kinh doanh Châu Á*, 31(11), 21–35.
- Bùi Xuân Huy, Nguyễn An Tế, & Trần Thị Song Minh. (2021). Phát triển mô hình tư vấn học tập trong đào tạo trực tuyến dựa trên cộng đồng người học đa tiêu chí. *Tạp chí Nghiên cứu Kinh tế và Kinh doanh Châu Á*, 32(7), 45–64.
- Geng, L. (2022). The recommendation system of innovation and entrepreneurship education resources in universities based on improved collaborative filtering model. *Computational Intelligence Neuroscience*, 2022. doi: 10.1155/2022/7228833
- JothiPrabha, A., Bhargavi, R., & Rani, B. D. (2023). Prediction of dyslexia severity levels from fixation and saccadic eye movement using machine learning. *Biomedical Signal Processing Control*, 79(1), 104094.
- Joy, J., Raj, N. S., & Renumol, V. G. (2021). Ontology-based E-learning content recommender system for addressing the pure cold-start problem. *ACM Journal of Data Information Quality*, 13(3), 1–27.
- Kaufman, L., & Rousseeuw, P. J. (2009). *Finding Groups in Data: An Introduction to Cluster Analysis*. New Jersey: John Wiley & Sons.
- Morsomme, R., & Alferez, S. V. (2019). *Content-based course recommender system for liberal arts education*. Proceedings of The 12th International Conference on Educational Data Mining (EDM 2019), July 2–5, 2019, Montréal, Canada.
- Rodrigues, H., Almeida, F., Figueiredo, V., & Lopes, S. L. (2019). Tracking e-learning through published papers: A systematic review. *Computers & Education*, 136, 87–98.
- Valverde-Berrocso, J., Garrido-Arroyo, M. d. C., Burgos-Videla, C., & Morales-Cevallos, M. B. J. S. (2020). Trends in educational research about e-learning: A systematic literature review (2009–2018). *Sustainability*, 12(12), 5153.
- van der Aalst, W. (2011). *Process Mining: Discovery, Conformance and Enhancement of Business Processes*. Berlin: Springer.

- Wu, Y. H., & Wu, E. H. (2020). *AI-based college course selection recommendation system: Performance prediction and curriculum suggestion*. Paper presented at The 2020 International Symposium on Computer, Consumer and Control (IS3C), 13–16 November, 2020, Taichung City, Taiwan.
- Yassine, A., Mohamed, L., & Al Achhab, M. (2021). Intelligent recommender system based on unsupervised machine learning and demographic attributes. *Simulation Modelling Practice Theory*, 107, 102198.
- Zhu, X., & Li, L. (2023). Estimating the number of clusters in high-dimensional large datasets. *International Journal of Data Warehousing Mining*, 19(2), 1–14.