

INFERRING PRE-SERVICE TEACHERS' LATENT DIGITAL COMPETENCE FROM DIGITAL LEARNING TRACES: A STUDY COMBINING HIDDEN MARKOV MODELS AND SELF - REGULATED LEARNING THEORY

Nguyen Hoai Nam^{1,2}, Nguyen Ha Nam^{*3}

* Corresponding author
Email: namnh@vnu.edu.vn

¹ VNU, University of Education
Vietnam National University, Hanoi
144 Xuan Thuy, Cau Giay ward, Hanoi, Vietnam

² Email: nam.moet@gmail.com
Ministry of Education and Training
35 Dai Co Viet, Bach Mai ward, Hanoi, Vietnam

³ Email: namnh@vnu.edu.vn
Vietnam National University, Hanoi
Hoa Lac, Hanoi, Vietnam

Received: 10/9/2025
Revised: 25/10/2025
Accepted: 05/11/2025
Published: 20/12/2025

Abstract: This study proposes an approach that integrates hidden Markov model and self-regulated learning theory to infer the latent states of digital competence among pre-service teachers based on behavioral data collected from a learning management system. Data were gathered from two distinct 15-week courses: "Introduction to Informatics" (216 students, 498,050 events) and "Application of Information Technology in Teaching" (51 students, 16,878 events). The results demonstrate that Hidden Markov Models effectively identified distinct latent learning states, which were meaningfully interpreted as different levels of digital competence through the lens of the three-phase self-regulated learning framework. This research contributes both methodologically, by introducing an integrated hidden Markov model - self-regulated learning analytical pipeline; and practically, by providing an objective, continuous, and behavior-based assessment tool for evaluating competence in higher education.

Keywords: Digital competence, hidden Markov model, self-regulated learning, learning analytics, learning management system, competence assessment.

SUY LUẬN NĂNG LỰC SỐ ẨN CỦA SINH VIÊN SƯ PHẠM TỪ DẤU VẾT HỌC TẬP SỐ: NGHIÊN CỨU VỚI MÔ HÌNH MARKOV ẨN VÀ KHUNG LÝ THUYẾT HỌC TẬP TỰ ĐIỀU CHỈNH

Nguyễn Hoài Nam^{1,2}, Nguyễn Hà Nam^{*3}

* Tác giả liên hệ
Email: namnh@vnu.edu.vn

¹ Trường Đại học Giáo dục, Đại học Quốc gia Hà Nội
144 Xuân Thủy, phường Cầu Giấy,
Hà Nội, Việt Nam

² Email: nam.moet@gmail.com
Bộ Giáo dục và Đào tạo
35 Đại Cồ Việt, phường Bạch Mai,
Hà Nội, Việt Nam

³ Email: namnh@vnu.edu.vn
Đại học Quốc gia Hà Nội
Xã Hòa Lạc, Hà Nội, Việt Nam

Nhận bài: 10/9/2025
Chỉnh sửa xong: 25/10/2025
Chấp nhận đăng: 05/11/2025
Xuất bản: 20/12/2025

Tóm tắt: Nghiên cứu này đề xuất một phương pháp tiếp cận kết hợp giữa mô hình Markov ẩn và Lý thuyết học tập tự điều chỉnh để suy luận trạng thái năng lực số của sinh viên sư phạm dựa trên dữ liệu hành vi thu thập từ Hệ thống Quản lý học tập Moodle (LMS). Dữ liệu được thu thập từ hai học phần: "Tin học đại cương" (216 sinh viên, 498.050 sự kiện) và "Ứng dụng Công nghệ thông tin trong dạy học" (51 sinh viên, 16.878 sự kiện) trong 15 tuần. Kết quả cho thấy, mô hình Markov ẩn có khả năng phân biệt các trạng thái học tập tiềm ẩn khác biệt, tương ứng với các mức độ năng lực số khác nhau, được diễn giải thông qua lăng kính các giai đoạn học tập tự điều chỉnh. Nghiên cứu đóng góp cả về phương pháp luận (tích hợp mô hình Markov ẩn - học tập tự điều chỉnh) và thực tiễn (cung cấp một công cụ đánh giá khách quan, liên tục dựa trên dữ liệu hành vi) cho lĩnh vực đánh giá năng lực số trong giáo dục đại học.

Từ khóa: Năng lực số, mô hình Markov ẩn, học tập tự điều chỉnh, phân tích học tập, hệ thống quản lý học tập, đánh giá năng lực.

1. Đặt vấn đề

Năng lực số không chỉ đơn thuần là kỹ năng sử dụng công nghệ mà là một tập hợp tích hợp của kiến thức, kỹ năng và thái độ cho phép cá nhân sử dụng công nghệ số một cách tự tin, có phê phán và có trách nhiệm. Các khung năng lực số như DigComp cung

cấp cấu trúc thành phần và mức độ thành thực để quy chiếu đánh giá, thiết kế học tập và bồi dưỡng năng lực (Carretero và cộng sự, 2017; Vuorikari và cộng sự, 2022). Ở Việt Nam, định hướng thúc đẩy đánh giá - phát triển năng lực số trong giáo dục đã được nêu rõ (MOET, 2025). Một số nghiên cứu đã đề

xuất khung/thang đo và khảo sát hiện trạng năng lực số ở người học (Lê Anh Vinh và cộng sự, 2021; Le Thai và cộng sự, 2022) nhưng phần lớn dựa trên tự báo cáo nên dễ chịu ảnh hưởng của thiên lệch nhận thức và mong muốn của xã hội, do đó chưa phản ánh đầy đủ hành vi học tập thực tế trong môi trường số (Sillat và cộng sự, 2021).

Các nghiên cứu gần đây đã khám phá tiềm năng của học máy trong nghiên cứu hành vi học tập theo thời gian. Các mô hình phân tích chuỗi thời gian nổi lên như một giải pháp đầy hứa hẹn để mô hình hóa tiến trình học tập (Baker và Yacef, 2009; Siemens và Baker, 2012). Trong đó, mô hình Markov ẩn (Hidden Markov Model) tỏ ra phù hợp để mô hình hóa các trạng thái ẩn (như sự tiến bộ về năng lực) dựa trên các quan sát là chuỗi hành vi tương tác trên LMS (Xing và Du, 2019). Ưu điểm của mô hình Markov ẩn là khả năng nắm bắt tính tuần tự và sự tiến triển theo thời gian của hành vi, đồng thời cung cấp một mô hình có thể diễn giải được so với các mô hình “hộp đen” phức tạp (Rabiner, 1989). Tuy nhiên, việc diễn giải ý nghĩa sự phạm của các trạng thái ẩn này vẫn là một thách thức lớn.

Lí thuyết Học tập tự điều chỉnh (Self-Regulated Learning theory) cung cấp một khung lí thuyết vững chắc để hiểu cách người học chủ động điều khiển quá trình học tập của mình thông qua ba giai đoạn chính: 1) Định hướng và đặt mục tiêu (Forethought); 2) Thực hiện và kiểm soát trong khi thực hiện nhiệm vụ (Performance); 3) Đánh giá và phản ánh sau khi hoàn thành (Self-Reflection) (Zimmerman, 2002; Pintrich, 2004). Các hành vi trên LMS (như xem tài liệu sớm, nộp bài đúng hạn, tham gia thảo luận, xem lại phản hồi) có thể được xem là những biểu hiện khách quan, có thể diễn giải được bằng các chiến lược học tập tự điều chỉnh (Winne, 2014; Matcha và cộng sự, 2019).

Nghiên cứu này đề xuất một khung diễn giải tích hợp, sử dụng lí thuyết học tập tự điều chỉnh để kết nối các trạng thái ẩn do mô hình Markov ẩn phát hiện với các chiến lược học tập cụ thể, từ đó suy luận về mức độ năng lực số của sinh viên sự phạm thông qua quy chiếu DigComp. Cụ thể, nghiên cứu này thực hiện: 1) Chuẩn hoá log sự kiện thành 10 nhóm hành vi theo chức năng LMS; 2) Huấn luyện và lựa chọn mô hình Markov ẩn theo tiêu chí thông tin; 3) Mô tả đặc trưng phát xạ, quỹ đạo chuyển dịch và độ ổn định của các trạng thái; 4) Diễn giải sự phạm theo lí thuyết học tập tự điều chỉnh, liên hệ với các cấu phần năng lực số của DigComp. Nghiên cứu kì vọng

đóng góp được thước đo bổ trợ, khách quan dựa trên hành vi cho đánh giá năng lực số, đồng thời gợi ý thiết kế can thiệp nhằm kéo dài lưu trú ở các trạng thái học tập “tích cực” của sinh viên sự phạm trong bối cảnh đào tạo giáo viên tại Việt Nam.

2. Phương pháp nghiên cứu

2.1. Bối cảnh và thiết kế nghiên cứu

Nghiên cứu sử dụng phương pháp hỗn hợp, kết hợp phân tích định lượng dữ liệu log LMS và diễn giải định tính các trạng thái ẩn dựa trên khung lí thuyết học tập tự điều chỉnh. Dữ liệu được thu thập từ hệ thống LMS của Trường Đại học Sư phạm Hà Nội trong học kì 1 năm học 2024 -2025 từ hai học phần bắt buộc: 1) “Tin học đại cương” là học phần nền tảng về công nghệ thông tin (mã THUD_5828, 216 sinh viên, 498.050 sự kiện); 2) “Ứng dụng CNTT trong dạy học” là học phần nghiệp vụ tập trung vào tích hợp công nghệ vào phương pháp giảng dạy (mã UDCNTT_5787, 51 sinh viên, 16.878 sự kiện). Cả hai học phần đều kéo dài 15 tuần. Toàn bộ hoạt động của sinh viên khi tham gia học hai học phần trên hệ thống LMS được tự động ghi lại vào nhật kí của hệ thống (Log system). Quá trình thu thập dữ liệu tuân thủ nguyên tắc tối thiểu xâm phạm và đảm bảo ẩn danh danh tính người học.

Khung nghiên cứu tổng thể đã được xây dựng như tại Hình 1, trong đó các khung lí thuyết về học tập tự điều chỉnh và năng lực số được vận dụng làm lí thuyết nền trong suốt các bước của nghiên cứu, giúp làm rõ mối tương quan giữa dữ liệu, phương pháp, lí giải và kiểm chứng kết quả. Đầu vào của toàn bộ luồng nghiên cứu là bộ dữ liệu thô về hành vi của từng sự kiện theo thời gian (bao gồm các thông tin: user_id, course_id, timestamp, component, action/eventname...) từ nhật kí hệ thống của LMS. Một bộ tiền xử lí dữ liệu và trích xuất đặc trưng dữ liệu được xây dựng để chuẩn hóa dữ liệu từ hàng nghìn sự kiện rời rạc thành các tập quan sát là các chuỗi hành vi mang ý nghĩa sự phạm dựa trên các khung lí thuyết nền (như CONTENT_VIEW, QUIZ_ATTEMPT...), làm đầu vào cho việc huấn luyện mô hình học máy. Tiếp theo, mô hình học máy được huấn luyện với tập dữ liệu đã chuẩn hóa nhằm phát hiện ra các trạng thái ẩn từ các tập quan sát hành vi. Việc lựa chọn mô hình học máy tối ưu nào (từ đó cho biết số lượng trạng thái ẩn của các quan sát) được đánh giá dựa trên kết hợp các tiêu chí: 1) Tiêu chí thông tin thống kê AIC (Akaike Information Criterion) và BIC (Bayesian Information Criterion) để tránh quá khớp dữ liệu (Overfitting); 2) Tính ổn định của chuỗi trạng

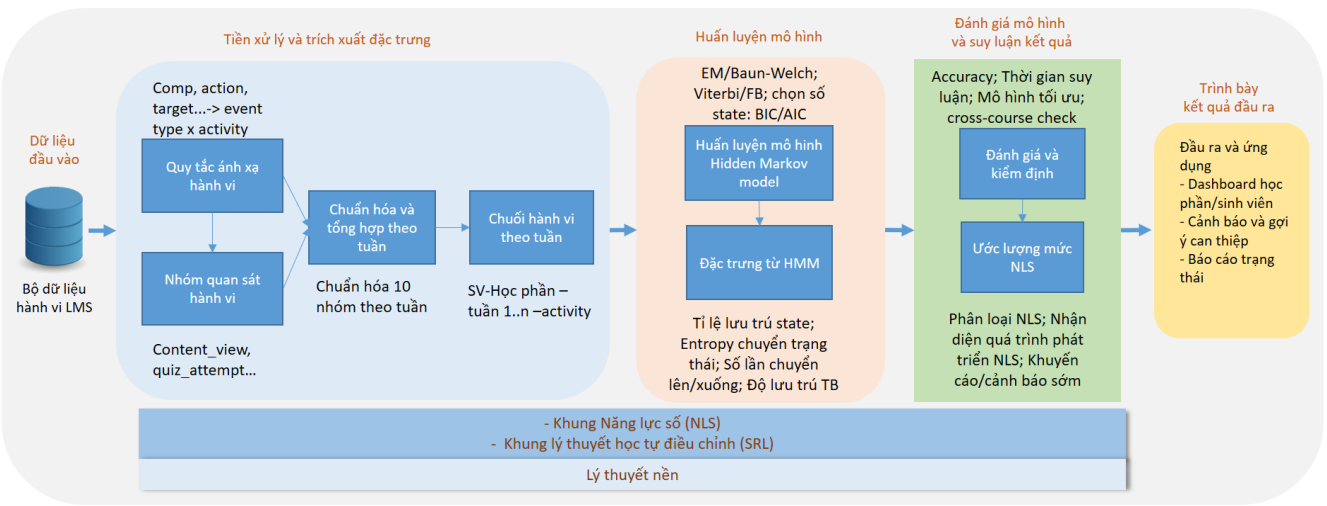
thái được giải mã bằng thuật toán Viterbi; 3) Tính diễn giải được về mặt sư phạm dựa trên lí thuyết nền. Cuối cùng, dữ liệu đầu ra được trình bày bằng dashboard báo cáo và các khuyến cáo tương ứng.

2.2. Tiền xử lí và trích xuất đặc trưng dữ liệu

Trong lĩnh vực Phân tích học tập và Khai phá dữ liệu giáo dục, dữ liệu vết học tập (learning traces) từ LMS được công nhận rộng rãi như một nguồn dữ liệu khách quan, giàu thông tin để suy luận về hành vi và chiến lược học tập của người học (Siemens và Baker, 2012). Tuy nhiên, dữ liệu log thô từ LMS thường là các sự kiện chi tiết, mức độ thấp (Ví dụ: course_viewed, assign_submitted), mang tính rời rạc, tạo nên một tập dữ liệu thừa và gây khó khăn cho diễn giải trực tiếp ý nghĩa sư phạm cũng như xây dựng các mô hình dự đoán ổn định (Blikstein và Worsley, 2016). Để giải quyết vấn đề này, nghiên cứu kế thừa các phương pháp được đề xuất trong lĩnh vực phân

tích học tập (Matcha và cộng sự, 2019) bằng cách tổng hợp và ánh xạ các sự kiện thô thành các nhóm hành vi có ý nghĩa dựa trên chức năng học tập. Cách tiếp cận này giúp giảm nhiều dữ liệu, cải thiện độ ổn định thống kê khi mô hình hóa chuỗi thời gian do giảm chiều của dữ liệu đầu vào, tăng khả năng so sánh và diễn giải kết quả. Dựa trên các nghiên cứu nền tảng, bản đồ kết nối khái niệm và thao tác hóa thành các loại hành vi có ý nghĩa sư phạm được xây dựng. Các hành vi này được sắp xếp theo các pha của lí thuyết học tập tự điều chỉnh và thể hiện năng lực số thông qua các đặc trưng có thể quan sát trên hệ thống LMS (xem Bảng 1).

Đối với hệ thống Moodle, nghiên cứu thực nghiệm đã ủng hộ việc sử dụng các nhóm biến được xác định dựa trên tính năng công cụ (Ví dụ: diễn đàn, tài nguyên, bài tập, bài kiểm tra, phản hồi) làm đơn vị phân tích cốt lõi và đã được chứng minh trong bài



Hình 1: Khung thiết kế nghiên cứu tổng thể

Bảng 1: Bản đồ khái niệm - thao tác hoá (Học tập tự điều chỉnh × năng lực số × Log Moodle)

Pha học tập tự điều chỉnh	Nhóm năng lực số	Dấu vết log Moodle	Chỉ số/đặc trưng minh họa
Định hướng và đặt mục tiêu	Thông tin và dữ liệu; Giải quyết vấn đề.	Xem đề cương, tài nguyên định hướng, rubric.	Tần suất xem tài liệu khởi động; độ phủ tài nguyên; thời điểm truy cập trước hạn.
Giám sát	Giao tiếp và hợp tác; An toàn và đạo đức số.	Theo dõi thông báo, kiểm tra tiến độ, đọc phản hồi.	Khoảng cách thời gian phản hồi đến hành động; kiểm tra tiến độ định kì.
Điều chỉnh	Sáng tạo nội dung; Ứng dụng công cụ số.	Làm quiz, nộp bài, sửa bài, workshop/peer grading.	Số vòng sửa bài; tỉ lệ đúng hạn; thời lượng tác vụ.
Phản ánh và Đánh giá	Năng lực siêu nhận thức (metacognition)	Xem thông tin phản hồi, xem “giải thích đáp án”, nhật kí học tập.	Tần suất xem thông tin phản hồi; thời gian đọc thông tin phản hồi.

toán phân tích tương quan với thành tích, cảnh báo sớm và mô hình hóa tiến trình học tập (Macfadyen và Dawson, 2010). Tính hợp lệ của phương pháp phân loại theo chức năng còn được củng cố bởi các chuẩn dữ liệu giáo dục. Theo đó, chuẩn IMS Caliper 1.2 định nghĩa các “Hồ sơ sự kiện” (Event profiles) với các kiểu sự kiện xoay quanh các miền hoạt động học tập chính như Đánh giá (Assessment), Bài tập có thể giao (Assignable), Diễn đàn (Forum) và Điều hướng (Navigation) (IMS Global, 2022). Chuẩn xAPI (Experience API) có cú pháp Tác nhân - Hành vi - Đối tượng và các “Hồ sơ” theo ngữ cảnh để chuẩn hóa cách mô tả hoạt động học tập (ADL, 2023). Cả hai chuẩn này cho thấy một bảng phân loại hành vi dựa trên chức năng hoạt động thay vì dựa trên mã sự kiện riêng lẻ là cần thiết để cả máy tính và con người đều hiểu được ý nghĩa. Về mặt kỹ thuật, kiến trúc Moodle Events API có logic tương thích hoàn toàn với tiếp cận trên. Mỗi sự kiện trong Moodle được mô tả như một thông tin nguyên tử về hành động đã xảy ra và được đặt tên theo mô-đun/chức năng cụ thể (Ví dụ: mod_forum/event/post_created, mod_

quiz/event/attempt_submitted) (Moodle, 2023). Thiết kế này cũng cho thấy việc ánh xạ trực tiếp các sự kiện thành các nhóm hành vi chức năng là khả thi và là cơ sở kỹ thuật cho việc xây dựng mô hình chuỗi thời gian.

Từ bản đồ khái niệm - thao tác hóa, kết hợp với các sự kiện log theo chức năng công cụ, một bộ quy tắc ánh xạ các trường dữ liệu nhật ký thô (Component, event name, action, target, objectType...) thành các loại hành vi sự phạm đã được xây dựng để gom hơn 500,000 sự kiện từ hai học phần vào 10 nhóm hành vi chính (kí hiệu hmm_group10), bao phủ toàn bộ các sự kiện rời rạc trong nhật ký Moodle, thể hiện được những hoạt động tiêu biểu của sinh viên (xem Bảng 2).

Để phù hợp với đầu vào của mô hình Markov ẩn, dữ liệu sau khi được ánh xạ được chuyển đổi thành chuỗi quan sát tuần tự theo tuần. Với mỗi tuần, nhóm hành vi có tần suất cao nhất (ưu thế) được chọn làm quan sát đại diện. Kết quả là một chuỗi gồm 15 kí hiệu (mỗi kí hiệu thuộc một trong 10 nhóm hành vi) tương ứng 15 tuần cho mỗi sinh viên. Để giảm thiểu

Bảng 2: Quy tắc ánh xạ chi tiết nhóm hành vi chính của sinh viên sự phạm trên LMS (hmm_group10)

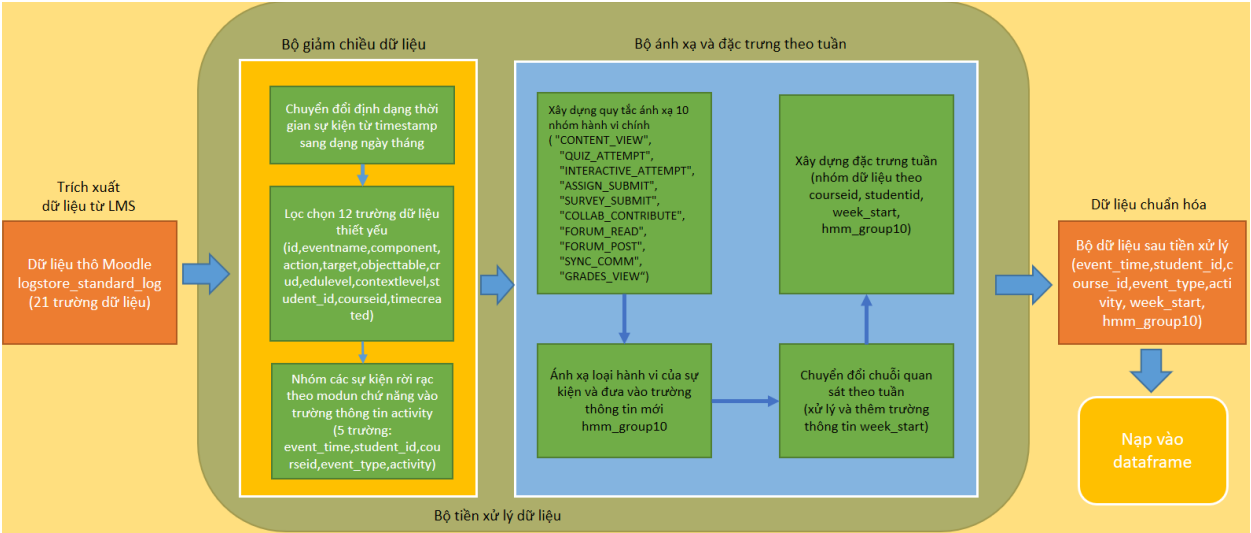
Điều kiện ánh xạ	Nhóm quan sát (hmm_group10)	Mô tả
view × (resource url page file book course grades) course_nav (nếu có)	CONTENT_VIEW (Bao gồm cả GRADES_VIEW)	Xem nội dung/ điều hướng học phần (resource, page, url, file, book, course, grades*)
attempt × quiz	QUIZ_ATTEMPT	Thực hiện bài kiểm tra trắc nghiệm/quiz (bắt đầu/tiến hành)
attempt × (h5p scorm lesson lti bigbluebutton)	INTERACTIVE_ATTEMPT	Tương tác nội dung tương tác (H5P/SCORM/Lesson/LTI/BigBlueButton...)
submit × (assign workshop attendance course) attendance submit (nếu có)	ASSIGN_SUBMIT	Nộp/ cập nhật bài tập/ workshop/ attendance...
submit × (feedback questionnaire survey choice)	SURVEY_SUBMIT	Nộp khảo sát/feedback/questionnaire/survey/choice
contribute × (wiki glossary database)	COLLAB_CONTRIBUTE	Đóng góp vào công cụ cộng tác (wiki, glossary, database)
forum_post × *	FORUM_POST	Tạo thảo luận/bài đăng trên diễn đàn
forum_read × * view × forum	FORUM_READ	Đọc/xem nội dung trên diễn đàn
chat_msg × chat sync_join × (bbb zoom...) (nếu có)	SYNC_COMM	Giao tiếp đồng bộ (tham gia chat...)
khác/không khớp	OTHER	Khác (không thuộc các nhóm trên)

sự sai lệch do chênh lệch quy mô hoạt động giữa các cá nhân (Ví dụ: một sinh viên hiếu động so với một sinh viên ít hoạt động), hai chiến lược chuẩn hóa đã được áp dụng là: 1) Chuẩn hóa theo tuần sử dụng Z-score scaling trên 10 biến quan sát của từng tuần độc lập để so sánh hành vi giữa các sinh viên trong cùng một khoảng thời gian. 2) Chuẩn hóa theo cá nhân bằng cách tính tỉ lệ phần trăm đóng góp của mỗi nhóm hành vi trong tổng số hoạt động của từng sinh viên xuyên suốt khóa học, giúp mô hình Markov ẩn tập trung vào phong cách hành vi và các mẫu chuyển tiếp đặc trưng của từng cá nhân hơn là vào cường độ hoạt động thuần túy, từ đó làm nổi bật

hơn sự tiến triển năng lực nội tại của sinh viên trong quá trình tham gia vào học phần. Toàn bộ quá trình tiền xử lý này được mô tả cụ thể tại ở Hình 2:

2.3. Ảnh xạ ý nghĩa sự phạm giữa hành vi trên LMS với lí thuyết học tập tự điều chỉnh và năng lực số

Để diễn giải ý nghĩa các trạng thái ẩn do mô hình Markov ẩn phát hiện sau này, nghiên cứu này xây dựng một bảng ánh xạ chi tiết giữa các nhóm hành vi trên LMS với các pha học tập tự điều chỉnh và các thành phần năng lực số dựa trên Khung DigComp 2.2 (xem Bảng 3).



Hình 2: Tiền xử lí dữ liệu

Bảng 3: Ánh xạ hành vi

Nhóm hành vi	Ý nghĩa hành vi	Pha học tập tự điều chỉnh chủ đạo	Năng lực số (DigComp)	Mã DigComp
CONTENT_VIEW	Truy cập, xem tài liệu học tập (PDF, video, trang web).	Forethought (Định hướng - Khởi động)	1.1. Duyệt, tìm kiếm và lọc dữ liệu, thông tin và nội dung số	1.1
QUIZ_ATTEMPT	Nỗ lực trả lời câu hỏi trong bài quiz/ kiểm tra, thể hiện việc thực hành, tự kiểm tra kiến thức.	Performance (Thực hiện)	1.2. Đánh giá dữ liệu, thông tin và nội dung số	1.2
ASSIGN_SUBMIT	Nộp bài tập, dự án. Hành động hoàn thành một sản phẩm học tập.	Performance (Thực hiện)	3.1. Phát triển nội dung số	3.1
GRADES_VIEW	Xem điểm số, phản hồi từ giảng viên. Hành vi phản ánh nhu cầu theo dõi kết quả và đánh giá bản thân.	Self-Reflection (Tự đánh giá)	4.2. Bảo vệ dữ liệu cá nhân và quyền riêng tư (xem thông tin cá nhân); 2.3. Tham gia vào xã hội thông qua công nghệ số (nhận phản hồi).	4.2, 2.3

Nhóm hành vi	Ý nghĩa hành vi	Pha học tập tự điều chỉnh chủ đạo	Năng lực số (DigComp)	Mã DigComp
FORUM_POST	Đăng bài mới, trả lời thảo luận. Thể hiện sự tham gia tích cực, đóng góp ý kiến và xây dựng cộng đồng học tập.	Performance (Thực hiện)	3.1. Phát triển nội dung số; 2.3. Tham gia vào xã hội thông qua công nghệ số.	3.1, 2.3
FORUM_READ	Đọc bài đăng trên diễn đàn. Hành vi tiếp nhận thông tin, quan sát và học hỏi từ cộng đồng.	Forethought (Định hướng)	1.1. Duyệt, tìm kiếm và lọc dữ liệu, thông tin và nội dung số.	1.1
SURVEY_SUBMIT	Phản hồi khảo sát. Thể hiện sự tham gia đóng góp ý kiến phản hồi cho khóa học.	Self-Reflection (Tự đánh giá)	2.1. Giao tiếp thông qua công nghệ số; 5.3. Sáng tạo và đổi mới với công nghệ số (đóng góp ý tưởng cải tiến).	2.1, 5.3
INTERACTIVE_ATTEMPT	Tương tác với các hoạt động học tập tương tác (H5P, simulation). Thể hiện sự tham gia thực hành, khám phá.	Performance (Thực hiện)	4.1. Giải quyết vấn đề kỹ thuật; 1.2. Đánh giá dữ liệu, thông tin và nội dung số.	4.1, 1.2
COLLAB_CONTRIBUTE	Đóng góp vào tài liệu cộng tác (Wiki, Google Docs). Thể hiện năng lực làm việc nhóm và đồng sáng tạo trong môi trường số.	Performance (Thực hiện)	3.1. Phát triển nội dung số; 2.2. Chia sẻ thông tin và nội dung thông qua công nghệ số; 2.4. Cộng tác thông qua công nghệ số.	3.1, 2.2, 2.4
SYNC_COMM	Tham gia giao tiếp đồng bộ (phòng học ảo, chat). Thể hiện năng lực giao tiếp và tham gia trực tiếp bằng công nghệ.	Performance (Thực hiện)	2.1. Giao tiếp thông qua công nghệ số; 2.4. Cộng tác thông qua công nghệ số.	2.1, 2.4
OTHERS	Các hành vi chung chung, khó gán ý nghĩa SRL rõ ràng. Có thể là tín hiệu của sự tham gia thụ động hoặc các hoạt động không liên quan trực tiếp đến học tập.			

2.4. Huấn luyện mô hình Markov ẩn

Với mỗi học phần, do tính chất rời rạc của dữ liệu quan sát, một mô hình Markov ẩn với giả định phát xạ đa thức độc lập được huấn luyện để mô hình hóa chuỗi hành vi chính theo tuần nhằm suy luận ra chuỗi các trạng thái năng lực số ẩn của sinh viên sự phạm theo thời gian. Mô hình Markov ẩn với phát xạ đa thức trong nghiên cứu này là một mô hình giả định rằng, khi một sinh viên sự phạm đang ở trong một trạng thái năng lực số ẩn nào đó (Ví dụ: “Thực hiện”), sinh viên đó sẽ có một xác suất nhất định để thực hiện mỗi loại hành vi quan sát được (Ví dụ: xác suất thực hiện ASSIGN_SUBMIT là rất cao, trong khi xác suất FORUM_READ lại rất thấp). Mô hình sẽ học các xác suất này từ dữ liệu thực tế.

Một mô hình Markov ẩn được xác định bởi: 1) Số trạng thái ẩn (K), 2) Tập hợp các xác suất chuyển

trạng thái, 3) Xác suất phát xạ quan sát từ mỗi trạng thái. Giá trị K tối ưu ở đây là chưa biết trước và cần được xác định dựa trên dữ liệu. Do đó, mô hình Markov ẩn được huấn luyện với số trạng thái K lần lượt thuộc tập {2,3,4,5} trên dữ liệu mỗi học phần. Phạm vi này được lựa chọn dựa trên cơ sở: K=2 là giá trị tối thiểu, phù hợp để kiểm tra giả thuyết về sự tồn tại của ít nhất hai “chế độ” học tập cơ bản (thụ động/chủ động hoặc định hướng/ thực hiện); K=3 đến K=5 có thể cho phép mô hình bắt được các sắc thái phức tạp hơn trong quá trình học tập, có thể tương ứng với các giai đoạn chi tiết hơn của khung lý thuyết học tập tự điều chỉnh (Ví dụ: tách biệt giai đoạn tự đánh giá); K>5 không được xem xét để tránh hiện tượng quá khớp dữ liệu và tạo ra các mô hình quá phức tạp, khó diễn giải ý nghĩa sự phạm của từng trạng thái. Thuật toán lặp để tìm cực đại của khả năng xảy ra

xác xuất Baum - Welch được sử dụng để ước lượng tham số cho từng mô hình Markov ẩn ứng với mỗi K, với nhiều lần khởi tạo ngẫu nhiên (mỗi mô hình cho một giá trị K cụ thể được huấn luyện 50 lần) để tránh cực trị địa phương. Toàn bộ quá trình huấn luyện được thực hiện bằng thư viện hmmlearn trong Python.

3. Kết quả nghiên cứu

3.1. Lựa chọn số trạng thái ẩn tối ưu

Để lựa chọn mô hình Markov ẩn tối ưu, các tiêu chí thông tin thống kê AIC và BIC đã được sử dụng để đánh giá các mô hình với số trạng thái ẩn (K) từ 2 đến 5. Trong đó, BIC được ưu tiên do có xu hướng chọn mô hình đơn giản hơn, giúp tránh hiện tượng quá khớp dữ liệu. Kết quả ở Bảng 4 cho thấy, đối với cả hai học phần, mô hình với 2 trạng thái ẩn (K=2) có giá trị BIC thấp nhất, chỉ ra rằng đây là mô hình phù hợp nhất với dữ liệu, cân bằng được độ phức tạp và khả năng giải thích.

3.2. Đặc trưng hành vi của trạng thái ẩn trong từng học phần

Phân tích ma trận phát xạ của mô hình Markov ẩn hai trạng thái sẽ diễn giải ý nghĩa sự phạm của từng trạng thái, qua đó cho thấy hình ảnh năng lực số khác nhau của sinh viên trong mỗi học phần.

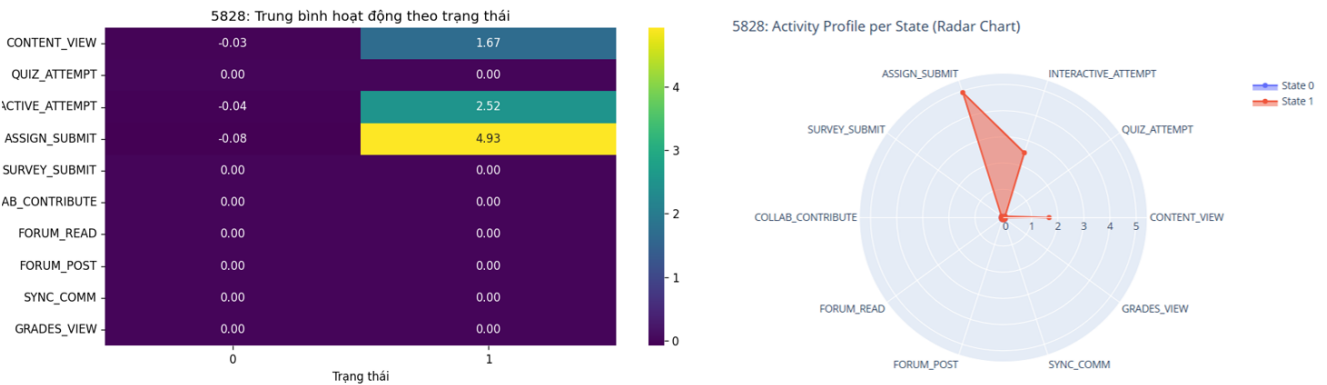
Đối với học phần “Tin học đại cương”, ma trận

phát xạ trạng thái cho thấy hai trạng thái ẩn được phân cực một cách rõ rệt (xem Hình 3). Trạng thái 1 được đặc trưng bởi mức độ hoạt động rất cao ở các hành vi thực thi cụ thể. Giá trị trung bình có trọng số của ASSIGN_SUBMIT (4.93) là cao vượt trội, theo sau là INTERACTIVE_ATTEMPT (2.52) và CONTENT_VIEW (1.67). Điều này cho thấy sinh viên ở trạng thái này đang tập trung cao độ vào việc hoàn thành và nộp các sản phẩm được yêu cầu, tương ứng với pha “Thực hiện - Tự đánh giá” của học tập tự điều chỉnh. Tuy nhiên, hành vi nộp bài là hành vi nổi bật cho thấy hoạt động này mang tính chất phản ứng theo móc/đến hạn hơn là một quá trình chủ động. Ngược lại, trạng thái 0 có mức hoạt động gần như bằng không ở tất cả hành vi, biểu thị trạng thái thụ động hoặc phụ thuộc vào tài liệu ngoại tuyến. Đây là trạng thái cơ bản hoạt động tối thiểu, nơi sinh viên có rất ít tương tác có thể đo lường được trên LMS, cho thấy có pha “Định hướng” rất hạn chế. Kết quả này cho thấy một mẫu hình học tập lưỡng cực, trong đó sinh viên hoặc là đang “bật” chế độ hoạt động cao độ để hoàn thành một nhiệm vụ cụ thể, hoặc là gần như hoàn toàn “tắt” trên môi trường số. Mẫu hình học tập gợi ý sự tham gia thiên về thực thi hơn là chuẩn bị có chủ đích; dấu hiệu của pha Performance/Reflection ngắn và thiếu Forethought theo lý thuyết học tập tự điều chỉnh (Pintrich, 2004). Trạng thái “Thực hiện” ở học phần lý thuyết này cũng phản ánh một dạng

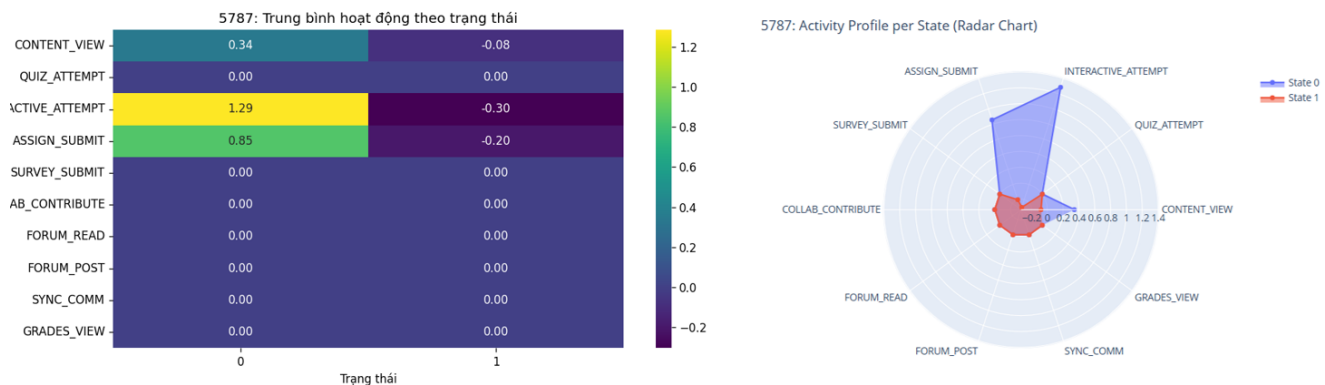
Bảng 4: Giá trị BIC của các mô hình HMM theo số trạng thái (K)

Học phần	K=2	K=3	K=4	K=5
UDCNTT_5787	-35592.42	-35477.92	-35328.71	-35132.58
THUD_5828	-149364.23	-149293.75	-149085.18	-148861.21

(Ghi chú: Giá trị BIC nhỏ nhất (tốt nhất) được in đậm)



Hình 3: Ma trận phát xạ trạng thái và biểu đồ hoạt động của sinh viên trong học phần “Tin học đại cương” (mã THUD_5828)



Hình 4: Ma trận phát xạ trạng thái và biểu đồ hoạt động của sinh viên trong học phần học phần “Ứng dụng công nghệ thông tin trong dạy học” (mã UDCNTT_5787)

năng lực số thủ tục thể hiện ở khả năng sử dụng các công cụ số để hoàn thành một nhiệm vụ đã được xác định rõ ràng (Ví dụ: tải lên một tệp).

Đối với học phần “Ứng dụng công nghệ thông tin trong dạy học”, các trạng thái ẩn cũng cho thấy sự khác biệt, nhưng theo một chiều hướng khác (xem Hình 4). Trạng thái 0 trong học phần này là trạng thái tích cực, được đặc trưng bởi mức độ hoạt động tích cực đồng thời ở nhiều loại hành vi: INTERACTIVE_ATTEMPT (1.29), ASSIGN_SUBMIT (0.85) và CONTENT_VIEW (0.34). Sự kết hợp giữa việc xem tài liệu (liên quan đến pha “Định hướng - Khởi động”), tương tác với nội dung thực hành và nộp bài (liên quan đến pha “Thực hiện - Tự đánh giá”) cho thấy một cách tiếp cận chủ động và cân bằng hơn. Trạng thái 1 ở học phần này tương tự như trạng thái nền ở học phần “Tin học đại cương”, được đặc trưng bởi các giá trị âm trên các hoạt động chính, cho thấy những giai đoạn ít tương tác trên LMS. Cấu hình này hàm ý một chu trình học tập tự điều chỉnh liên mạch hơn (khởi động - thực thi - tự đánh giá) so với học phần trên (Pintrich, 2004). Trạng thái “tích cực” cũng phản ánh một năng lực số tích hợp hơn, trong đó năng lực số được phát triển không chỉ giới hạn ở việc hoàn thành nhiệm vụ mà còn bao gồm cả việc khám phá và thực hành. Cách tiếp cận này phù hợp hơn với tầm nhìn toàn diện của các khung năng lực như DigComp, vốn nhấn mạnh việc sử dụng công nghệ một cách sáng tạo và có phê phán. Điều này củng cố quan điểm rằng, học tập tự điều chỉnh không phải là một kỹ năng đơn lẻ mà là một quá trình lặp đi lặp lại. So sánh hai học phần cho thấy mô hình Markov ẩn trong nghiên cứu này thể hiện được các quá trình học tập khác nhau và đôi khi không hoàn chỉnh trong các bối cảnh học tập khác nhau.

Các biểu đồ radar kèm theo củng cố các nhận định: trọng số quan sát được chuẩn hoá theo từng

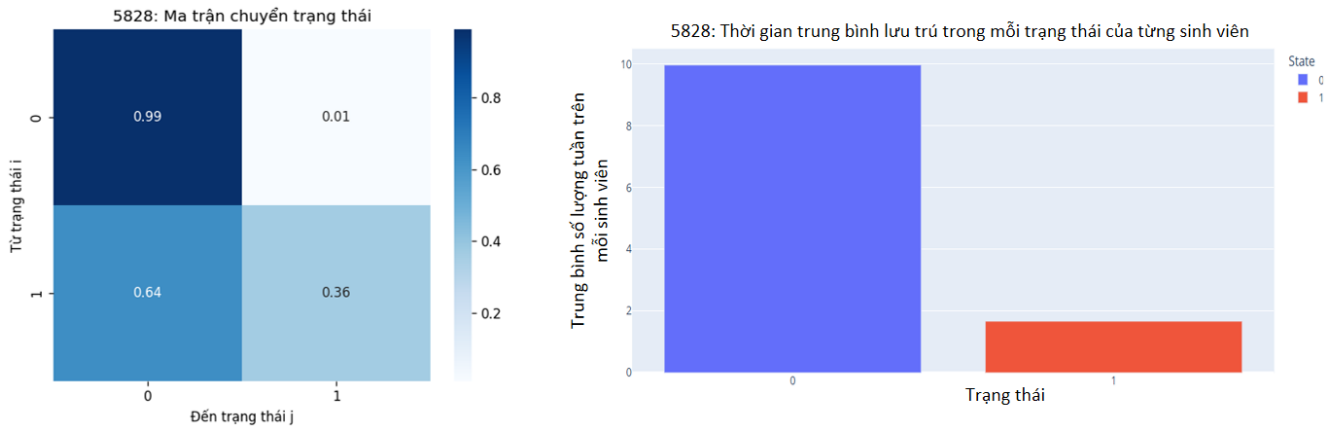
nhóm hành vi, giúp đối chiếu trực quan mức độ trội/yếu của mỗi nhóm trong từng trạng thái để diễn giải sự phạm nhất quán giữa hai học phần (xem Hình 3 và Hình 4).

3.3. Phân tích động lực chuyển dịch trạng thái

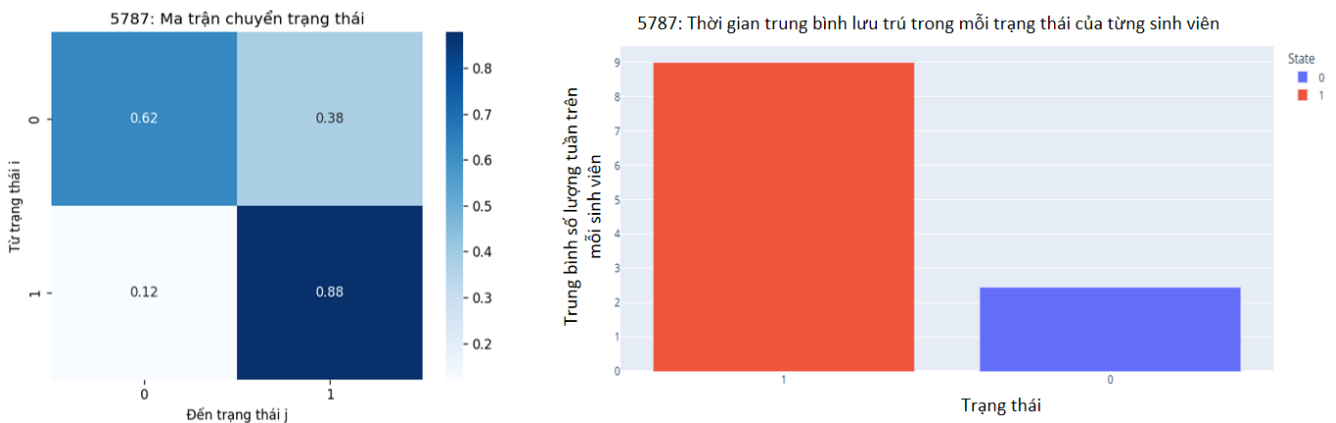
Phân tích ma trận chuyển trạng thái và thời gian lưu trung bình trong mỗi trạng thái cho thấy những động lực học tập hoàn toàn khác biệt giữa hai học phần (xem Hình 5).

Động lực học tập trong học phần “Tin học đại cương” này cho thấy trạng thái “Định hướng - Khởi động” - hoạt động tối thiểu (Trạng thái 0) có tính bền vững cực cao (xem Hình 5). Một sinh viên ở trạng thái này có xác suất lên tới 99% sẽ tiếp tục ở lại trạng thái đó trong tuần tiếp theo. Trạng thái này hoạt động như một “điểm hút” mạnh mẽ. Ngược lại, trạng thái “Thực hiện - Tự đánh giá” (Trạng thái 1) rất tạm thời. Một khi sinh viên chuyển sang trạng thái hoạt động này, xác suất sẽ quay trở lại trạng thái tối thiểu ngay trong tuần sau là 64%. Biểu đồ về thời gian lưu trung bình củng cố nhận định này khi sinh viên dành trung bình 9.97 tuần ở trạng thái tối thiểu nhưng chỉ 1.67 tuần ở trạng thái hoạt động. Đây là một quỹ đạo học tập đặc trưng bởi sự tuân thủ theo thời hạn của nhiệm vụ, hoạt động học tập trong học phần này mang tính chất “bùng nổ” theo mốc thời gian hơn là một quá trình tham gia liên tục (xem Hình 6).

Động lực học tập ở học phần “Ứng dụng công nghệ thông tin trong dạy học” hoàn toàn khác (xem Hình 6). Mặc dù trạng thái “Định hướng - Khởi động” - hoạt động thấp (Trạng thái 1) cũng khá bền vững (xác suất 88% ở lại), trạng thái “Thực hiện - Tự đánh giá” - tham gia tích cực (Trạng thái 0) lại có độ ổn định tương đối cao. Một sinh viên đang tích cực có xác suất 62% sẽ tiếp tục tích cực trong tuần tiếp theo. Thời gian lưu trung bình trong trạng thái tích



Hình 5: Ma trận chuyển trạng thái và biểu đồ thời gian lưu trú trung bình trong mỗi trạng thái của học phần “Tin học đại cương” (mã THUD_5828)



Hình 6: Ma trận chuyển trạng thái và biểu đồ thời gian lưu trú trung bình trong mỗi trạng thái của học phần “Ứng dụng công nghệ thông tin trong dạy học” (mã UDCNTT_5787)

cực là 2.45 tuần, dài hơn đáng kể so với học phần kia. Điều này cho thấy một quỹ đạo học tập trong đó sinh viên có sự tham gia chủ động và có xu hướng chuyển dịch linh hoạt giữa hai trạng thái. Khi đã vào trạng thái “Thực hiện”, sinh viên có xu hướng duy trì khá bền vững, phản ánh đúng tính chất thực hành liên tục của học phần, nơi mà các dự án và bài tập được phân bổ đều đặn khiến sinh viên giữ được nhịp độ tham gia cao suốt học phần.

Sự tương phản này cho thấy hai quỹ đạo học tập về cơ bản là khác nhau: Một quỹ đạo được thúc đẩy bởi sự tuân thủ phản ứng và một quỹ đạo khác được đặc trưng bởi sự tham gia chủ động và bền vững hơn. Sự khác biệt này gợi ý một mối liên hệ nhân quả với cấu trúc thiết kế sự phạm của mỗi học phần. Đây cũng là một điểm cốt lõi và có ý nghĩa cao của nghiên cứu này với việc phát hiện ra vai trò quyết định của thiết kế sự phạm trong việc định hình các mẫu hành vi tự điều chỉnh dẫn đến hình thành các quỹ đạo học tập của sinh viên.

4. Kết luận

Nghiên cứu đã chứng minh một phương pháp tiếp cận tích hợp, kết hợp mô hình Markov ẩn và Lí thuyết học tập tự điều chỉnh để suy luận trạng thái năng lực số của sinh viên sự phạm từ dấu vết học tập trên LMS. Kết quả chính chỉ ra rằng mô hình Markov ẩn có thể xác định hiệu quả hai trạng thái học tập tiềm ẩn, được diễn giải một cách có ý nghĩa là “Định hướng” và “Thực hiện - Tự giám sát” thông qua lăng kính lí thuyết học tập tự điều chỉnh. Quan trọng hơn, sự tương phản rõ rệt về động lực chuyển dịch trạng thái giữa hai học phần đã cho thấy thiết kế sự phạm đóng vai trò then chốt trong việc định hình quỹ đạo phát triển năng lực số. Điều này cũng cho thấy năng lực số không phải là một đặc điểm cố định, mà là một trạng thái năng động, chịu ảnh hưởng sâu sắc bởi bối cảnh học tập. Do đó, các nỗ lực nhằm cải thiện năng lực số của sinh viên sự phạm phải tập trung vào việc thiết kế các trải nghiệm học tập tốt hơn, cũng như việc đào tạo kĩ năng trực tiếp.

Về mặt lí luận, nghiên cứu đóng góp một khung phương pháp luận mới, kết nối giữa học máy, lí thuyết giáo dục và phân tích dữ liệu hành vi để đánh giá một năng lực phức hợp như năng lực số.

Về mặt thực tiễn, nghiên cứu cung cấp một công cụ khách quan và tự động hóa tiềm năng cho việc chẩn đoán và hỗ trợ quá trình phát triển năng lực số, từ đó hỗ trợ đưa ra các quyết định can thiệp kịp thời và phù hợp. So với các phương pháp đánh giá

năng lực số truyền thống (như bảng hỏi tự đánh giá), phương pháp này cung cấp một đánh giá khách quan, liên tục và ít chịu ảnh hưởng bởi thiên kiến của người đánh giá. Nghiên cứu này cũng mở ra một hướng phát triển cho các hệ thống phản hồi, can thiệp tự động dựa trên kết quả đầu ra của mô hình Markov ẩn như là một hướng đi đầy hứa hẹn, góp phần hiện thực hóa tiềm năng của lĩnh vực phân tích học tập trong việc cá nhân hóa và nâng cao chất lượng giáo dục.

Tài liệu tham khảo

- ADL.(2023). *xAPI Profiles*. Advanced Distributed Learning Initiative. <https://adlnet.github.io/xapi-profiles/xapi-profiles-about.html>.
- Baker, Ryan. S. & Yacef, Kalina. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining*, 1(1), 3-17. <https://doi.org/10.5281/zenodo.3554657>.
- Blikstein, Paulo & Worsley, Marcelo. (2016). Multimodal learning analytics and education data mining: Using computational technologies to measure complex learning tasks. *Journal of Learning Analytics*, 3(2), 220-238. <https://doi.org/10.18608/jla.2016.32.11>.
- Carretero, S., Vuorikari, R. & Punie, Y. (2017). *DigComp 2.1: The Digital Competence Framework for Citizens with eight proficiency levels and examples of use*. Publications Office of the European Union. <https://doi.org/10.2760/38842>.
- IMS Global. (2022). *IMS Caliper Analytics® Specification*. IMS Global Learning Consortium. <https://www.imsglobal.org/spec/caliper/v1p2>.
- Lê Anh Vinh, Bùi Diệu Quỳnh, Đỗ Đức Lân, Đào Thái Lai & Tạ Ngọc Trí. (2021). Xây dựng khung năng lực số cho học sinh phổ thông Việt Nam. *Tạp chí Khoa học Giáo dục Việt Nam*, 17(1), 1-11.
- Le Thai, Hung, Dinh Thi Kim Thoa, Vu Phuong Lie & Nguyen Thi Phuong Vy. (2022). ICT competence of pre-service teachers in Vietnam: Structure and impact model. *Journal of Educational and Social Research*, 12(3), 172. <https://doi.org/10.36941/jesr-2022-0076>.
- Macfadyen, Leah. P. & Dawson, Shane. (2010). Mining LMS data to develop an “early warning system” for educators: A proof of concept. *Computers & Education*, 54(2), 588-599. <https://doi.org/10.1016/j.compedu.2009.09.008>.
- Matcha Wannisa, Gasevic Dragan, Nora Ayu Ahmad Uziz, Jelena Jovanovic & Pardo Abelardo. (2019). Analytics of learning strategies: Associations with academic performance and feedback. In *Proceedings of the 9th International Conference on Learning Analytics & Knowledge* (pp. 461-470). ACM. <https://doi.org/10.1145/3303772.3303787>.
- Ministry of Education and Training (MOET). (2025). *Khung năng lực số cho người học*. Hanoi.
- Moodle. (2023). *Events API*. Moodle Developer Resources. https://docs.moodle.org/dev/Events_API
- Pintrich, Paul R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review*, 16(4), 385-407. <https://doi.org/10.1007/s10648-004-0006-x>.
- Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2), 257-286. <https://doi.org/10.1109/5.18626>.
- Siemens George, & Baker Ryan S. (2012). Learning analytics and educational data mining: Towards communication and collaboration. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 252-254). ACM. <https://doi.org/10.1145/2330601.2330661>.
- Sillat Linda Helene, Tammets Kairit & Laanpere Mart. (2021). Digital competence assessment methods in higher education: A systematic literature review. *Education Sciences*, 11(8), 402. <https://doi.org/10.3390/educsci11080402>.
- Vuorikari, R., Kluzer, S. & Punie, Y. (2022). *DigComp 2.2: The Digital Competence Framework for Citizens - With new examples of knowledge, skills and attitudes*. Publications Office of the European Union. <https://doi.org/10.2760/490274>.
- Winne, P. H. (2014). Issues in researching self-regulated learning as patterns of events. *Metacognition and Learning*, 9(2), 229-237. <https://doi.org/10.1007/s11409-014-9113-3>.
- Xing Wanli & Du Dongping. (2019). Dropout prediction in MOOCs: Using deep learning for personalized intervention. *Journal of Educational Computing Research*, 57(3), 547-570. <https://doi.org/10.1177/0735633118757015>.
- Zimmerman Barry J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice*, 41(2), 64-70. https://doi.org/10.1207/s15430421tip4102_2.