Ứng dụng Trí tuệ nhân tạo trong xây dựng hệ thống Học tăng cường hỗ trợ dạy và học STEM

1st abc
Khoa Công nghệ Thông tin
Trường Đại học Sài Gòn
TP. Hồ Chí Minh, Việt Nam
emaik@sgu.edu.vn

2nd ABC
Khoa Công nghệ Thông tin
Trường Đại học Sài Gòn
TP. Hồ Chí Minh, Việt Nam
email@sgu.edu.vn

3rd ABC
Khoa Công nghệ Thông tin
Trường Đại học Sài Gòn
TP. Hồ Chí Minh, Việt Nam
email@sgu.edu.vn

***Abstract*—Trong kỷ nguyên công nghiệp 4.0, giáo dục STEM đóng vai trò then chốt trong việc đào tạo nguồn nhân lực chất lượng cao. Tuy nhiên, các phương pháp giảng dạy truyền thống và hệ thống quản lý học tập (LMS) hiện hành thường áp dụng cách tiếp cận "một kích cỡ cho tất cả", thất bại trong việc đáp ứng nhu cầu cá nhân hóa của từng người học. Bài báo này đề xuất một hệ thống gợi ý lộ trình học tập thông minh sử dụng kỹ thuật Học tăng cường (Reinforcement Learning - RL), cụ thể là thuật toán Q-learning, được tích hợp vào nền tảng Moodle qua chuẩn LTI 1.3. Hệ thống mô hình hóa quá trình học tập dưới dạng Quy trình quyết định Markov (MDP), sử dụng dữ liệu hành vi thực tế để phân cụm người học và tối ưu hóa chiến lược gợi ý. Kết quả thực nghiệm mô phỏng trên 500 vòng lặp cho thấy thuật toán giúp tăng 22.5% điểm số trung bình và giảm 51.0% số lượng kỹ năng yếu so với phương pháp truyền thống.**

***Index Terms*—Học tăng cường, Q-learning, Cá nhân hóa học tập, Giáo dục STEM, Microservices, LTI 1.3.**

I. Giới thiệu

Sự phát triển mạnh mẽ của Trí tuệ nhân tạo (AI) đang định hình lại nhiều lĩnh vực, trong đó có giáo dục. [cite_start]Theo nghiên cứu của Frey và Osborne, khoảng 47% các công việc truyền thống có nguy cơ bị tự động hóa, đặt ra yêu cầu cấp thiết về việc trang bị các kỹ năng mới cho người lao động, đặc biệt là các kỹ năng STEM (Khoa học, Công nghệ, Kỹ thuật và Toán học) [2]. Giáo dục STEM chú trọng phát triển tư duy phản biện và khả năng giải quyết vấn đề, tuy nhiên, việc triển khai hiệu quả gặp nhiều rào cản do sự đa dạng về năng lực và tốc độ tiếp thu của học viên.

Thách thức lớn nhất hiện nay là cá nhân hóa trải nghiệm học tập (Personalized Adaptive Learning - PAL) trên quy mô lớn. [cite_start]Các hệ thống LMS truyền thống như Moodle, Blackboard chủ yếu đóng vai trò lưu trữ tài liệu và quản lý điểm số, thiếu khả năng phân tích hành vi để đưa ra các can thiệp sư phạm kịp thời [1]. [cite_start]Tại Việt Nam, các nghiên cứu về ứng dụng AI trong giáo dục chủ yếu tập trung vào bài toán dự báo (prediction) - ví dụ như dự báo nguy cơ bỏ học hoặc dự đoán điểm số cuối kỳ - mà chưa chú trọng nhiều đến bài toán đưa ra khuyến nghị hành động (prescription) để cải thiện kết quả đó [4].

Để giải quyết vấn đề này, nhu cầu về một hệ thống hỗ trợ dạy và học STEM cá nhân hóa ứng dụng Học tăng cường (Reinforcement Learning - RL) trở nên cấp thiết. RL, với cơ chế học thử-sai (trial-and-error), cho phép hệ thống tự động khám phá và tối ưu hóa chiến lược giảng dạy dựa trên phản hồi liên tục từ người học.

Nghiên cứu này đề xuất xây dựng một hệ thống gợi ý thông minh tích hợp vào Moodle LMS, với các đóng góp chính sau:

1) Đề xuất kiến trúc Microservices tích hợp qua chuẩn LTI 1.3, đảm bảo khả năng mở rộng và tương thích với nhiều nền tảng LMS.
2) Xây dựng quy trình xử lý dữ liệu và phân cụm người học để giải quyết bài toán không gian trạng thái trong RL.
3) Thiết kế và tối ưu hóa thuật toán Q-learning với hàm phần thưởng đa mục tiêu, thích ứng với đặc điểm của từng nhóm người học.
4) Kiểm chứng hiệu quả của hệ thống thông qua mô phỏng so sánh (A/B testing) với dữ liệu tham số hóa từ thực tế.

II. Tổng quan nghiên cứu

A. Hệ thống học tập thích ứng (Adaptive Learning)

Các hệ thống PAL điều chỉnh lộ trình học tập dựa trên nhu cầu riêng biệt của sinh viên. [cite_start]Theo báo cáo tổng quan của du Plooy và cộng sự (2024), khảo sát 69 công trình nghiên cứu, 59% số nghiên cứu ghi nhận sự cải thiện về kết quả học tập và 36% chỉ ra sự gia tăng mức độ tham gia khi áp dụng PAL [1]. Các nền tảng như Moodle và McGraw-Hill's Connect LearnSmart là những môi trường phổ biến để triển khai các giải pháp này.

B. Học tăng cường trong giáo dục

Trong những năm gần đây, Học tăng cường (RL) đã nổi lên như một phương pháp hiệu quả để xây dựng các gia sư thông minh (Intelligent Tutoring Systems). [cite_start]Theo Riedmann (2025), xu hướng sử dụng RL trong giáo dục tăng mạnh từ năm 2018, đặc biệt trong lĩnh vực STEM [3]. [cite_start]Về mặt thuật toán, Q-learning và Deep Q-Network (DQN) là phổ biến nhất nhờ tính linh hoạt của phương pháp Model-free, cho phép hệ thống học chiến lược tối ưu mà

không cần mô hình hóa chính xác quy luật phức tạp của hành vi con người [6].

[cite_start]Tuy nhiên, Riedmann cũng chỉ ra hạn chế lớn của các nghiên cứu hiện tại là sự phụ thuộc vào dữ liệu mô phỏng và thiếu các triển khai thực tế tại các quốc gia đang phát triển [3].

C. Tình hình nghiên cứu tại Việt Nam

Tại Việt Nam, các nghiên cứu chủ yếu tập trung vào khai phá dữ liệu (Data Mining) để dự báo. [cite_start]Ví dụ, nghiên cứu của Trần Bá Thuấn (2021) sử dụng Random Forest để dự báo cảnh báo học vụ, hay Lưu Hoài Sang (2020) sử dụng Deep Learning để dự báo điểm số [4]. Mặc dù đạt độ chính xác cao, các mô hình này mang tính chất "dự báo tĩnh" và thiếu cơ chế gợi ý động theo thời gian thực. Đề tài này nhằm lấp đầy khoảng trống đó bằng cách xây dựng một hệ thống RL có khả năng tương tác và điều hướng người học.

III. Kiến trúc hệ thống và Xử lý dữ liệu

A. Kiến trúc Microservices

Hệ thống được thiết kế theo kiến trúc Microservices để đảm bảo tính module hóa và khả năng mở rộng độc lập. Các thành phần chính bao gồm:

- **LTI Integration Service:** Đảm nhiệm việc xác thực và kết nối an toàn với Moodle LMS thông qua chuẩn LTI 1.3 và giao thức OAuth 2.0. [cite_start]Service này giúp hệ thống hoạt động như một công cụ độc lập (Tool Provider) mà không cần can thiệp vào mã nguồn của LMS [7].
- **Recommend Service:** Là trái tim của hệ thống, chứa thuật toán Q-learning (được viết bằng Python). [cite_start]Service này chịu trách nhiệm tính toán giá trị Q, cập nhật bảng tri thức và trả về danh sách gợi ý [5].
- **Course Service & User Service:** Quản lý metadata cấu trúc khóa học và thông tin người dùng, đồng bộ hóa từ LMS.
- **Frontend:** Ứng dụng phía người dùng xây dựng bằng ReactJS, hiển thị dashboard cá nhân hóa.

[cite_start]Dữ liệu được lưu trữ phân tán, với MongoDB được sử dụng cho Recommend Service để xử lý dữ liệu log hành vi phi cấu trúc và trạng thái người học thay đổi liên tục [5].

B. Quy trình Xử lý dữ liệu và Phân cụm

[cite_start]Để xây dựng không gian trạng thái cho thuật toán RL, chúng tôi thực hiện quy trình khai phá dữ liệu từ log hệ thống Moodle của một khóa học STEM thực tế (Course ID 670) với 13,995 sự kiện tương tác và 23 sinh viên tham gia đầy đủ [5].

1) Tiền xử lý dữ liệu: Dữ liệu thô từ Moodle chứa nhiều nhiễu. Quá trình làm sạch bao gồm việc loại bỏ các sự kiện hệ thống tự động (như *webservice_function_called*) và chuẩn hóa thời gian. [cite_start]Sau đó, chúng tôi trích xuất 114 đặc trưng hành vi và áp dụng kỹ thuật lọc tương quan (Correlation Filtering) để loại bỏ hiện tượng đa cộng tuyến, giữ lại 15 đặc trưng quan trọng nhất như: tần suất xem tài liệu, số lần nộp bài, thời gian phản hồi, v.v. [5].

2) Phân cụm người học (Clustering): Việc cá nhân hóa đòi hỏi hệ thống phải nhận diện được đặc điểm của từng nhóm đối tượng. Chúng tôi sử dụng thuật toán K-means để phân nhóm sinh viên dựa trên 15 đặc trưng đã trích xuất. [cite_start]Số lượng cụm tối ưu $K = 6$ được xác định thông qua phương pháp Elbow và chỉ số Silhouette [5]. Các nhóm người học được định danh bao gồm:

- **Cluster 0 (Nhóm cần hỗ trợ - 34.8%):** Tương tác thấp, thụ động, điểm số thấp.
- **Cluster 1 (Nhóm tự giác):** Thường xuyên theo dõi tiến độ và bảng điểm.
- **Cluster 2 (Nhóm chủ động - 30.4%):** Tương tác cao, nộp bài đầy đủ.
- **Cluster 4 (Nhóm nghiên cứu):** Tập trung vào việc xem và tải tài liệu. [cite_start]
- **Cluster 5 (Nhóm thành tích):** Quan tâm đến huy hiệu và xếp hạng [5].

Thông tin Cluster ID này sẽ được đưa vào vector trạng thái của thuật toán Q-learning.

IV. Mô hình hóa bài toán Q-learning

Bài toán gợi ý lộ trình học tập được mô hình hóa dưới dạng Quy trình Quyết định Markov (MDP), bao gồm bộ ba $< S, A, R >$.

A. Không gian trạng thái (State Space)

Trạng thái S_t tại thời điểm t đại diện cho tình trạng học tập hiện tại của sinh viên, được định nghĩa là một vector 6 chiều:

$$S_t = (C, M, P, Sc, Ph, E) \tag{1}$$

Trong đó:

- C (Cluster): Nhóm người học $(0 \dots 5)$ từ kết quả phân cụm.
- M (Module): Chỉ số bài học hiện tại trong lộ trình tuần tự.
- P (Progress): Mức độ hoàn thành module, được rời rạc hóa thành 4 mức (0.25, 0.5, 0.75, 1.0).
- Sc (Score): Điểm số tích lũy, chia thành 4 mức tương ứng (Yếu, TB, Khá, Giỏi).
- Ph (Phase): Giai đoạn học tập (0: Pre-learning, 1: Active-learning, 2: Reflective-learning). [cite_start]
- E (Engagement): Mức độ tương tác (Thấp, TB, Cao), tính toán dựa trên trọng số hành động theo khung ICAP [5].

Kích thước không gian trạng thái là $5 \times 6 \times 4 \times 4 \times 3 \times 3 \approx 4,320$ trạng thái, đảm bảo tính khả thi cho việc hội tụ của bảng Q-table.

B. Không gian hành động (Action Space)

Thay vì sử dụng hàng trăm sự kiện thô của Moodle, chúng tôi thiết kế 15 hành động sư phạm cốt lõi, được phân loại theo ngữ cảnh thời gian để hỗ trợ các chiến lược học tập khác nhau (Bảng I).

Table I
KHÔNG GIAN HÀNH ĐỘNG THEO NGỮ CẢNH THỜI GIAN

Ngữ cảnh	Hành động	Mục tiêu sư phạm
PAST (Quá khứ)	view_content_past	Ôn tập kiến thức cũ (Remediation)
	review_quiz_past	Phân tích lỗi sai
	attempt_quiz_past	Cải thiện kỹ năng còn yếu
	post_forum_past	Thảo luận lại chủ đề cũ
CURRENT (Hiện tại)	view_assignment	Hiểu yêu cầu nhiệm vụ
	view_content	Tiếp thu kiến thức mới
	attempt_quiz	Thực hành kiến tạo (Constructive)
	submit_quiz	Hoàn thành đánh giá
	post_forum	Tương tác xã hội
FUTURE (Tương lai)	view_content_fut	Chuẩn bị bài trước (Preview)
	attempt_quiz_fut	Thử thách nâng cao (Exploration)

[cite_start]Việc phân chia này cho phép AI đưa ra các chiến lược linh hoạt: Gợi ý hành động Past cho sinh viên yếu cần củng cố nền tảng, và hành động Future cho sinh viên giỏi muốn học vượt [5].

C. Hàm phần thưởng (Reward Function)

Hàm thưởng là thành phần quan trọng nhất định hướng hành vi của tác nhân. Chúng tôi thiết kế hàm thưởng tổng hợp R_{total}:

$$R_{total} = R_{base} + R_{LO} + R_{seq} - P_{penalty} \quad (2)$$

1) Phần thưởng cơ bản thích ứng (R_{base}): Áp dụng chiến lược Cluster-Adaptive Reward. Nhóm sinh viên Yếu (Weak Cluster) nhận được điểm thưởng cao hơn (+10) khi hoàn thành một hành động so với nhóm Giỏi (+5). [cite_start]Cơ chế này nhằm tạo động lực ngoại sinh, khuyến khích nhóm yếu duy trì tương tác thay vì bỏ cuộc [6].

2) Phần thưởng dựa trên Chuẩn đầu ra (R_{LO}): Được tính dựa trên mức độ cải thiện năng lực ($\Delta Mastery$) của sinh viên đối với các Chuẩn đầu ra (LO).

$$R_{LO} = \sum (\Delta Mastery_i \times W_{midterm,i} \times K_{cluster}) \quad (3)$$

[cite_start]Trong đó, $W_{midterm}$ là trọng số quan trọng của LO trong bài thi, và $K_{cluster}$ là hệ số khích lệ riêng cho từng nhóm [5].

3) Phần thưởng chuỗi hành động (R_{seq}): Hệ thống thưởng thêm điểm cho các chuỗi hành động hợp lý về mặt sư phạm (Beneficial Sequences), ví dụ: *view_content* $\rightarrow$ *attempt_quiz* (Học đi đôi với hành), hoặc *review_quiz* $\rightarrow$ *view_content* (Phản tư và học lại).

D. Thuật toán Q-learning

Hệ thống sử dụng thuật toán Q-learning tiêu chuẩn để cập nhật bảng giá trị Q:

$$Q(s,a) \leftarrow Q(s,a) + \alpha[r + \gamma \max_{a'} Q(s',a') - Q(s,a)] \quad (4)$$

Với tốc độ học $\alpha = 0.1$ và hệ số chiết khấu $\gamma = 0.95$. [cite_start]Chiến lược lựa chọn hành động là ϵ-greedy, với ϵ giảm dần từ 1.0 xuống 0.01 sau 400 episodes để cân bằng giữa khám phá và khai thác [6].

V. THỰC NGHIỆM VÀ KẾT QUẢ

Do giới hạn về việc triển khai trên sinh viên thực tế trong thời gian ngắn, nghiên cứu sử dụng phương pháp mô phỏng dựa trên dữ liệu (Data-driven Simulation) để giải quyết vấn đề "khởi động lạnh" và kiểm chứng thuật toán.

A. Thiết lập môi trường mô phỏng

[cite_start]Môi trường giả lập được xây dựng dựa trên các tham số thống kê (xác suất chuyển trạng thái, phân phối điểm số) trích xuất từ dữ liệu khóa học 670. Các tác nhân ảo (Virtual Agents) được sinh ra với phân phối năng lực mô phỏng thực tế: 20% Yếu, 60% Trung bình, 20% Giỏi [5]. Quá trình huấn luyện diễn ra qua 500 vòng lặp (episodes), mỗi vòng gồm 100 tác nhân.

B. Đánh giá so sánh (A/B Testing)

Chúng tôi so sánh hiệu quả giữa hai chính sách:

- **Nhóm Q-learning (Thực nghiệm):** Hành động dựa trên bảng Q-table đã huấn luyện.
- **Nhóm Historical Policy (Đối chứng):** Hành động dựa trên xác suất ngẫu nhiên mô phỏng lại thói quen của sinh viên các khóa trước.

Kết quả định lượng (Bảng II) cho thấy sự vượt trội rõ rệt của thuật toán Q-learning trên tất cả các chỉ số đo lường.

Table II
KẾT QUẢ SO SÁNH HIỆU SUẤT GIỮA HAI NHÓM

Chỉ số (Metrics)	Đối chứng	Thực nghiệm	Cải thiện (%)
Tổng phần thưởng tích lũy	88.4	389.6	**+340.8%**
Điểm số trung bình (thang 10)	6.25	7.66	**+22.5%**
Mức độ thành thạo LO (0-1)	0.58	0.66	**+13.9%**
Số kỹ năng yếu (Weak LOs)	3.02	1.48	**-51.0%**

C. Phân tích tác động sư phạm

Bên cạnh các con số thống kê, hệ thống thể hiện khả năng thích ứng thông minh với từng nhóm đối tượng, giải quyết được vấn đề "một kích cỡ cho tất cả":

1) Đối với nhóm sinh viên Yếu: Chỉ số quan trọng nhất là **Số kỹ năng yếu (Avg Weak LO Count)** đã giảm mạnh 51.0% (từ 3.02 xuống 1.48). [cite_start]Điều này chứng minh rằng AI đã nhận diện được các lỗ hổng kiến thức và chủ động gợi ý các hành động khắc phục (Remedial Actions) như xem lại bài giảng cũ, thay vì đẩy sinh viên vào làm các bài kiểm tra quá sức gây nản lòng [5]. Mức chênh lệch phần thưởng lớn nhất cũng được ghi nhận ở nhóm này, phản ánh chiến lược "khích lệ" của hàm thưởng.

2) Đối với nhóm sinh viên Giỏi: Nhóm này đạt điểm số trung bình tuyệt đối cao nhất (8.18/10). [cite_start]Hệ thống đã nhận diện năng lực vượt trội và chuyển sang chiến lược đề xuất các nội dung thách thức hơn (Future Actions) hoặc các bài tập khó, giúp duy trì hứng thú học tập và tránh sự nhàm chán [5].

3) Độ tin cậy thống kê: [cite_start]Kiểm định T-test độc lập giữa hai nhóm cho kết quả $T - statistic = 67.74$ và $P - value \approx 0.000$, khẳng định sự khác biệt về hiệu quả là có ý nghĩa thống kê với độ tin cậy 99.9% [5].

VI. Kết luận và Hướng phát triển

Bài báo đã trình bày một giải pháp toàn diện để cá nhân hóa giáo dục STEM thông qua việc ứng dụng Học tăng cường. Các đóng góp chính bao gồm: (1) Kiến trúc hệ thống mở dựa trên Microservices và LTI 1.3; (2) Quy trình mô hình hóa dữ liệu người học chi tiết; và (3) Thuật toán Q-learning với cơ chế thưởng thích ứng.

Kết quả thực nghiệm cho thấy hệ thống không chỉ cải thiện điểm số (+22.5%) mà quan trọng hơn là giúp lấp đầy các lỗ hổng kiến thức cho sinh viên yếu (-51% số kỹ năng yếu), hiện thực hóa mục tiêu "không ai bị bỏ lại phía sau".

Tuy nhiên, nghiên cứu vẫn còn hạn chế khi chưa được triển khai trên lớp học thực tế (Live Deployment) và không gian trạng thái bị giới hạn bởi phương pháp rời rạc hóa.

Hướng phát triển trong tương lai bao gồm: cite_start

- **Deep Reinforcement Learning (DRL):** Áp dụng mạng nơ-ron sâu (DQN, PPO) để xử lý không gian trạng thái liên tục và phức tạp hơn [6].
- **Triển khai thực tế:** Tích hợp hệ thống vào các khóa học STEM tại trường đại học để thu thập dữ liệu phản hồi thực và tinh chỉnh mô hình. [cite_start]
- **Federated Learning:** Nghiên cứu cơ chế học tập liên kết để bảo vệ quyền riêng tư dữ liệu người học khi triển khai trên nhiều cơ sở giáo dục [5].

Lời cảm ơn

Nhóm tác giả xin chân thành cảm ơn TS. Đỗ Như Tài đã hướng dẫn tận tình. Nghiên cứu được thực hiện tại Khoa Công nghệ Thông tin, Trường Đại học Sài Gòn.

References

[1] E. du Plooy et al., "Personalized adaptive learning in higher education: A scoping review," *Heliyon*, vol. 10, no. 21, p. e39630, 2024.

[2] C. B. Frey and M. A. Osborne, "The future of employment: How susceptible are jobs to computerisation?," *Tech. Forecast. Soc. Change*, vol. 114, pp. 254–280, 2017.

[3] A. Riedmann et al., "Reinforcement Learning in Education: A Systematic Literature Review," *Int. J. Artif. Intell. Educ.*, 2025.

[4] T. B. Thuan, "Ứng dụng machine learning dự báo sinh viên diện cảnh báo học tập," *Hue Uni. Journal of Science*, 2021.

[5] I. Gligorea et al., "Adaptive Learning Using Artificial Intelligence in e-Learning," *Educ. Sci.*, vol. 13, no. 12, 2023.

[6] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. MIT Press, 2018.

[7] IMS Global, "LTI 1.3 Implementation Guide," [Online]. Available: https://www.imsglobal.org/spec/lti/v1p3.