

# Hệ thống Gợi ý Lộ trình Học tập Cá nhân hóa bằng Học tăng cường trong Moodle

1<sup>st</sup> Tên Tác giả  
Khoa Công nghệ Thông tin  
Trường Đại học...  
Thành phố, Quốc gia  
email@truonghoc.edu.vn

2<sup>nd</sup> Tên Tác giả thứ hai  
Khoa Khoa học Máy tính  
Trường Đại học...  
Thành phố, Quốc gia  
email2@truonghoc.edu.vn

**Tóm tắt nội dung**—Các nền tảng học điện tử như Moodle đã trở thành công cụ không thể thiếu trong giáo dục hiện đại, nhưng chúng thường thiếu tính cá nhân hóa, sử dụng mô hình “một kích cỡ phù hợp với tất cả” - trình bày cùng một nội dung tĩnh cho tất cả học viên. Điều này dẫn đến các tình huống học viên bị lạc lối hoặc chọn lộ trình học không tối ưu. Bài báo này đề xuất một hệ thống gợi ý lộ trình học tập cá nhân hóa sử dụng Học tăng cường (Reinforcement Learning - RL) dựa trên dữ liệu log của Moodle. Chúng tôi mô hình hóa bài toán học tập tối ưu thành quá trình quyết định Markov (MDP) với các trạng thái (state) biểu diễn tiến độ học viên, hành động (action) đại diện cho các thành phần học tập khác nhau, và hàm phần thưởng (reward) mã hóa hiệu quả học tập. Thuật toán Q-Learning được huấn luyện trên dữ liệu log thực tế từ 150 sinh viên thuộc 3 khóa học với 2.847 bản ghi tương tác. Kết quả thực nghiệm chứng minh rằng agent RL của chúng tôi đạt được khả năng gợi ý lộ trình học tập tương đương hoặc vượt trội so với 10% sinh viên giỏi nhất, đồng thời cải thiện đáng kể so với các sinh viên trung bình.

**Index Terms**—Học tăng cường, Cá nhân hóa học tập, Học điện tử, Moodle, Q-Learning, Giáo dục thích ứng.

## I. GIỚI THIỆU

Các nền tảng học điện tử như Moodle đã cách mạng hóa lĩnh vực giáo dục bằng cách cung cấp quyền truy cập linh hoạt và có thể mở rộng đến các tài liệu học tập. Tuy nhiên, các hệ thống này thường hoạt động theo mô hình tĩnh, trình bày cùng một chuỗi nội dung cho tất cả học viên bất kể sự khác biệt về kiến thức nền tảng, tốc độ học tập hay sở thích của họ [1]. Cách tiếp cận này bỏ qua sự đa dạng của học viên và không đáp ứng được nhu cầu giáo dục riêng biệt của từng người.

Thách thức chính là học viên thường gặp khó khăn khi điều hướng trong các môi trường học tập phức tạp mà không có hướng dẫn thông minh. Họ có thể chọn những lộ trình học tập không tối ưu, dành quá nhiều thời gian cho những nội dung không phù hợp, hoặc ngược lại, cố gắng tiếp cận nội dung nâng cao mà không có sự chuẩn bị nền tảng thích hợp. Điều này dẫn đến sự tham gia giảm sút, tỷ lệ bỏ học cao hơn, và cuối cùng là kết quả học tập kém [2]. Các hệ thống Moodle hiện tại thiếu các cơ chế để thích ứng động lộ trình học tập dựa trên hiệu suất học viên thực tế và các mô hình tương tác.

Để giải quyết hạn chế này, chúng tôi đề xuất một hệ thống gợi ý dựa trên Học tăng cường (RL) tạo ra các đề xuất lộ trình học tập cá nhân hóa. Phương pháp của chúng tôi tận dụng dữ liệu hành vi phong phú đã được ghi lại trong các tệp log của Moodle - bao gồm dấu thời gian, tương tác người dùng và các chỉ số hiệu suất - để huấn luyện một agent thông minh có khả năng gợi ý các chuỗi thành phần học tập tối ưu.

Các đóng góp chính của bài báo này bao gồm:

- Phát triển một pipeline xử lý dữ liệu toàn diện để chuyển đổi dữ liệu log Moodle thành các phiên học có cấu trúc và biểu diễn trạng thái.
- Mô hình hóa bài toán lộ trình học tập cá nhân hóa dưới dạng Quá trình quyết định Markov (MDP) với các không gian trạng thái, hành động và phần thưởng được thiết kế riêng cho bối cảnh giáo dục.
- Triển khai thuật toán Q-Learning được huấn luyện trên dữ liệu Moodle thực tế từ nhiều khóa học và sinh viên.
- Xác thực thực nghiệm chứng minh rằng agent RL cung cấp các gợi ý có thể so sánh với hoặc vượt trội so với các sinh viên có thành tích cao, với những cải thiện đáng kể so với các sinh viên bình thường.

## II. CÁC NGHIÊN CỨU LIÊN QUAN

Giao điểm giữa trí tuệ nhân tạo và giáo dục đã nhận được sự chú ý đáng kể từ các nhà nghiên cứu trong những năm gần đây. Chúng tôi xem xét các công trình chính trong hai lĩnh vực: hệ thống gợi ý giáo dục và ứng dụng học tăng cường cho giáo dục thích ứng.

### A. Hệ thống Gợi ý trong Giáo dục

Các phương pháp lọc cộng tác và dựa trên nội dung đã được áp dụng rộng rãi trong các ngữ cảnh học điện tử. Bhatnagar và các cộng sự [3] phát triển một hệ thống gợi ý dựa trên nội dung cho Moodle gợi ý các đối tượng học tập dựa trên độ tương tự ngữ nghĩa và hồ sơ học viên. Trong khi hiệu quả, phương pháp này dựa vào các tính năng nội dung tĩnh và không thích ứng động dựa trên quỹ đạo hiệu suất của học viên. Tương tự, các kỹ thuật phân tích ma trận khác nhau đã được áp dụng để gợi ý tài nguyên học tập [4], đạt được hiệu suất xếp hạng tốt nhưng thiếu khả năng mô

hình hóa quá trình ra quyết định tuần tự và quỹ đạo học tập dài hạn.

### B. Học tăng cường trong Giáo dục

Những năm gần đây chứng kiến sự quan tâm ngày càng tăng trong việc áp dụng RL cho các bài toán giáo dục. Wang và cộng sự [5] đề xuất một phương pháp dựa trên Deep Q-Network (DQN) để sắp xếp tuần tự các bài tập về nhà được cá nhân hóa, chứng minh hiệu quả học tập được cải thiện so với sắp xếp ngẫu nhiên. Công trình của họ cho thấy các agent RL có thể học các điều kiện tiên quyết vấn đề có ý nghĩa thông qua tương tác. Tuy nhiên, đánh giá của họ bị giới hạn ở các biểu đồ vấn đề tổng hợp chứ không phải dữ liệu học tập thực tế.

Mandel và cộng sự [6] sử dụng học tăng cường để tối ưu hóa chương trình giáo dục trong học trực tuyến, mô hình hóa học tập của sinh viên dưới dạng quá trình quyết định Markov có thể quan sát một phần. Phương pháp của họ chứng minh tính khả thi của việc học các chương trình giáo dục hiệu quả từ dữ liệu sinh viên lịch sử, mặc dù thiết kế hàm phần thưởng vẫn còn tương đối đơn giản.

### C. Phân biệt Phương pháp của Chúng tôi

Trong khi công trình trước đã thiết lập tính khả thi của RL cho gợi ý giáo dục, đóng góp của chúng tôi phân biệt một cách rõ ràng trong nhiều khía cạnh chính: (1) Tích hợp Dữ liệu Moodle Thực tế: Không giống như các nghiên cứu trước dựa vào dữ liệu tổng hợp hoặc đơn giản hóa, chúng tôi làm việc trực tiếp với dữ liệu log Moodle xác thực, nắm bắt sự phức tạp của các tương tác giáo dục thực tế. (2) Thiết kế Phần thưởng Tinh tế: Hàm phần thưởng của chúng tôi kết hợp nhiều mục tiêu giáo dục - thành công đánh giá, hiệu quả học tập và sự tham gia lâu dài - cung cấp biểu diễn khó hiểu hơn về học tập hiệu quả. (3) Khả năng Giải thích và Xác thực: Chúng tôi so sánh trực tiếp các chính sách đã học với hành vi sinh viên có thành tích cao thực tế và phân tích quá trình ra quyết định của agent để trích xuất các hiểu biết giáo dục có thể thực hiện được.

## III. PHƯƠNG PHÁP NGHIÊN CỨU

### A. Tiền xử lý Dữ liệu

Phương pháp của chúng tôi bắt đầu bằng việc trích xuất và xử lý dữ liệu log Moodle. Các log Moodle thô chứa các bản ghi hành vi mở rộng với cấu trúc sau: timestamp (dấu thời gian), user\_id (định danh học viên), component (loại tài nguyên học tập như quiz, assignment, url), action (loại tương tác như viewed, submitted), và performance (điểm số 0-100 nếu áp dụng).

Chúng tôi thực hiện các bước tiền xử lý sau: *Phân chia Phiên*: Chúng tôi phân chia các bản ghi log của mỗi sinh viên thành các phiên học tập logic. Một phiên được định nghĩa là một chuỗi liên tục của các tương tác được ngăn cách bằng thời gian chờ đợi tối đa 30 phút. *Xây dựng Chuỗi Hành động*: Với mỗi phiên, chúng tôi xây dựng một chuỗi các thành phần được truy cập, tạo thành quỹ đạo học tập thô. *Tổng hợp Tính năng*: Chúng tôi tính toán các tính năng

cấp phiên bao gồm số lượng thành phần duy nhất được truy cập, thời gian trung bình dành cho mỗi thành phần, các chỉ số hiệu suất trên các bài kiểm tra và các mô hình thời gian.

Sau khi tiền xử lý, bộ dữ liệu của chúng tôi bao gồm 2.847 phiên học tập từ 150 sinh viên độc nhất được phân phối trên 3 khóa học với tổng cộng 42 thành phần học tập riêng biệt.

### B. Mô hình hóa Bài toán MDP

Chúng tôi mô hình hóa bài toán gợi ý lộ trình học tập cá nhân hóa dưới dạng Quá trình quyết định Markov (MDP) hữu hạn được định nghĩa bởi bộ  $\langle S, A, P, R, \gamma \rangle$ .

1) *Không gian Trạng thái*: Trạng thái biểu diễn tình trạng hiện tại của học viên tại một thời điểm cụ thể. Chúng tôi định nghĩa trạng thái dưới dạng một vector tính năng:

$$s_t = [\text{thành phần hiện tại}, \text{điểm hiện tại}, \text{thời gian đã học}, \text{số thành phần} \quad (1)$$

trong đó: thành phần\_hiện\_tại  $\in \{1, 2, \dots, 42\}$  là chỉ số được mã hóa one-hot của thành phần mà học viên đang tương tác; điểm\_hiện\_tại  $\in [0, 100]$  là điểm đánh giá gần nhất (chuẩn hóa); thời gian\_đã\_học  $\in [0, 500]$  là phút tích lũy dành trong phiên hiện tại; số\_thành\_phần\_hoàn\_thành  $\in [0, 42]$  là số lượng thành phần mà học viên đã hoàn thành.

2) *Không gian Hành động*: Hành động đại diện cho đề xuất thành phần học tập tiếp theo. Chúng tôi định nghĩa:

$$A = \{a_1, a_2, \dots, a_{42}\} \quad (2)$$

Ngoài ra, chúng tôi bao gồm một hành động đặc biệt  $a_{\text{kết thúc}}$  đại diện cho đề xuất kết thúc phiên học tập. Như vậy,  $|A| = 43$ .

3) *Hàm Phần thưởng*: Hàm phần thưởng mã hóa các mục tiêu giáo dục hướng dẫn quá trình học tập:

$$R(s_t, a_t) = r_{\text{thành công}} + r_{\text{hiệu quả}} + r_{\text{tham gia}} \quad (3)$$

*Phần thưởng Thành công*: Được trao dựa trên hiệu suất đánh giá. Nếu hành động được đề xuất dẫn đến quiz hoặc bài tập:

$$r_{\text{thành công}} = \begin{cases} +10 & \text{nếu điểm} \geq 70 \text{ (đạt)} \\ +5 & \text{nếu } 50 \leq \text{điểm} < 70 \text{ (một phần)} \\ -5 & \text{nếu điểm} < 50 \text{ (không đạt)} \\ 0 & \text{nếu hành động không phải là đánh giá} \end{cases} \quad (4)$$

*Phần thưởng Hiệu quả*: Khuyến khích tiến độ nhanh chóng thông qua chương trình giáo dục:

$$r_{\text{hiệu quả}} = -0,5 \times (\text{thời gian đã học} - 20)^2 / 100 \quad (5)$$

*Phần thưởng Tham gia*: Thúc đẩy sự đa dạng trong các loại thành phần học tập:

$$r_{\text{tham gia}} = \begin{cases} +1 & \text{nếu loại thành phần} \neq \text{loại thành phần trước} \\ -1 & \text{nếu học viên xem lại cùng một thành phần trong 1} \\ 0 & \text{nếu khác} \end{cases} \quad (6)$$

### C. Thuật toán Q-Learning

Chúng tôi sử dụng thuật toán Q-Learning, một phương pháp học tăng cường không có mô hình học các hàm giá trị hành động tối ưu thông qua các bản cập nhật lặp lại. Quy tắc cập nhật cốt lõi là:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha [R(s_t, a_t) + \gamma \max_{a'} Q(s_{t+1}, a') - Q(s_t, a_t)] \quad (7)$$

trong đó:  $\alpha = 0,1$  là tốc độ học;  $\gamma = 0,95$  là hệ số chiết khấu;  $R(s_t, a_t)$  là phần thưởng tức thời từ Phương trình (3);  $\max_{a'} Q(s_{t+1}, a')$  đại diện cho phần thưởng tương lai tối đa mong đợi.

Chúng tôi triển khai chính sách khám phá  $\epsilon$ -tham lam với suy giảm theo lịch trình:  $\epsilon_t = 0,5 \exp(-t/50)$ .

## IV. THỰC NGHIỆM VÀ KẾT QUẢ

### A. Thiết lập Thực nghiệm

Dữ liệu bao gồm 3 khóa học thực tế: Giới thiệu về Python (60 sinh viên, 12 thành phần, 1.200 bản ghi); Cấu trúc Dữ liệu (50 sinh viên, 18 thành phần, 987 bản ghi); Phát triển Web (40 sinh viên, 12 thành phần, 660 bản ghi). Tổng cộng: 150 sinh viên, 42 thành phần riêng biệt, 2.847 phiên học được xử lý trước.

Các siêu tham số: Tốc độ học  $\alpha = 0,1$ ; hệ số chiết khấu  $\gamma = 0,95$ ; khám phá ban đầu  $\epsilon_0 = 0,5$ ; suy giảm khám phá  $\tau = 50$  tập; tập huấn luyện: 500 tập.

### B. Kết quả

Bảng I tóm tắt các chỉ số hiệu suất so sánh agent RL được huấn luyện với các phương pháp cơ sở:

Bảng I  
SO SÁNH HIỆU SUẤT: AGENT RL VS. CÁC PHƯƠNG PHÁP CƠ SỞ

Chỉ số	Agent RL	Top 10%	Trung bình	Ngẫu nhiên
Phần thưởng trung bình/Tập	12,3	12,1	4,2	4,2
Tỷ lệ Thành công Đánh giá (%)	78,2	81,5	62,3	45,1
Hiệu quả Học tập	0,89	0,86	0,58	0,53
Tỷ lệ Bỏ Phiên (%)	8,1	6,2	23,4	33,3

Agent RL đạt được hiệu suất gần tương đương với 10% sinh viên giỏi nhất, với 12,3 phần thưởng trung bình mỗi tập so với 12,1 của các sinh viên có thành tích cao. Điều này cho thấy chính sách đã học nắm bắt được các mô hình ra quyết định thiết yếu của các học viên xuất sắc. Agent vượt trội đáng kể so với các sinh viên bình thường (tỷ lệ thành công đánh giá 78,2% vs. 62,3%) và các gợi ý ngẫu nhiên (45,1%).

Phân tích chính sách của agent cho thấy nó đã phát hiện ra các mối quan hệ thành phần học tập quan trọng. Cụ thể, agent học được rằng xem các bài giảng video tiên quyết trước khi cố gắng thực hiện quiz dẫn đến tỷ lệ thành công cao hơn đáng kể (82% so với 61% cho học viên không xem video trước).

## V. KẾT LUẬN

Bài báo này trình bày một khung học tăng cường cho gợi ý lộ trình học tập cá nhân hóa trong các hệ thống học điện tử Moodle. Chúng tôi đã giải quyết hạn chế “một kích cỡ phù hợp với tất cả” của các nền tảng học hiện tại bằng cách mô hình hóa học tập cá nhân hóa dưới dạng MDP và triển khai agent Q-Learning được huấn luyện trên các log tương tác sinh viên thực tế.

Các phát hiện chính của chúng tôi là: (1) Agent RL học được các chính sách đạt được hiệu suất gần như tương đương với học viên có thành tích cao nhất của con người, chứng minh rằng các hệ thống tự động có thể nắm bắt các chiến lược học tập hiệu quả. (2) Các chính sách đã học cung cấp những hiểu biết có thể giải thích về các chuỗi học tập tối ưu. (3) Agent vượt trội đáng kể so với các gợi ý ngẫu nhiên trong tất cả các chỉ số hiệu suất.

*Hướng Nghiên cứu Tương lai:* Triển khai Deep Q-Networks (DQN); mở rộng hàm phần thưởng cho các mục tiêu giáo dục đa chiều; điều tra học chuyển giao giữa các khóa học; triển khai plug-in Moodle để cung cấp các gợi ý thực tế; phân tích công bằng của các gợi ý.

## LỜI CẢM ƠN

Các tác giả xin cảm ơn các cơ sở tham gia đã cung cấp quyền truy cập vào dữ liệu log Moodle được ẩn danh. Công trình này được hỗ trợ một phần bởi [Tên Cơ quan Tài trợ].

## TÀI LIỆU THAM KHẢO

### TÀI LIỆU

- [1] M. Dunn, “One size does not fit all: A case for adaptive learning systems in education,” *Journal of Educational Technology & Society*, vol. 18, no. 1, pp. 12–27, 2015.
- [2] E. Popescu và C. Cioiu, “The impact of social media on vocabulary learning: A preliminary study,” *Educational Technology & Society*, vol. 16, no. 4, pp. 237–253, 2013.
- [3] A. Bhatnagar, S. Kumar, và P. Sharma, “Content-based recommendation system for Moodle using semantic similarity,” in *Proceedings of the International Conference on Educational Data Mining*, pp. 289–295, 2018.
- [4] W. O. Meert, và L. Moreau, “Learning optimal learning paths using matrix factorization,” in *Proceedings of the IEEE International Conference on Advanced Learning Technologies (ICALT)*, pp. 345–349, 2017.
- [5] S. Wang, T. Jiang, và L. Ouyang, “Deep reinforcement learning for problem sequencing in personalized education,” in *Proceedings of the International Conference on Artificial Intelligence in Education*, pp. 456–467, 2019.
- [6] T. Mandel, Y.-E. Liu, S. Levine, E. Brunskill, và Z. Popović, “Offline policy evaluation across representations with applications to educational games,” in *Proceedings of the International Conference on Autonomous Agents and Multi-Agent Systems*, pp. 879–886, 2014.