Nghiên cứu ứng dụng thuật toán học tăng cường xây dựng hệ thống gợi ý cho các môn học thi trắc nghiệm tại trường Đại học Kinh tế quốc dân

*Trần Thị Mỹ Diệp, Mai Vĩnh Khang, Nguyễn Diệp Anh, Bùi Quang Huy, Trần Duy Việt, Hồ Thị Minh Ngân, Phạm Xuân Lâm*

Khoa Công nghệ thông tin, Trường Đại học Kinh tế quốc dân

***Tóm tắt****: Trong bối cảnh chuyển đổi số mạnh mẽ và sự phát triển nhanh chóng của trí tuệ nhân tạo, việc cải thiện chất lượng học tập thông qua các phương pháp tiếp cận mới ngày càng trở nên quan trọng. Nghiên cứu này đề xuất việc ứng dụng một hệ thống gợi ý thông minh sử dụng kỹ thuật học tăng cường để hỗ trợ sinh viên trong các môn học thi trắc nghiệm tại Trường Đại học Kinh tế Quốc dân. Hệ thống này không chỉ cá nhân hóa quá trình học tập, mà còn tăng cường sự hứng thú và động lực học tập của sinh viên bằng cách cung cấp các gợi ý và phản hồi thông minh dựa trên kết quả làm bài. Bên cạnh đó, hệ thống giúp giảng viên tối ưu hóa quá trình giảng dạy và quản lý, cung cấp thông tin chi tiết về điểm mạnh và điểm yếu của sinh viên. Nghiên cứu đã thực hiện thiết kế hệ thống, và đang trong bước thử nghiệm cho môn học “Nhập môn công nghệ thông tin” cho sinh viên kỳ mùa Thu năm học 2024-2025 tại trường Đại học Kinh tế quốc dân. Kết quả thử nghiệm ban đầu cho thấy hệ thống gợi ý hoạt động tốt và có tiềm năng trong việc ứng dụng để cải thiện kết quả học tập và sự tham gia của sinh viên.*

***Từ khoá****: Cá nhân hóa học tập, Gợi ý học tập, Hệ thống gợi ý, Học tăng cường, Thi trắc nghiệm*

# Giới thiệu

Hiện nay, 100% các môn học thuộc khối kiến thức cơ sở ở Trường Đại học Kinh tế Quốc dân đã áp dụng hình thức thi trắc nghiệm trên máy tính trong bài kiểm tra kết thúc học phần; bên cạnh đó, một số giảng viên cũng áp dụng hình thức làm bài này trong việc kiểm tra đánh giá giữa kỳ cho sinh viên. Việc chuyển đổi các môn học từ hình thức thi tự luận truyền thống sang hình thức thi trắc nghiệm trên máy tính mang lại nhiều tiện ích (Scalise & Gifford, 2006) (DeSouza & Fleming, 2003), song một trong những thách thức lớn nhất của hình thức kiểm tra này là làm thế nào để có thể thực sự phản ánh đúng năng lực học tập của sinh viên và giúp họ cải thiện các nội dung kiến thức họ chưa nắm vững (Wood và cộng sự, 2022).

Nhằm hỗ trợ người học chuẩn bị cho các bài kiểm tra theo hình thức trắc nghiệm, nhiều giảng viên đã tích hợp các câu hỏi trắc nghiệm vào bài giảng trên lớp hoặc đưa chúng lên hệ thống quản lý học tập (LMS) để sinh viên có thể tự ôn luyện tại nhà. Tuy nhiên, nhìn chung, cách tiếp cận này chưa thực sự hiệu quả, bởi việc ôn tập chưa có mức độ cá nhân hóa cao. Các bài tập trắc nghiệm thông thường chỉ cung cấp phản hồi đơn giản như "đúng" hoặc "sai", thiếu đi sự tương tác hoặc giải thích chi tiết (Ryan và cộng sự, 2020). Điều này khiến sinh viên không được giải thích rõ ràng về những lỗi sai mình đang gặp phải để khắc phục, đồng thời có ít cơ hội học hỏi từ sai lầm do các câu hỏi được hiển thị một cách ngẫu nhiên từ bộ câu hỏi được soạn trước. Nhìn chung, các hệ thống hay nền tảng hỗ trợ ôn tập trắc nghiệm thông thường chưa có chức năng phân tích quá trình học tập của sinh viên một cách liên tục, và từ đó không thể thường xuyên đưa ra những đề xuất hay phản hồi cụ thể giúp sinh viên cải thiện nhược điểm trong việc ôn tập của mình. Hơn nữa, việc thiếu các yếu tố khuyến khích và động lực trong quá trình làm bài trắc nghiệm cũng là một vấn đề đáng lưu ý. Nếu ngân hàng câu hỏi không đủ phong phú hoặc không có sự thay đổi trong cách đặt câu hỏi, việc ôn tập có thể mang tính lặp lại, dễ khiến người học nhàm chán và chỉ tập trung vào việc ghi nhớ câu trả lời thay vì hiểu rõ bản chất của vấn đề (Elpidorou, 2023). Ngoài ra, nếu sinh viên chỉ tập trung vào việc đạt điểm cao trong các bài trắc nghiệm mà không thực sự hiểu sâu kiến thức, việc học trở thành một nhiệm vụ áp lực thay vì một trải nghiệm học tập thú vị (Mingo và cộng sự, 2018).

Khác với các nền tảng học tập trực tuyến thông thường, một hệ thống học tập tích hợp tính năng gợi ý bài tập sẽ góp phần nâng cao trải nghiệm của sinh viên. Người học không cần thiết phải hoàn thiện tất cả các bài tập mà giảng viên cung cấp mới có thể nắm vững các khái niệm và thành thạo các kỹ năng của môn học; thay vào đó, họ chỉ cần làm các bài tập được đề xuất dựa trên mức độ phù hợp với năng lực của từng cá nhân. Điều này thúc đẩy sự hào hứng của sinh viên với việc học, từ đó đóng vai trò quan trọng trong việc cải thiện kết quả học tập của sinh viên. Tuy nhiên tại Đại học Kinh tế Quốc dân (NEU), chưa có hệ thống gợi ý nội dung học tập nào được áp dụng trên nền tảng quản l‎ý học tập LMS của trường (theo Trung tâm Ứng dụng Công nghệ Thông tin – CAIT).

Nhận thấy thực tiễn tại Đại học Kinh tế Quốc dân và tầm quan trọng của vấn đề cá nhân hóa học tập, nhóm nghiên cứu đã phát triển một hệ thống gợi ý thông minh dựa trên kỹ thuật Học tăng cường cho các môn học thi trắc nghiệm, trước mắt thực nghiệm với học phần “Nhập môn Công nghệ thông tin” thuộc Khoa Công nghệ thông tin. Hệ thống gợi ý này sẽ được thử nghiệm trong môi trường học tập thực tế tại trường với sự tham gia của hai nhóm đối tượng: sinh viên và giảng viên. Nhóm sinh viên sẽ được yêu cầu đánh giá hiệu quả của hệ thống thông qua bộ câu hỏi khảo sát, trong khi nhóm giảng viên sẽ cung cấp phản hồi về cách hệ thống hỗ trợ và cải thiện quá trình giảng dạy. Dữ liệu từ hai nhóm sẽ đóng vai trò quan trọng trong việc đánh giá hiệu quả và tính khả thi của hệ thống, giúp nhóm nghiên cứu hiểu rõ hơn về cách hệ thống hỗ trợ người học và giảng dạy. Hiệu quả của hệ thống trong việc cải thiện điểm số sinh viên sẽ được đánh giá bằng phương pháp thử nghiệm A/B, so sánh giữa các nhóm sinh viên có và không sử dụng hệ thống. Phản hồi từ sinh viên sẽ được thu thập để tối ưu hóa hệ thống, và hiệu quả của thuật toán học tăng cường sẽ được đánh giá nhằm đóng góp vào sự phát triển của các hệ thống hỗ trợ học tập trong tương lai.

# Tổng quan lý thuyết

Với sự phát triển mạnh mẽ của giáo dục trực tuyến, các nguồn tài liệu học tập đang ngày một phong phú, đa dạng. Việc học tập trực tuyến vì thế trở nên thuận lợi hơn cho người học. Tuy nhiên, đứng trước nguồn học liệu dường như vô tận này, người học phải đối mặt với một thách thức không nhỏ: làm thế nào để tìm được tài liệu phù hợp với nhu cầu kiến thức hay trình độ của bản thân? (Adil và cộng sự, 2024). Đôi khi, việc tìm kiếm và truy vấn thông tin đơn thuần không thể đáp ứng được trọn vẹn nhu cầu của từng cá nhân, thậm chí còn có thể gây lãng phí thời gian không cần thiết, phần nào làm giảm hứng thú và động lực học tập ở người học (Chitrakar & Nisanth, 2023). Chính vì vậy, việc xây dựng hệ thống gợi ý bài tập phù hợp với từng cá nhân đã trở thành một nhu cầu thiết yếu, giúp người học tiết kiệm thời gian và cải thiện hiệu quả học tập, nhất là trong bối cảnh việc làm hết lượng lớn bài tập có sẵn không phù hợp với mọi người.

## Hệ thống gợi ý và thuật toán gợi ý

Các hệ thống gợi ý sử dụng kĩ thuật gợi ý phổ biến như: lọc theo nội dung (Content-based Filtering), lọc cộng tác (Collaborative Filtering), lọc hỗn hợp (Hybrid Filtering), học sâu (Deep Learning) và Học tăng cường (Reinforcement Learning) cùng một số kỹ thuật khác đã được sử dụng với bộ dữ liệu học tập của học sinh được thu thập từ nền tảng học tập trực tuyến. Huang và cộng sự (2019) đã đưa ra DRE - một framework tân tiến áp dụng kỹ thuật Học tăng cường sâu (Deep Reinforcement Learning) với ba mục tiêu khác nhau: giúp học sinh ôn tập & khám phá, đề xuất các bài tập với độ khó phân bố hợp lí và duy trì sự hứng thú của họ khi luyện tập. Hay như nghiên cứu của Li và cộng sự (2021) trình bày một phương pháp khác sử dụng hồi quy tuyến tính, thuật toán EM và K-nearest neighbor để đề xuất nội dung luyện tập phù hợp. Hiệu quả của phương pháp này đã được chứng minh khi các học sinh có kết quả học tập được cải thiện rõ rệt. Một mô hình gợi ‎ý ‎lai đã được đưa ra bởi Gong and Yao (2019), kết hợp giữa hai thành phần: deep collaborative filtering (DeepCF) và tuyến tính rộng (wide linear). DeepCF sử dụng bộ giải mã tự động khử nhiễu xếp chồng (SDAE) để học các đặc trưng tiềm ẩn và dự đoán đánh giá của học sinh. Thành phần tuyến tính rộng tích hợp các thuộc tính meta của bài tập như độ khó, xếp loại, và các thành phần kiến thức; sau đó, hai thành phần này được kết hợp bằng phương pháp tuyến tính. Kết quả thử nghiệm đã cho thấy mô hình đề xuất này đã cải thiện 10% chỉ số AUC so với phương pháp lọc cộng tác truyền thống.

Về các nghiên cứu có liên quan trong nước, nghiên cứu của Nghe and Nhàn (2013) đề cập đến ứng dụng kỹ thuật Phân rã ma trận thiên vị (Biased Matrix Factorization) trong dự đoán kết quả học tập của sinh viên, từ đó đưa ra những gợi ý phù hợp cho họ trong việc lựa chọn các môn tự chọn. Nhàn và cộng sự của mình tiến hành xử lý dữ liệu điểm của sinh viên, chuyển chúng theo định dạng phù hợp cho giải thuật BMF, tìm kiếm siêu tham số và đánh giá, so sánh với các giải thuật khác. Nghiên cứu của Vu và cộng sự (Vu và cộng sự, 2020) sử dụng phương pháp Bản thể học (Ontology), thu thập các thông tin về sở thích, vòng tròn bạn bè, mối quan hệ xã hội,... xoay quanh từng học sinh trung học để đưa ra gợi ý cho học sinh đó về việc lựa chọn trường Đại học và chuyên ngành.

## Thuật toán học tăng cường và ứng dụng thuật toán học tăng cường trong giáo dục

Kĩ thuật học tăng cường (Reinforcement Learning) là một trong ba mô hình học máy cơ bản, bên cạnh học có giám sát (Supervised Learning) và học không giám sát (Unsupervised Learning); kĩ thuật này dựa trên việc các tác nhân phần mềm nên thực hiện hành động nào trong một môi trường nhằm tối đa hóa phần thưởng tích lũy (Mahesh, 2020). Mỗi khi hệ thống học tăng cường tương tác với môi trường, đầu tiên nó sẽ nhận đầu vào là trạng thái môi trường s, sau đó là đầu ra với một hành động a theo cơ chế suy luận nội bộ (Qiang & Zhongli, 2011). Cuối cùng, môi trường sẽ thay đổi sang trạng thái **s’** sau khi chấp nhận hành động. Hệ thống nhận đầu vào của trạng thái mới **s’** và nhận được tín hiệu phần thưởng và hình phạt r của môi trường cho hệ thống. Tóm lại, một tác nhân học tăng cường (Reinforcement Learning Agent) có mục tiêu là học được cách tốt nhất để hoàn thành một nhiệm vụ cụ thể thông qua các tương tác lặp đi lặp lại của nó với môi trường (Sutton, 2018). Kĩ thuật này khác với kĩ thuật học có giám sát ở chỗ các cặp vào/ ra (input/ output) không được đưa ra, đồng thời các hành động không tối ưu cũng không được chỉnh sửa một cách rõ ràng (Das và cộng sự, 2015).

Hiện tại, kĩ thuật học tăng cường có một số ứng dụng quan trọng trong các lĩnh vực chăm sóc sức khỏe, robot và điều khiển tự động, truyền thông và mạng, xử l‎ý ngôn ngữ tự nhiên, trò chơi và hệ thống tự tổ chức, quản lý lịch trình và cấu hình tài nguyên, hệ thống Internet kết nối vạn vật và thị giác máy tính (Naeem và cộng sự, 2020). Tuy nhiên, học tăng cường không chỉ có những ứng dụng nêu trên, nó là một kĩ thuật đầy hứa hẹn và là tương lai của trí tuệ nhân tạo với một số những ứng dụng mới trong các lĩnh vực: giáo dục (Liu và cộng sự, 2014; Mandel và cộng sự, 2014), tài chính (Deng và cộng sự, 2016; Khandani và cộng sự, 2010), quản trị kinh doanh (Li và cộng sự, 2015; Silver và cộng sự, 2013) và phát triển xe tự lái (Bojarski, 2016; Bojarski và cộng sự, 2017).

Các thuật toán và phương pháp học tăng cường đã được sử dụng rộng rãi trong nhiều hướng nghiên cứu khác nhau trong việc ứng dụng vào giáo dục (Fahad Mon và cộng sự, 2023). Nghiên cứu của (He-Yueya & Singla, 2021) đã tìm hiểu cách sử dụng chính sách dựa trên học tăng cường để đưa ra câu hỏi cho học sinh và tìm hiểu mức độ kiến thức của họ. Một nghiên cứu khác của (Barnes & Stamper, 2008) đã sử dụng khung MDP (Quá trình Quyết định Markov) để tự động tạo ra các gợi ý logic dựa trên dữ liệu của học sinh. Bên cạnh đó, kĩ thuật học tăng cường có thể được sử dụng để tạo ra các nội dung giáo dục. (Minoofam và cộng sự, 2022) đã phát triển RALF, một khung học tăng cường thích ứng dựa trên Automata Học Tế bào (Cellular Learning Automata – CLA) để tự động tạo ra các nội dung học tập cho học sinh mắc chứng khó đọc (Dyslexia). Với việc công nghệ thông tin và truyền thông (ICT) ngày càng trở nên quan trọng trong giáo dục, (Fok & Ip, 2004) ‎đã đưa ra một hệ thống cá nhân hóa giáo dục (PES) nhằm tận dụng lượng tài nguyên khổng lồ trên Internet. Việc giáo dục cá nhân hóa bao gồm việc xác định và tìm hiểu nhu cầu cũng như năng lực của từng học sinh trước khi áp dụng và triển khai các phương pháp giảng dạy và đưa ra nội dung học tập hiệu quả nhất (Fahad Mon et al., 2023).

Kĩ thuật học tăng cường cũng có thể được sử dụng để tối ưu hóa trình tự giảng dạy (instructional sequence) như trong nghiên cứu của (Atkinson, 1972); và việc này có thể ảnh hưởng tới mức độ hiệu quả trong học tập (Ritter và cộng sự, 2007). Nghiên cứu của (Tzeng và cộng sự, 2023) đã phát triển một hệ thống sử dụng học tăng cường với phương pháp Actor – Critic để gợi ‎ý bài tập cho sinh viên; trong đó, Actor sẽ gợi ý các bài tập phù hợp cho sinh viên dựa trên trạng thái học tập hiện tại của họ, còn Critic sẽ đánh giá quyết định của Actor dựa trên sự cải thiện trong điểm số của sinh viên, giúp cải thiện các gợi ý trong tương lai. Quá trình đề xuất của hệ thống áp dụng Quá‎ trình Quyết định Markov (Markov Decision Process), bao gồm các trạng thái, hành động và phần thưởng nhằm điều chỉnh và tối ưu hóa việc gợi ‎ý bài tập. Các sinh viên sử dụng hệ thống có tỉ lệ hoàn thành bài tập và điểm kiểm tra giữa kì cao hơn những sinh viên không sử dụng, đồng thời họ cũng cho phản hồi tích cực về tính hữu dụng và hiệu quả của hệ thống và mong muốn được tiếp tục sử dụng trong tương lai.

### Phương pháp Actor – Critic

Actor-Critic là một trong những phương pháp phổ biến trong Học tăng cường (Reinforcement Learning), kết hợp cả hai phương pháp chính sách (policy-based) và phương pháp giá trị (value-based) để cải thiện hiệu suất học tập (Tzeng et al., 2023). Trong Actor-Critic, có hai thành phần chính: Actor (tác nhân) và Critic (nhà phê bình). Phương pháp Actor-Critic kết hợp ưu điểm của phương pháp critic và actor-only, cho phép tạo ra các hành động liên tục. Actor nhận trạng thái hiện tại làm đầu vào và trả về tham số cho hàm chấm điểm. Tác nhân gợi ý (recommender agent) chấm điểm và chọn các mục. Critic học hàm giá trị (Q value) để đánh giá sự phù hợp giữa hành động và trạng thái. Actor cập nhật tham số chính sách (policy) dựa trên đánh giá của critic.

# Giải pháp gợi ý cho các môn học thi trắc nghiệm

## Đặt vấn đề

Hệ thống gợi ý bài tập trắc nghiệm nhóm nghiên cứu đề xuất bao gồm |U| (sinh viên) và |E| (bài tập), quá trình học tập của học sinh u được ghi lại là u = {(e1, p1), ..., (et, pt)}}. Ở đây, et là bài tập thực hiện tại thời điểm t, pt là câu trả lời cho bài et của sinh viên đó (pt = 1 nếu trả lời đúng, ngược lại pt = 0). Mỗi bài tập e được mô tả bằng bộ ba e = {d, k, b}, trong đó d là độ khó, k là phạm trù kiến thức, và b là trạng thái đánh dấu xác định người học nhận thấy bài tập này cần được ôn luyện nhiều hơn (b = 1 nếu bài tập được đánh dấu, b = 0 nếu không).

Trong nghiên cứu này, nhóm mô hình hóa việc đề xuất bài tập như một Quá trình Quyết định Markov (MDP). Quá trình bao gồm trạng thái, hành động và phần thưởng như sau:

* **Trạng thái S (state):** Biến S mô tả đặc điểm của câu hỏi trắc nghiệm mà người học đang thực hành tại thời điểm t: bao gồm
* **Hành động A (action):** Biến A là không gian hành động bao gồm tất cả các bài tập của môn học. Thực hiện hành động tương đương gợi ý bài học .
* **Phần thưởng R (reward):** là hàm phần thưởng. Hàm sẽ tính toán phần thưởng khi hệ thống gợi ý thực hiện hành động từ trạng thái dựa trên kết quả trả lời của người học.
* **Xác suất chuyển tiếp P (Probability transition):** thể hiện xác suất chuyển tiếp từ trạng thái St sang trạng thái St+1 sau khi thực hiện hành động At.

## Mô tả hàm phần thưởng

Hàm phần thưởng đóng vai trò quan trọng trong quy trình huấn luyện mô hình, giúp cải thiện chính sách gợi ý. Nhóm quyết định thiết kế hàm phần thưởng kết hợp 3 tiêu chí là độ khó, phạm trù kiến thức và mức độ chủ động ôn luyện lại bài tập. Phương trình hàm phần thưởng được mô tả bằng công thức dưới đây:

* Mục tiêu đánh giá R1 là khuyến khích gợi ý những bài tập có độ khó cao hơn nếu họ đã làm đúng. Nếu học sinh không làm đúng, hệ thống sẽ tiếp tục gợi ý các câu hỏi có độ khó dễ hơn để giúp người học củng cố lại kiến thức.
* Mục tiêu đánh giá R2 nhằm khuyến khích người học cải thiện kiến thức của họ. Trong trường hợp hệ thống gợi ý các câu hỏi thuộc phạm trù kiến thức khác nhưng người học vẫn trả lời sai, hệ thống cần gợi ý những câu thuộc phạm trù kiến thức cũ để người học có thể củng cố lại kiến thức.
* Mục tiêu đánh giá R3 khuyến khích việc gợi ý độ khó một cách ổn định, không thay đổi quá lớn giữa các lần học. Việc thay đổi đột ngột độ khó (tăng hoặc giảm đột ngột) có thể gây mất ổn định trong việc học (Tzeng et al., 2023). Hàm phần thưởng âm khi sự khác biệt giữa độ khó ở hai thời điểm t và t+1 càng lớn, khuyến khích học sinh tiến bộ từng bước thay vì thay đổi đột ngột.
* Trước đây, việc đánh dấu một bài tập được xem là dấu hiệu của việc học sinh cần ôn luyện bài tập nhiều hơn. Tuy nhiên, nếu việc đánh dấu bài tập được xem là hành động tích cực (ví dụ: để lưu lại những bài tập khó cần xem lại sau), thì việc thưởng khi đánh dấu sẽ khuyến khích học sinh có thái độ chủ động và chiến lược học tập rõ ràng hơn.

Ta có phương trình hàm phần thưởng như sau:

Hàm phần thưởng tổng hợp này cân bằng giữa việc khuyến khích tiến bộ, ổn định và sự nghiêm túc trong học tập, bằng cách sử dụng trọng số 0.25 cho mỗi phần thưởng con, nó đảm bảo các yếu tố quan trọng được xem xét đồng đều, từ đó cải thiện quá trình học tập một cách toàn diện.

## Phương pháp Actor-Critic:

* **Chính sách ():** Trong Actor-Critic, Actor chịu trách nhiệm quyết định hành động nào cần thực hiện dựa trên trạng thái hiện tại. Chính sách (policy), được ký hiệu là π, xác định chiến lược chọn hành động dựa trên xác suất cho mỗi trạng thái cụ thể.
* **Hàm giá trị ():** Critic chịu trách nhiệm đánh giá giá trị của trạng thái hiện tại khi tác nhân tuân theo chính sách. Hàm giá trị, được ký hiệu là , đo lường giá trị kỳ vọng của việc bắt đầu tại trạng thái và tuân theo chính sách . Hàm giá trị được tính như sau:

Trong đó, là hệ số chiết khấu (discount factor), là phần thưởng tại thời điểm t

* Ngoài hàm giá trị - trạng thái, Critic cũng có thể sử dụng hàm giá trị - hành động, ký hiệu là , để đánh giá giá trị của cặp trạng thái-hành động. Hàm giá trị hành động đo lường giá trị kỳ vọng khi tác nhân ở trạng thái , thực hiện hành động , và sau đó tiếp tục tuân theo chính sách . Hàm giá trị hành động được biểu diễn như sau:
* Phương trình Bellman đóng vai trò quan trọng trong việc tính toán giá trị của các trạng thái và cặp trạng thái-hành động trong Actor-Critic framework. Nó cho phép tính toán đệ quy giá trị kỳ vọng, giúp Critic đưa ra đánh giá chính xác hơn về chính sách của Actor. Dưới đây là hai dạng chính của phương trình Bellman:
  + Phương trình Bellman cho hàm giá trị:
  + Phương trình Bellman cho hàm giá trị - hành động:
* Trong quy trình huấn luyện mô hình, để tối thiểu hàm mất mát (, chúng tôi áp dụng công thức sau:

Trong đó:

Phương pháp Actor-Critic kết hợp giữa việc quyết định hành động thông qua Actor và đánh giá chính sách thông qua Critic. Actor sử dụng chính sách để chọn hành động, trong khi Critic sử dụng hàm giá trị để đánh giá và điều chỉnh chính sách của Actor nhằm tối ưu hóa phần thưởng kỳ vọng. Phương trình Bellman cung cấp cơ sở lý thuyết cho việc tính toán giá trị của các trạng thái và hành động, giúp cải thiện sự hiệu quả của quá trình học tập.

### Cá nhân hóa hệ thống gợi ý

Trong bài báo này, nhóm nghiên cứu áp dụng thuật toán Online Memory (Tzeng et al., 2023) để lưu trữ trạng thái của các bài tập mà sinh viên đã thực hiện, giúp tối ưu hóa quá trình học tập cá nhân. Khi một người học hoàn thành một bài tập, trạng thái hiện tại, hành động đã chọn, phần thưởng nhận được, trạng thái tiếp theo, và trạng thái hoàn thành sẽ được lưu trữ vào bộ nhớ này. Nếu bộ nhớ đạt đến kích thước tối đa đã định, các trải nghiệm cũ nhất sẽ được loại bỏ để nhường chỗ cho các trải nghiệm mới. Điều này đảm bảo rằng hệ thống chỉ học từ các trải nghiệm gần đây, phản ánh sát nhất quá trình học tập hiện tại của người học. Các trải nghiệm được lấy mẫu từ bộ nhớ để huấn luyện mô hình Actor-Critic. Bằng cách này, hệ thống có thể học cách tối ưu hóa hành động để cải thiện phần thưởng dài hạn cho sinh viên. Thuật toán Online Memory có thể được mô tả như sau:

* 1. Khởi tạo bộ nhớ : Sử dụng cấu trúc dữ liệu ‘deque’ để lưu trữ trải nghiệm với kích thước giới hạn. Bộ nhớ này có chức năng như một buffer lưu các trải nghiệm gần đây nhất của sinh viên.
  2. Khi sinh viên trả lời câu hỏi, hệ thống lưu một trải nghiệm mới dưới dạng (S, A, R, ). Nếu bộ nhớ đã đầy, trải nghiệm cũ nhất sẽ bị loại bỏ để nhường chỗ cho trải nghiệm mới.

## Quy trình huấn luyện mô hình

Trong lĩnh vực học tăng cường, các bài toán liên quan đến việc liên tục phải đưa ra quyết định thường đặt ra những thách thức lớn do không gian hành động vô hạn và sự phức tạp của môi trường. Deep Deterministic Policy Gradient (DDPG) là một thuật toán học tăng cường dựa trên mạng nơ-ron sâu, được thiết kế để giải quyết hiệu quả các vấn đề này. Bằng cách kết hợp những ưu điểm của cả học sâu và học tăng cường, DDPG cho phép tác nhân học hỏi trực tiếp từ tương tác với môi trường, từ đó đưa ra những quyết định tối ưu.

Thuật toán Deep Deterministic Policy Gradient (DDPG) bắt đầu bằng việc khởi tạo hai mạng nơ-ron: Actor Network (μ) để chọn hành động dựa trên trạng thái hiện tại, và Critic Network (Q) để ước lượng giá trị của hành động đó. Song song với đó, các Target Networks (μ' và Q') được khởi tạo với các trọng số ban đầu từ hai mạng chính để giúp việc cập nhật mô hình trở nên ổn định hơn. Quá trình huấn luyện diễn ra trong từng episode, tại mỗi bước của episode, hệ thống sẽ chọn một hành động dựa trên Actor Network (Dòng 4). Trải nghiệm thu được từ hành động này (gồm trạng thái hiện tại, hành động, phần thưởng và trạng thái tiếp theo) sẽ được lưu trữ trong Replay Buffer (Dòng 6). Để huấn luyện mô hình, một batch ngẫu nhiên từ Replay Buffer sẽ được lấy mẫu. Critic Network được huấn luyện bằng cách tối ưu hóa sự sai lệch giữa giá trị ước lượng và giá trị mục tiêu, giá trị mục tiêu này được tính toán dựa trên phần thưởng nhận được và giá trị của trạng thái tiếp theo do Target Critic Network ước lượng. Sau đó, Actor Network được cập nhật thông qua gradient ascent, tối ưu hóa các hành động để tăng giá trị ước lượng của Critic Network. Cuối cùng, các Target Networks sẽ được cập nhật dần dần bằng cách kết hợp trọng số từ các mạng chính và mạng mục tiêu trước đó, đảm bảo sự ổn định trong suốt quá trình huấn luyện.

| **Thuật toán Deep Deterministic Policy Gradient** |
| --- |
| Khởi tạo mạng critic và mạng actor với trọng số ngẫu nhiên Khởi tạo mạng Q-Target với trọng số  Khởi tạo bộ nhớ đệm M  1: **for u = 1, 2, …, do** 2: Khởi tạo trạng thái ngẫu nhiên 3: **for t = 1, 2, …, T do**  4: Chọn hành động dựa trên chính sách  5: Thực hiện hành động , quan sát phần thưởng và trạng thái kế tiếp  6: Lưu trải nghiệm  7: Cập nhật state  8: Lấy mẫu một batch các trải nghiệm từ bộ nhớ đệm 9: Tính  10: Cập nhật mạng critic bằng cách tối thiểu hóa hàm mất mát  11: Cập nhật mạng target: 12:  13:  14: **end** 15: **end** |

## Đề xuất phương pháp đánh giá thuật toán

Trong tương lai, để đánh giá hiệu quả của hệ thống gợi ý bài tập, chúng tôi sẽ sử dụng các chỉ số tỷ lệ các bài tập được hoàn thành bởi người học trên tổng số bài tập của hệ thống. Chỉ số này cho thấy mức độ hiệu quả của hệ thống trong việc thúc đẩy sinh viên tham gia và hoàn thành tất cả các bài tập trong khoá học. Chỉ số được tính bởi công thức:

Chỉ số độ chính xác tại K (Precision@K) cũng sẽ được sử dụng để đo lường phần trăm số bài tập phù hợp với người học trong số K bài tập hàng đầu mà hệ thống gợi ý. Bên cạnh đó, nhóm cũng sẽ sử dụng chỉ số khả năng thu hồi tại K (Recall@K) nhằm đánh giá tỷ lệ số bài tập phù hợp được gợi ý trong K bài tập hàng đầu.

Sau khi hoàn thành triển khai hệ thống gợi ý, nhóm nghiên cứu đề xuất quy trình thu thập dữ liệu được thực hiện thông qua một RESTful API được xây dựng bằng Python, sử dụng Flask để kết nối với cơ sở dữ liệu MongoDB. API này phục vụ giao tiếp giữa trang web học tập và cơ sở dữ liệu, đảm bảo việc lưu trữ và truy xuất dữ liệu một cách hiệu quả.

Trang web học tập hiện tại chỉ cung cấp một môn học duy nhất là "Nhập môn Công nghệ Thông tin." Tất cả các lớp tham gia vào nghiên cứu đều truy cập được vào trang web này. Tuy nhiên, cấu trúc của nghiên cứu được thiết kế để chia các lớp thành hai nhóm nhằm đánh giá hiệu quả của hệ thống gợi ý:

* **Nhóm 1**: Bao gồm một nửa số lớp học, được tích hợp hệ thống gợi ý. Học sinh trong các lớp này có quyền lựa chọn sử dụng hệ thống gợi ý hoặc không. Việc này giúp chúng ta đánh giá sự ảnh hưởng của hệ thống gợi ý đến hiệu suất học tập và mức độ hài lòng của học sinh.
* **Nhóm 2**: Bao gồm các lớp còn lại, không được cung cấp hệ thống gợi ý. Học sinh trong các lớp này chỉ thực hành làm bài tập trắc nghiệm một cách thông thường mà không có sự hỗ trợ từ hệ thống gợi ý (các câu hỏi được cung cấp một cách ngẫu nhiên).

Dữ liệu từ cả hai nhóm sẽ được thu thập tự động thông qua API và lưu trữ trong MongoDB. Các thông tin bao gồm điểm số, thời gian hoàn thành bài tập, và quyết định có trả lời câu hỏi được gợi ý hay không (đối với nhóm 1). Dữ liệu này sẽ được sử dụng để so sánh, phân tích hiệu quả của hệ thống gợi ý đối với quá trình học tập của sinh viên.

A diagram of a computer process

Description automatically generated

*Hình 1. Sơ đồ hệ thống gợi ý bài tập trắc nghiệm*

# Kết luận và đề xuất

Nghiên cứu này đề xuất một giải pháp ứng dụng thuật toán Học tăng cường trong xây dựng hệ thống gợi ý cho các môn học thi trắc nghiệm tại Đại học Kinh tế Quốc dân. Hệ thống được triển khai nhằm hỗ trợ việc học mang tính cá nhân hoá cao thông qua việc đưa ra các gợi ý dựa trên kết quả làm bài của sinh viên, giúp họ nắm bắt được điểm mạnh, điểm yếu của bản thân trong quá trình ôn tập, từ đó có thể tập trung vào phần kiến thức mình chưa vững. Không chỉ củng cố động lực học tập và tăng thêm hứng thú học cho sinh viên, hệ thống còn kỳ vọng có thể được sử dụng như một công cụ hỗ trợ giảng viên trong việc quản lý, đảm bảo chất lượng học tập của người học. Trong bối cảnh hình thức thi trắc nghiệm trên máy tính được ứng dụng phổ biến như hiện nay ở Đại học Kinh tế Quốc dân, nhà trường chưa có một hệ thống hay nền tảng nào hỗ trợ sinh viên trong việc ôn tập các câu hỏi trắc nghiệm trước kỳ thi. Việc triển khai hệ thống này cho sinh viên trong Trường không chỉ trao cho người học cơ hội tự trau dồi, đào sâu kiến thức môn học, mà còn giúp họ có sự chuẩn bị tốt cho các bài thi, bài kiểm tra. Để thực hiện được điều này, các câu hỏi ôn tập được đề xuất sử dụng hệ thống gợi ý cần phải được gắn vào các khối kiến thức tương ứng trong hệ thống thi của Nhà trường. Hệ thống gợi ý nêu trên nếu được triển khai thành công sẽ có tiềm năng ứng dụng vào nhiều môn học thi theo hình thức trắc nghiệm khác, đặc biệt là các môn học khác thuộc khối kiến thức cơ sở của Trường Đại học Kinh tế Quốc dân.

**Tài liệu tham khảo**

Adil, H. M., Ali, S., Sultan, M., Ashiq, M., & Rafiq, M. (2024). Open education resources’ benefits and challenges in the academic world: a systematic review. *Global Knowledge, Memory and Communication*, *73*(3), 274-291.

Atkinson, R. C. (1972). Ingredients for a theory of instruction. *American Psychologist*, *27*(10), 921.

Barnes, T., & Stamper, J. (2008). Toward automatic hint generation for logic proof tutoring using historical student data. International conference on intelligent tutoring systems,

Bojarski, M. (2016). End to end learning for self-driving cars. *arXiv preprint arXiv:1604.07316*.

Bojarski, M., Yeres, P., Choromanska, A., Choromanski, K., Firner, B., Jackel, L., & Muller, U. (2017). Explaining how a deep neural network trained with end-to-end learning steers a car. *arXiv preprint arXiv:1704.07911*.

Chitrakar, N., & Nisanth, P. (2023). Frustration and its influences on Student Motivation and Academic Performance. *International Journal of Scientific Research in Modern Science and Technology*, *2*(11), 01-09.

Das, S., Dey, A., Pal, A., & Roy, N. (2015). Applications of artificial intelligence in machine learning: review and prospect. *International Journal of Computer Applications*, *115*(9).

Deng, Y., Bao, F., Kong, Y., Ren, Z., & Dai, Q. (2016). Deep direct reinforcement learning for financial signal representation and trading. *IEEE transactions on neural networks and learning systems*, *28*(3), 653-664.

DeSouza, E., & Fleming, M. (2003). A comparison of in-class and online quizzes on student exam performance. *Journal of Computing in Higher Education*, *14*, 121-134.

Elpidorou, A. (2023). Boredom and cognitive engagement: A functional theory of boredom. *Review of Philosophy and Psychology*, *14*(3), 959-988.

Fahad Mon, B., Wasfi, A., Hayajneh, M., Slim, A., & Abu Ali, N. (2023). Reinforcement Learning in Education: A Literature Review. Informatics,

Fok, A. W. P., & Ip, H. H. (2004). Personalized Education (PE) œ Technology Integration for Individual Learning.

Gong, T., & Yao, X. (2019). Deep exercise recommendation model. *International Journal of Modeling and Optimization*, *9*(1), 18-23.

He-Yueya, J., & Singla, A. (2021). Quizzing Policy Using Reinforcement Learning for Inferring the Student Knowledge State. *International Educational Data Mining Society*.

Huang, Z., Liu, Q., Zhai, C., Yin, Y., Chen, E., Gao, W., & Hu, G. (2019). Exploring multi-objective exercise recommendations in online education systems. Proceedings of the 28th ACM International Conference on Information and Knowledge Management,

Khandani, A. E., Kim, A. J., & Lo, A. W. (2010). Consumer credit-risk models via machine-learning algorithms. *Journal of Banking & Finance*, *34*(11), 2767-2787.

Li, X., Li, L., Gao, J., He, X., Chen, J., Deng, L., & He, J. (2015). Recurrent reinforcement learning: a hybrid approach. *arXiv preprint arXiv:1509.03044*.

Li, Z., Hu, H., Xia, Z., Zhang, J., Li, X., Wang, Z., Huang, X., Zeng, S., & Qiu, B. (2021). Exercise recommendation method based on machine learning. 2021 International Conference on Advanced Learning Technologies (ICALT),

Liu, Y.-E., Mandel, T., Brunskill, E., & Popovic, Z. (2014). Trading Off Scientific Knowledge and User Learning with Multi-Armed Bandits. EDM,

Mahesh, B. (2020). Machine learning algorithms-a review. *International Journal of Science and Research (IJSR).[Internet]*, *9*(1), 381-386.

Mandel, T., Liu, Y.-E., Levine, S., Brunskill, E., & Popovic, Z. (2014). Offline policy evaluation across representations with applications to educational games. AAMAS,

Mingo, M. A., Chang, H.-H., & Williams, R. L. (2018). Undergraduate students’ preferences for constructed versus multiple-choice assessment of learning. *Innovative Higher Education*, *43*(2), 143-152.

Minoofam, S. A. H., Bastanfard, A., & Keyvanpour, M. R. (2022). RALF: an adaptive reinforcement learning framework for teaching dyslexic students. *Multimedia Tools and Applications*, *81*(5), 6389-6412.

Naeem, M., Rizvi, S. T. H., & Coronato, A. (2020). A gentle introduction to reinforcement learning and its application in different fields. *IEEE Access*, *8*, 209320-209344.

Nhàn, H. L. T. 2013 Hệ thống dự đoán kết quả học tập và gợi ý lựa chọn môn học. Kỷ yếu hội thảo quốc gia lần thứ XVI: Một số vấn đề chọc lọc của CNTT&TT

Qiang, W., & Zhongli, Z. (2011). Reinforcement learning model, algorithms and its application. 2011 International Conference on Mechatronic Science, Electric Engineering and Computer (MEC),

Ritter, F. E., Nerb, J., Lehtinen, E., & O'Shea, T. M. (2007). *In order to learn: How the sequence of topics influences learning*. Oxford University Press.

Ryan, A., Judd, T., Swanson, D., Larsen, D. P., Elliott, S., Tzanetos, K., & Kulasegaram, K. (2020). Beyond right or wrong: More effective feedback for formative multiple-choice tests. *Perspectives on Medical Education*, *9*, 307-313.

Scalise, K., & Gifford, B. (2006). Computer-Based Assessment in E-Learning: A Framework for Constructing "Intermediate Constraint" Questions and Tasks for Technology Platforms. *The Journal of Technology, Learning and Assessment*, *4*(6). <https://ejournals.bc.edu/index.php/jtla/article/view/1653>

Silver, D., Newnham, L., Barker, D., Weller, S., & McFall, J. (2013). Concurrent reinforcement learning from customer interactions. International conference on machine learning,

Sutton, R. S. (2018). Reinforcement learning: An introduction. *A Bradford Book*.

Tzeng, J.-W., Huang, N.-F., Chuang, A.-C., Huang, T.-W., & Chang, H.-Y. (2023). Massive open online course recommendation system based on a reinforcement learning algorithm. *Neural Computing and Applications*, 1-12.

Vu, T. N., Nguyen, T. D. A., & Le, T. D. (2020). Develop High School Students Recommendation System Based on Ontology. Future Data and Security Engineering. Big Data, Security and Privacy, Smart City and Industry 4.0 Applications: 7th International Conference, FDSE 2020, Quy Nhon, Vietnam, November 25–27, 2020, Proceedings 7,

Wood, E., Klausz, N., & MacNeil, S. (2022). Examining the influence of multiple-choice test formats on student performance. *Innovative Higher Education*, 1-17.