

MÔ HÌNH TRUY VẾT TRI THỨC NGƯỜI HỌC KẾT HỢP ĐỘNG LỰC THỜI GIAN: PHÂN TÍCH KHẢ NĂNG DỰ ĐOÁN TỪ SAINT+ VÀ DỮ LIỆU EDNET

Phạm Hoàng An¹
Email: an.ph@ou.edu.vn

Ngày tòa soạn nhận được bài báo: 19/08/2025

Ngày phản biện đánh giá: 13/10/2025

Ngày bài báo được duyệt đăng: 24/10/2025

DOI: 10.59266/houjs.2025.811

Tóm tắt: Trí tuệ nhân tạo đang trở thành công cụ mạnh mẽ trong việc định hình trải nghiệm học tập cá nhân hóa, đặc biệt thông qua các mô hình truy vết và dự đoán trạng thái tri thức của người học theo thời gian. Trong số đó, mô hình SAINT (Separated Attention-based Knowledge Tracing) đạt độ chính xác cao nhờ khai thác các đặc trưng thời gian như thời lượng làm bài và khoảng cách giữa các lần phản hồi. Tuy nhiên, tính chất quyết định không minh bạch của mô hình này đặt ra thách thức đối với các nhà giáo dục trong việc tìm kiếm sự diễn giải và niềm tin vào hệ thống AI. Nghiên cứu này phân tích khả năng giải thích của cơ chế chú ý theo thời gian trong SAINT, sử dụng bộ dữ liệu quy mô lớn EdNet với hơn 130 triệu tương tác từ học sinh trong nhiều môi trường học tập khác nhau. Chúng tôi tập trung vào cách các động lực thời gian ảnh hưởng đến chất lượng dự đoán, đồng thời trực quan hóa trọng số chú ý để diễn giải cách các mẫu hành vi theo thời gian định hình mô hình hóa người học. Kết quả cho thấy việc mô hình hóa thời gian một cách minh bạch không chỉ cải thiện hiệu suất dự đoán mà còn thúc đẩy niềm tin và sự tương thích với trực giác sư phạm. Công trình này đặc biệt có ý nghĩa đối với các nhà giáo dục tại các khu vực thiếu nguồn lực như Việt Nam, nơi các công cụ AI có tính giải thích và dễ tiếp cận có thể góp phần thúc đẩy giáo dục cá nhân hóa một cách công bằng và khả thi hơn.

Từ khóa: SAINT+, Ednet, cá nhân hóa học tập, knowledge tracing

I. Đặt vấn đề

Trí tuệ nhân tạo (AI) đang ngày càng tái cấu trúc hệ thống giáo dục thông qua việc tạo ra các môi trường học tập cá nhân hóa, phù hợp với nhu cầu riêng biệt

của từng người học. Một trong những cơ chế cốt lõi thúc đẩy sự chuyển đổi này là truy vết tri thức (Knowledge Tracing - KT), mô hình hóa quá trình hiểu biết tiến triển của người học bằng cách phân

¹ Trường Đại học Mở TP HCM

tích lịch sử tương tác của họ với tài liệu học tập.

Các mô hình dựa trên kiến trúc Transformer đã chứng minh hiệu quả vượt trội trong lĩnh vực này, đặc biệt là mô hình SAINT (*Separated Attention-based Knowledge Tracing*), vốn sử dụng cơ chế chú ý kép để xử lý riêng biệt chuỗi câu hỏi và chuỗi phản hồi (Choi et al., 2020). Phiên bản kế tiếp, SAINT+, tích hợp thêm các đặc trưng thời gian như thời lượng thực hiện nhiệm vụ và khoảng cách giữa các câu hỏi, từ đó nâng cao độ chính xác trong dự đoán (Shin et al., 2020).

Mặc dù các mô hình này đạt hiệu suất cao, một hạn chế quan trọng vẫn tồn tại: thiếu khả năng diễn giải. Các cơ chế phức tạp đằng sau trọng số chú ý, đặc biệt là những yếu tố liên quan đến động lực thời gian, vẫn còn mờ昧 đối với các nhà giáo dục và nhà hoạch định chính sách. Tại các khu vực như Việt Nam, nơi công nghệ giáo dục đang phát triển nhanh nhưng vẫn còn thiếu thốn về hạ tầng, AI có khả năng diễn giải đóng vai trò then chốt trong việc xây dựng niềm tin và định hướng triển khai có trách nhiệm.

Nghiên cứu này hướng đến việc phân tích và trực quan hóa cơ chế chú ý theo thời gian trong mô hình SAINT+, sử dụng bộ dữ liệu EdNet quy mô lớn. Mặc dù dữ liệu không xuất phát từ Việt Nam, các đặc điểm hành vi học tập được ghi nhận có thể cung cấp những gợi ý có giá trị cho các hệ thống giáo dục đang trong quá trình chuyển đổi số. Cụ thể, chúng tôi khảo sát cách các đặc trưng nhạy cảm với thời gian ảnh hưởng đến chất lượng dự đoán, và cách việc diễn giải phân bố chú ý có thể làm cho quá trình cá nhân hóa dựa trên AI trở nên minh bạch hơn. Bằng cách kết hợp giữa phân tích kỹ thuật và ý nghĩa

sư phạm, nghiên cứu này đóng góp vào việc phát triển các mô hình học tập có khả năng diễn giải, hướng đến khả năng triển khai thực tế trong các môi trường giáo dục đang phát triển.

II. Cơ sở lý thuyết

Học tập cá nhân hóa qua trí tuệ nhân tạo đang trở thành xu hướng chuyển đổi trong giáo dục hiện đại. Bằng cách thích ứng với hành vi và hiệu suất của từng người học, các hệ thống AI giúp nâng cao kết quả học tập và hiệu quả giảng dạy. Một kỹ thuật cốt lõi trong các hệ thống này là KT, nhằm ước lượng trạng thái tri thức hiện tại và dự đoán hiệu suất tương lai dựa trên lịch sử tương tác.

KT đã phát triển từ các mô hình thống kê có khả năng diễn giải sang các kiến trúc học sâu phức tạp, phản ánh xu hướng biểu diễn thông tin phong phú và mô hình hóa hành vi linh hoạt. Đặc biệt, việc tích hợp đặc trưng thời gian và cơ chế chú ý giúp mô hình như SAINT+ trở nên chính xác hơn và có tiềm năng ứng dụng thực tiễn trong giáo dục, đặc biệt tại Việt Nam.

2.1. Mô hình truy vết tri thức

Truy vết tri thức là kỹ thuật dự đoán xác suất người học trả lời đúng một bài tập dựa trên chuỗi tương tác trước đó. Các mô hình KT ban đầu thường mang tính xác suất có cấu trúc, giúp theo dõi trạng thái tri thức ở cấp độ khái niệm. Dù dễ diễn giải, chúng gặp hạn chế khi xử lý hành vi phức tạp và yếu tố thời gian như độ trễ phản hồi hay sự thay đổi trong quá trình học. Công thức KT cơ bản được biểu diễn dưới dạng:

$$P[c_t=1|I_p, I_r, \dots, I_{(t-1)}, E_t]$$

Trong đó $c_t \in \{0,1\}$ là kết quả chính xác ở thời điểm hiện tại, Mỗi tương tác $I_{(t-1)} = (E_p, R_p)$ gồm thông tin bài tập và phản

hồi của người học, kèm theo siêu dữ liệu như loại bài tập, độ chính xác, thời gian phản hồi và khoảng cách giữa các lần tương tác. Dưới đây là tổng quan ngắn gọn về một số phương pháp KT tiêu biểu đã góp phần định hình lĩnh vực này.

2.1.1. Các mô hình truyền thống

- *Bayesian Knowledge Tracing* (BKT) mô hình hóa trạng thái tri thức của người học như một biến ẩn, và cập nhật niềm tin dựa trên độ chính xác của phản hồi. Mô hình này sử dụng các tham số cố định như tốc độ học và tốc độ quên, nhưng không thích ứng tốt với lộ trình học tập cá nhân hóa.

- *Learning Factors Analysis* (LFA) áp dụng hồi quy logistic để ước lượng ảnh hưởng của các yếu tố khác nhau đến hiệu suất học tập, bao gồm độ khó của khái niệm và số lần thử trước đó. Phương pháp này cải thiện tính linh hoạt của tham số trong khi vẫn giữ được khả năng diễn giải.

- *Performance Factors Analysis* (PFA hoặc LPA) phát triển từ LFA bằng cách phân biệt giữa phản hồi đúng và sai nhằm mô hình hóa quá trình tiến bộ của người học một cách chính xác hơn. Mặc dù hiệu quả trên các bộ dữ liệu có cấu trúc, các mô hình này bị giới hạn bởi khả năng xử lý dữ liệu quy mô lớn và có tính thời gian chi tiết.

2.1.2. Các phương pháp học sâu

- *Deep Knowledge Tracing* (DKT) sử dụng mạng nơ-ron hồi tiếp (RNN) để mô hình hóa chuỗi tương tác học tập. Phương pháp này mang lại khả năng dự đoán mạnh mẽ hơn nhưng đánh đổi khả năng diễn giải.

- *Self-Attentive Knowledge Tracing* (SAKT) thay thế RNN bằng kiến trúc Transformer và sử dụng cơ chế tự chú ý

để đánh trọng số cho các bài tập trước khi dự đoán hiệu suất hiện tại, từ đó cải thiện khả năng mở rộng và độ nhạy ngữ cảnh.

- *Separated Attention-based Knowledge Tracing* (SAINT) nâng cấp SAKT bằng cách bổ sung hai luồng chú ý riêng biệt một cho biểu diễn câu hỏi và một cho biểu diễn phản hồi cho phép mô hình hóa quá trình học tập một cách biểu cảm hơn.

- SAINT+ là phiên bản tinh chỉnh của SAINT, tích hợp thêm các đặc trưng thời gian như thời gian đã trôi qua, thời lượng phiên học, và khoảng cách giữa các mục học tập. Những cải tiến này giúp tăng đáng kể độ chính xác dự đoán bằng cách làm cho cơ chế chú ý trở nên nhạy cảm với yếu tố thời gian.

Nghiên cứu này kế thừa mô hình SAINT+, với trọng tâm đặc biệt vào cơ chế chú ý theo thời gian, nhằm khám phá cách các đặc trưng này có thể được diễn giải và trực quan hóa để thúc đẩy tính minh bạch và sự hiểu biết trong môi trường giáo dục.

2.2. Động lực thời gian trong giáo dục

Động lực thời gian trong giáo dục là tập hợp các yếu tố liên quan đến thời gian bao gồm thời gian thực hiện nhiệm vụ, khoảng cách giữa các sự kiện học tập, độ trễ phản hồi, và thời lượng phiên học có ảnh hưởng trực tiếp đến quá trình tiếp nhận, ghi nhớ và duy trì tri thức của người học. Khái niệm này phản ánh cách thời gian tác động đến nhận thức, hành vi học tập và hiệu suất học tập trong môi trường giáo dục.

2.2.1. Các lý thuyết nhận thức và hành vi

Nhiều nghiên cứu đã chỉ ra rằng các biến số liên quan đến thời gian đóng vai trò quan trọng trong cách người học xử lý và ghi nhớ thông tin. Ví dụ, hiệu ứng

giãn cách (spacing effect) cho thấy việc phân bổ thời gian học tập hợp lý theo từng giai đoạn giúp cải thiện khả năng ghi nhớ dài hạn so với phương pháp học dồn hoặc nhồi nhét (Carpenter et al., 2012; Santoro, 2021). Hiện tượng này, lần đầu tiên được quan sát bởi Ebbinghaus và sau đó được xác nhận qua hàng trăm nghiên cứu, giúp tăng cường quá trình mã hóa và truy xuất thông tin trong não bộ.

Bên cạnh đó, lý thuyết suy giảm trí nhớ (memory decay theory) cho rằng thông tin sẽ dần mất đi theo thời gian nếu không được củng cố thông qua việc ôn tập hoặc truy xuất (Teschner, 2024; Ebbinghaus, 1913). Sự xói mòn tự nhiên của dấu vết trí nhớ này được ghi nhận trong cả hệ thống trí nhớ ngắn hạn và dài hạn.

Thời gian thực hiện nhiệm vụ (time-on-task) tức khoảng thời gian người học dành để tập trung vào một nhiệm vụ cụ thể liên quan mật thiết đến mức độ nỗ lực nhận thức và sự chú ý. Thời gian thực hiện dài hơn có thể phản ánh quá trình xử lý sâu hơn, nhưng cũng có thể là dấu hiệu của sự bối rối hoặc mệt mỏi tùy theo ngữ cảnh (Leppink & Pérez-Fuster, 2019; Lee, 2019).

2.2.2. Ý nghĩa trong giáo dục

Trong bối cảnh truy vết tri thức, động lực thời gian không chỉ là một khái niệm lý thuyết mà còn đóng vai trò thiết thực trong việc xây dựng các đặc trưng đầu vào cho mô hình SAINT+. Các lý thuyết nhận thức như hiệu ứng giãn cách (spacing effect), suy giảm trí nhớ (memory decay), và thời gian thực hiện nhiệm vụ (time-on-task) được chuyển hóa thành các đặc trưng định lượng nhằm mô hình hóa tiến trình học tập của người học.

- Spacing effect được phản ánh qua đặc trưng khoảng cách giữa các lần tương tác hoặc độ trễ giữa các câu hỏi. Mô hình sử dụng thông tin này để đánh giá khả năng ghi nhớ dài hạn, giả định rằng khoảng cách hợp lý giữa các lần học sẽ cải thiện độ chính xác dự đoán.

- Memory decay theory được tích hợp thông qua đặc trưng thời gian đã trôi qua kể từ lần tương tác trước đó. Mô hình giả định rằng nếu khoảng thời gian quá dài mà không có sự củng cố, xác suất trả lời đúng sẽ giảm, từ đó điều chỉnh trọng số trong quá trình dự đoán.

- Time-on-task được sử dụng như một chỉ báo về mức độ nỗ lực nhận thức. Thời gian làm bài dài hơn có thể phản ánh sự tập trung cao hoặc sự bối rối, tùy theo ngữ cảnh. Mô hình SAINT+ sử dụng đặc trưng này để điều chỉnh mức độ chú ý (attention) trong quá trình mã hóa phản hồi của người học.

Việc tích hợp các đặc trưng này không chỉ giúp mô hình phản ánh sát thực hành vi học tập mà còn tạo ra các dự đoán mang tính cá nhân hóa và nhạy cảm với ngữ cảnh. Đây là cầu nối giữa lý thuyết tâm lý học và mô hình hóa AI, góp phần nâng cao giá trị sư phạm của hệ thống truy vết tri thức.

III. Phương pháp nghiên cứu

3.1. Tổng quan về bộ dữ liệu

Nghiên cứu này sử dụng bộ dữ liệu Riid Education, ban đầu được công bố trong khuôn khổ một cuộc thi dự đoán câu trả lời quy mô lớn. Bộ dữ liệu bao gồm ba thành phần chính: train.csv, questions.csv, và lectures.csv. Mỗi tệp cung cấp thông tin riêng biệt nhưng bổ trợ lẫn nhau trong việc mô hình hóa tương tác giữa người học và hệ thống, như được trình bày trong Bảng 1.

Cấu trúc phân cấp của EdNet. Bộ dữ liệu gốc EdNet được tổ chức theo bốn cấp độ trừu tượng: KT1, KT2, KT3 và KT4. Mỗi cấp độ bổ sung mức độ chi tiết cao hơn về tương tác của người học:

- KT1: Nhật ký câu hỏi-phản hồi cơ bản (thời gian, mã câu hỏi, câu trả lời của người dùng, thời gian thực hiện).

- KT2: Nhật ký hành động (vào bài, phản hồi, nộp bài) kèm theo siêu dữ liệu về nguồn và nền tảng sử dụng.

- KT3: Bao gồm các hoạt động học tập như đọc lời giải và xem bài giảng.

- KT4: Nhật ký tương tác đầy đủ, bao gồm sự kiện giao diện người dùng, điều khiển phương tiện và hành vi thanh toán.

Phiên bản được sử dụng trong nghiên cứu này tương ứng gần nhất với

cấp độ KT1, tập trung vào các tương tác cơ bản giữa người học và câu hỏi. Mặc dù bộ dữ liệu được công bố công khai thông qua nền tảng thi đấu khoa học dữ liệu, các bước xác thực và tiền xử lý bổ sung đã được thực hiện nhằm đảm bảo độ tin cậy và tính phù hợp với bối cảnh nghiên cứu.

Bảng 1. Nội dung chính của bộ dữ liệu và các cột chính

Dataset	Cột dữ liệu	Tác dụng trong mô hình
Train	User_id, content_id, timestamp, prior_question_elapsed_time, Answered_correctly	Lưu nhật ký tương tác chính. Được sử dụng để xây dựng chuỗi đầu vào cho từng người học. Các đặc trưng thời gian được trích xuất từ dấu thời gian. Mục tiêu dự đoán là độ chính xác của answered_correctly.
Questions	Question_id, bundle_id, correct_answer, tags	Cung cấp dữ liệu cho từng câu hỏi. Các thẻ (tags) mã hóa đặc trưng về mức độ kỹ năng. Đáp án đúng (correct_answer) được sử dụng để lọc và xác thực dữ liệu.
Lectures	lecture id, tag, part	Chứa nội dung học tập bổ trợ. Không được sử dụng trong quá trình huấn luyện SAINT+, nhưng có thể hữu ích cho các mở rộng đa phương thức trong tương lai.

3.2. Khai phá dữ liệu (Exploratory Data Analysis -EDA)

Trước khi xây dựng và huấn luyện mô hình SAINT+, việc khám phá dữ liệu đóng vai trò quan trọng trong việc hiểu rõ cấu trúc, đặc điểm và hành vi tiềm ẩn trong bộ dữ liệu EdNet. Phân tích EDA giúp làm rõ các yếu tố như phân phối thời gian phản hồi, độ trễ giữa các tương tác, tần suất kỹ năng, và xu hướng trả lời đúng/sai của người học. Những trực quan hóa này không chỉ hỗ trợ tiền xử lý dữ liệu mà còn cung cấp bằng chứng thực nghiệm cho việc tích hợp các đặc trưng thời gian vào mô hình truy vết tri thức. Qua đó, EDA đóng vai trò nền tảng trong việc kết nối dữ liệu thực tế với kiến trúc SAINT+ và định hướng cải tiến mô hình trong bối cảnh giáo dục Việt Nam.

3.2.1. Dữ liệu tập Train

Biểu đồ phân phối timestamp ở Hình 1 thể hiện khoảng thời gian (tính bằng mili giây) giữa mỗi tương tác của người học và sự kiện đầu tiên của họ trên nền tảng. Dữ

liệu cho thấy phần lớn người dùng tương tác trong giai đoạn đầu, với mật độ cao ở các giá trị timestamp thấp. Điều này phản ánh rằng đa số người học chưa hoạt động lâu trên hệ thống tại thời điểm ghi nhận. Việc lọc theo timestamp = 0 xác nhận rằng chúng ta có lịch sử đầy đủ cho tất cả 393,656 người dùng trong tập huấn luyện, với mỗi người đều có một dòng timestamp bằng 0 đại diện cho lần tương tác đầu tiên.



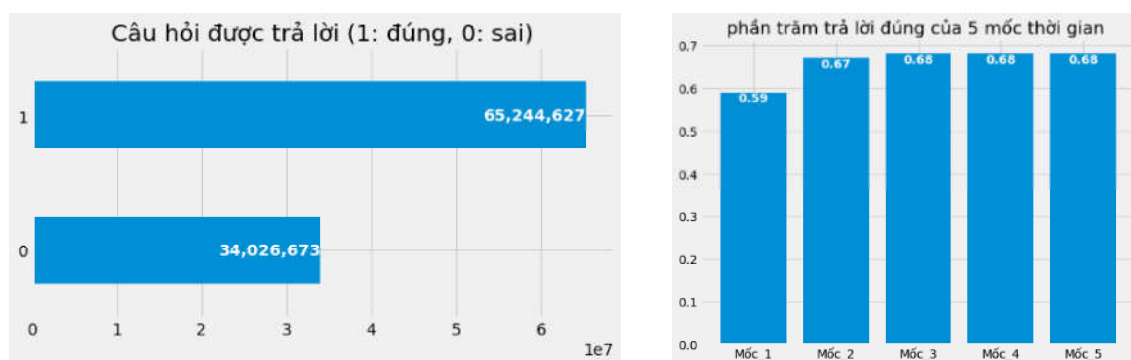
Hình 1: Biểu đồ giá trị Timestamp

Biến answered_correctly là mục tiêu dự đoán chính trong mô hình SAINT+, phản ánh việc người học trả lời đúng (1) hoặc sai (0) cho từng câu hỏi (Hình 2). Khi loại bỏ các tương tác dạng bài giảng (giá trị -1), dữ liệu cho thấy khoảng 1/3 số

câu hỏi được trả lời sai, cho thấy độ khó tương đối của bài tập trong EdNet.

Biểu đồ thứ hai của Hình 2 phân tích mối quan hệ giữa timestamp và answered_correctly bằng cách chia dữ liệu thành 5 nhóm thời gian. Kết quả cho thấy người học mới đăng ký gần đây có tỷ lệ trả lời

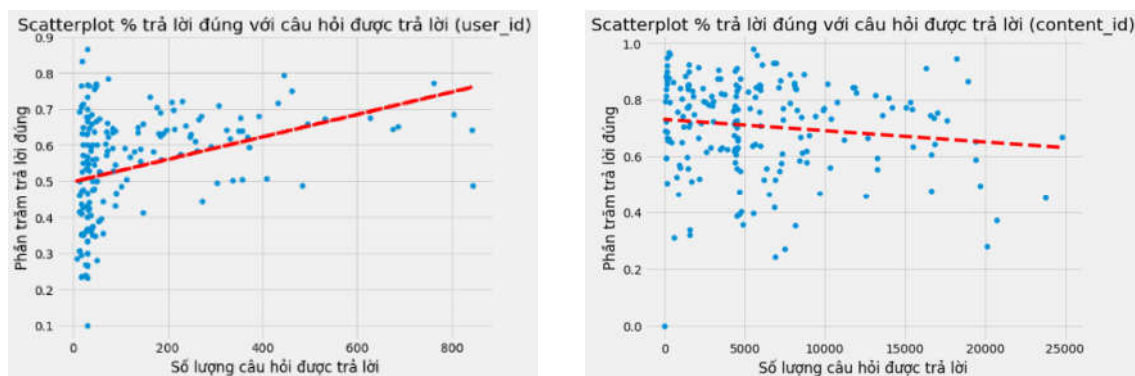
đúng thấp hơn so với những người hoạt động lâu hơn trên nền tảng. Mặc dù xu hướng này không quá rõ rệt, nó gợi ý rằng thời gian hoạt động có thể liên quan đến mức độ thành thạo, từ đó hỗ trợ việc tích hợp đặc trưng thời gian vào mô hình truy vết tri thức.



Hình 2: Số liệu về câu hỏi trong tập train

Hai biểu đồ scatter (Hình 3) cho thấy mối quan hệ giữa số lượng câu hỏi và tỷ lệ trả lời đúng, xét theo từng người học và từng nội dung câu hỏi. Với người học (user_id), xu hướng chung là tỷ lệ trả lời đúng tăng theo số lượng câu hỏi đã thực hiện, dù vẫn có sự biến động lớn ở nhóm ít tương tác. Ngược lại, với câu hỏi

(content_id), tỷ lệ trả lời đúng có xu hướng giảm nhẹ khi số lượt tương tác tăng, gợi ý rằng các câu hỏi phổ biến có thể khó hơn hoặc mang tính phân loại cao hơn. Những quan sát này giúp làm rõ sự đa dạng trong hành vi học tập và độ khó nội dung, từ đó hỗ trợ việc thiết kế mô hình truy vết tri thức hiệu quả hơn.



Hình 3: Scatterplot tỷ lệ trả lời đúng với mỗi câu hỏi được trả lời

Phân tích biến prior_question_had_explanation (Hình 4) cho thấy tỷ lệ trả lời đúng cao hơn khoảng 17% khi người học được cung cấp lời giải thích trước đó. Điều này cho thấy việc có lời giải thích đóng vai trò hỗ trợ rõ rệt trong quá trình học. Mặc

dù phần lớn câu hỏi đều có lời giải thích đi kèm, việc không có lời giải nên được xem là một bất lợi tiềm tàng. Đáng chú ý, các giá trị thiếu (missing) của biến này có tỷ lệ trả lời đúng gần với nhóm có lời giải thích, cho thấy khả năng dữ liệu bị thiếu chủ yếu

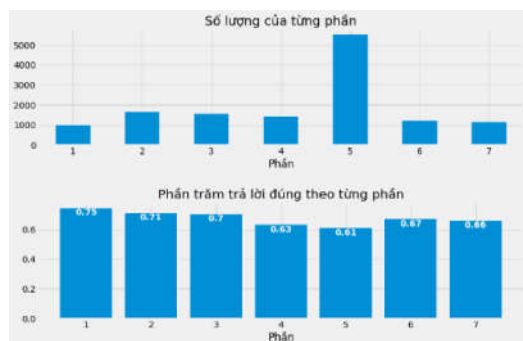
đến từ các tương tác vốn có lời giải nhưng không được ghi nhận đầy đủ.



Hình 4: Biểu đồ tỷ lệ câu hỏi được trả lời đúng khi được lý giải

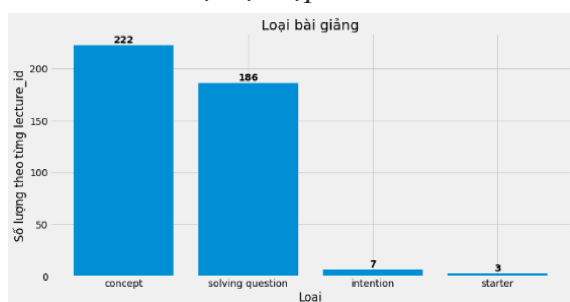
3.2.2. Dữ liệu tập Questions

Phân tích số lượng câu hỏi và tỷ lệ trả lời đúng theo từng phần (part) cho thấy sự phân bố không đồng đều giữa các nhóm nội dung (Hình 5). Đáng chú ý, Part 5 có số lượng câu hỏi lớn nhất nhưng đồng thời cũng ghi nhận tỷ lệ trả lời đúng thấp nhất, phản ánh mức độ khó cao hơn so với các phần còn lại. Quan sát này gợi ý rằng việc gắn thẻ kỹ năng theo part có thể hỗ trợ mô hình hóa độ khó và hành vi học tập trong truy vết tri thức.



Hình 5: Biểu đồ cấu trúc chung của bộ câu hỏi dữ liệu

3.2.3. Dữ liệu tập Lectures



Hình 6: Biểu đồ về loại bài giảng

Phân tích dữ liệu bài giảng (Hình 6) cho thấy có tổng cộng 418 lecture_id, với các loại bài giảng (type_of) được phân bố theo nhiều nhóm nội dung khác nhau.

Mặc dù số lượng bài giảng không lớn so với câu hỏi, biểu đồ ở Hình 7 cho thấy việc người học xem bài giảng có ảnh hưởng tích cực rõ rệt đến tỷ lệ trả lời đúng. Điều này cho thấy vai trò hỗ trợ của bài giảng trong việc củng cố kiến thức và cải thiện hiệu suất học tập, từ đó gợi ý rằng hành vi xem bài giảng nên được đưa vào mô hình truy vết tri thức như một đặc trưng bổ sung.



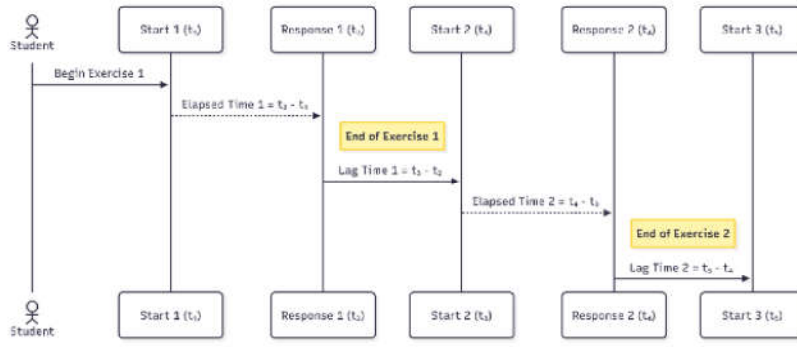
Hình 7: Biểu đồ tỷ lệ trả lời đúng khi người học xem bài giảng

3.3. Tổng quan mô hình

3.3.1. Cơ chế động lực thời gian

Để nâng cao năng lực dự đoán, SAINT+ tích hợp các tín hiệu thời gian được trích xuất từ nhật ký hoạt động của người học. Cụ thể, mô hình mã hóa hai đặc trưng liên tục (Hình 8): thời gian thực hiện mỗi câu hỏi (*elapsed time*) và độ trễ giữa các phản hồi liên tiếp (*time lag*). Các giá trị này được phân loại thành các nhóm (bucket) và ánh xạ thành các vector nhúng thời gian thông qua bảng tra học được.

Các vector nhúng thời gian này được cộng vào các vector nhúng đặc trưng phân loại tương ứng, và ảnh hưởng trực tiếp đến trọng số chú ý trong cả bộ mã hóa (encoder) và bộ giải mã (decoder). Bằng cách tập trung vào các mẫu hành vi nhạy cảm với thời gian chẳng hạn như sự do dự kéo dài hoặc phản hồi quá nhanh mô hình có thể nắm bắt các trạng thái nhận thức tiềm ẩn, vốn rất quan trọng trong việc mô hình hóa người học.



Hình 8: Biên độ thời gian của thời gian phải hỏi và độ trễ

Nghiên cứu này phân tích cách các thành phần thời gian ảnh hưởng đến độ chính xác dự đoán và khả năng diễn giải, thông qua các kỹ thuật trực quan hóa được trình bày trong Mục 4.

3.3.2. Kiến trúc SAINT+

Trong phần này, chúng tôi trình bày tóm tắt về SAINT+, một mô hình truy vết tri thức dựa trên kiến trúc Transformer được điều chỉnh chuyên biệt cho giáo dục cá nhân hóa (Choi et al., 2020). SAINT+ sử dụng kiến trúc hai luồng gồm bộ mã hóa và bộ giải mã, nhằm tách biệt chuỗi tương tác của người học thành hai phần: chuỗi bài tập và chuỗi phản hồi.

Bộ mã hóa nhận chuỗi embedding bài tập:

$$E_e = [E_{e1}, E_{e2}, \dots, E_{et}]$$

và tạo ra chuỗi đầu ra:

$$O_{enc} = [O_{enc1}, O_{enc2}, \dots, O_{enc_t}]$$

Bộ giải mã nhận chuỗi phản hồi đã dịch:

$$R_e = [S_e, R_{e1}, R_{e2}, \dots, R_{et-1}]$$

trong đó S_e là embedding của token bắt đầu. Mục tiêu của mô hình là dự đoán xác suất người học trả lời đúng tại thời điểm t , được biểu diễn dưới dạng:

$$P(c_t = 1 \mid E_t, I_1, I_2, \dots, I_{t-1})$$

Thành phần cốt lõi của SAINT+ là lớp attention đa đầu. Với đầu vào là ma

trận truy vấn Q , khóa K , và giá trị V , mỗi đầu attention được tính như sau:

$$Q_i = QW_{Q_i}, \quad K_i = KW_{K_i}, \quad V_i = VW_{V_i}$$

$$head_i = \text{SoftmaxMask}\left(\frac{Q_i K_i^T}{\sqrt{d}}\right) V_i$$

$$\text{Multihead}(Q, K, V) = \text{Concat}(head_1, \dots, head_h) W_o$$

trong đó d là chiều của vector truy vấn/ khóa, và W_o là ma trận trọng số đầu ra. Cơ chế mặt nạ nhân quả (causal mask) được áp dụng để ngăn việc truy cập thông tin tương lai trong chuỗi.

Mỗi khối mã hóa gồm N lớp encoder, với mỗi lớp thực hiện:

$$M = X + \text{Multihead}(\text{LayerNorm}(X), \text{LayerNorm}(X), \text{LayerNorm}(X))$$

$$O = M + \text{FFN}(\text{LayerNorm}(M))$$

trong đó FFN là mạng truyền thẳng hai lớp:

$$\text{FFN}(x) = \text{ReLU}(xW_1 + b_1)W_2 + b_2$$

Bộ giải mã gồm N lớp decoder, mỗi lớp thực hiện hai bước attention: đầu tiên là self-attention trên chuỗi phản hồi, sau đó là attention giữa phản hồi và đầu ra từ encoder:

$$M_1 = X + \text{Multihead}(\text{LayerNorm}(X), \text{LayerNorm}(X), \text{LayerNorm}(X))$$

$$M_2 = M_1 + \text{Multihead}(\text{LayerNorm}(M_1), O_{enc}, O_{enc})$$

$$O = M_2 + \text{FFN}(\text{LayerNorm}(M_2))$$

Đầu ra cuối cùng được đưa qua một lớp tuyến tính để tạo ra chuỗi xác suất dự đoán $\hat{c} = [\hat{c}_1, \hat{c}_2, \dots, \hat{c}_t]$. Kiến trúc này cho phép mô hình hóa hành vi người học một cách chi tiết, đồng thời tận dụng đặc trưng thời gian và nội dung để cải thiện độ chính xác trong truy vết tri thức.

3.4. Quy trình huấn luyện và phân tích khả năng diễn giải

3.4.1. Chi tiết môi trường huấn luyện

Mô hình SAINT+ được triển khai bằng thư viện PyTorch và huấn luyện trên nền tảng Google Development Cloud với GPU NVIDIA P100 và phiên bản CUDA 12.4. Mỗi chuỗi tương tác của người học được cắt ngắn hoặc bổ sung (padding) đến độ dài tối đa là 100 để đảm bảo hiệu quả tính toán. Quá trình huấn luyện sử dụng kích thước lô (batch size) là 64 và tốc độ học ban đầu là 10^{-3} , kết hợp với bộ tối ưu hóa AdamW và lịch trình giảm tốc độ học theo hình sin (cosine annealing scheduler). Kỹ thuật regularization bằng dropout được áp dụng với tỷ lệ 0.2 tại các lớp nhúng và lớp chú ý nhằm giảm thiểu hiện tượng quá khớp.

Mô hình sử dụng embedding có chiều 512, với 4 lớp encoder và 4 lớp decoder, mỗi lớp có 8 đầu attention. Tập dữ liệu huấn luyện được lấy từ file train.csv, bao gồm 13,523 câu hỏi và 10,000 kỹ năng.

Mục tiêu huấn luyện là phân loại nhị phân về độ chính xác của phản hồi của người học, sử dụng hàm mất mát Binary Cross Entropy (BCE).

$$\mathcal{L}_{BCE} = -\frac{1}{N} \sum_{i=1}^N [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)]$$

Trong đó y_i là nhãn thực tế và $\hat{y}_i$ là xác suất dự đoán từ mô hình đối với người học thứ i .

3.4.2. Quy trình huấn luyện

Mô hình SAINT+ được huấn luyện theo quy trình tuần tự gồm ba giai đoạn: embedding đầu vào, xử lý qua encoder-decoder, và dự đoán xác suất trả lời đúng (Thuật toán 1). Quá trình huấn luyện được tổ chức thông qua lớp PlusSAINTModule, kế thừa từ pl.LightningModule, giúp đơn giản hóa việc quản lý vòng lặp huấn luyện, tối ưu hóa và đánh giá.

Ở mỗi bước huấn luyện, mô hình nhận một batch dữ liệu gồm chuỗi bài tập, phản hồi, thời gian làm bài, và nhãn độ chính xác. Dữ liệu được embedding, sau đó đi qua encoder để trích xuất ngữ cảnh, và decoder để căn chỉnh phản hồi hiện tại với lịch sử tương tác. Đầu ra được đưa qua hàm sigmoid để dự đoán xác suất, và được so sánh với nhãn thực tế bằng hàm mất mát Binary Cross Entropy. Dưới đây là mã giả mô tả quy trình huấn luyện:

Algorithm 1: Huấn luyện mô hình SAINT+

Input: Batch dữ liệu gồm chuỗi bài tập, phản hồi, thời gian làm bài, và nhãn độ chính xác
Output: Loss huấn luyện, xác suất dự đoán, và các chỉ số đánh giá

```

1 Khởi tạo mô hình:
2   - Hàm mất mát BCEWithLogitsLoss
3   - Encoder và Decoder với multi-head attention
4   - Lớp embedding cho bài tập, phản hồi và thời gian
5   - Lớp tuyến tính đầu ra
6 Forward(x, y):
7   Embed bài tập và kỹ năng → enc
8   Embed phản hồi và thời gian → dec
9   encoder_output ← Encoder(enc)
10  decoder_output ← Decoder(dec, encoder_output)
11  output ← FC(decoder_output)
12  return output.square()
13 TrainingStep(batch):
14  Tách batch thành input, response, labels
15  output ← Forward(input, response)
16  loss ← BCE(output, labels)
17  Áp dụng mặt nạ để loại bỏ padding
18  output ← Sigmoid(output)
19  return loss, output, labels
20 ValidationStep(batch):
21  Tách batch thành input, response, labels
22  output ← Forward(input, response)
23  loss ← BCE(output, labels)
24  Áp dụng mặt nạ và sigmoid
25  Lưu output và labels để tính chỉ số
26  return output, labels
27 OnValidationEpochEnd():
28  Tổng hợp toàn bộ output và labels
29  Tính AUC, Accuracy và BCE Loss
30  Ghi nhận các chỉ số vào lịch sử
31  Hiển thị kết quả epoch
32  Xóa bộ đệm validation

```

Thuật toán 1: Các bước huấn luyện mô hình SAINT+

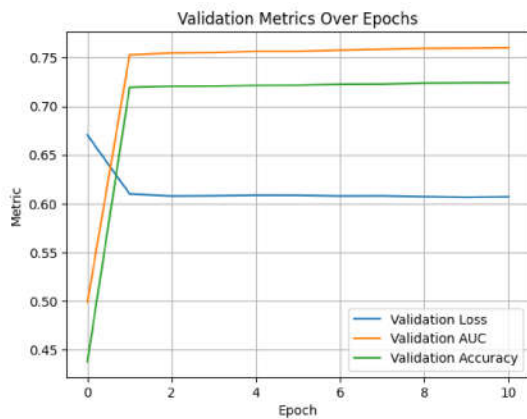
IV. Kết quả và thảo luận

4.1. Đánh giá hiệu suất mô hình

Để đánh giá năng lực dự đoán của mô hình SAINT+, chúng tôi tiến hành

phân tích hiệu suất dựa trên các chỉ số chuẩn gồm: diện tích dưới đường cong ROC (AUC), độ chính xác (accuracy), và độ mất mát (loss) theo từng vòng huấn luyện. Các kết quả tổng hợp được trình bày trong Hình 9, phản ánh hiệu suất tổng thể của mô hình trên tập kiểm định.

Sau 10 vòng lặp huấn luyện, SAINT+ đạt AUC = 0.760, accuracy = 0.724, và loss kiểm định = 0.607, cho thấy mô hình có khả năng phân biệt tương đối tốt giữa các phản hồi đúng và sai. Đáng chú ý, Hình 9 minh họa quá trình huấn luyện thông qua biểu đồ đường thể hiện sự thay đổi của loss, AUC và accuracy theo từng epoch. Đường cong loss giảm đều và hội tụ sớm, trong khi AUC và accuracy tăng ổn định, không có dao động lớn giữa các lần chạy. Điều này phản ánh khả năng tổng quát hóa tốt và độ ổn định của mô hình trong quá trình huấn luyện.



Hình 9: Kết quả sau khi huấn luyện mô hình của tập kiểm thử

4.2. So sánh với các mô hình cơ sở

4.2.1. So sánh kết quả về độ chính xác

Để đánh giá hiệu quả của SAINT+ một cách khách quan, chúng tôi tiến hành so sánh với các mô hình truy vết tri thức phổ biến. Các mô hình được huấn luyện trên cùng tập dữ liệu và quy trình tiền xử lý thống nhất nhằm đảm bảo tính công bằng

trong so sánh. Bảng 2 dưới đây trình bày độ chính xác (Accuracy) của từng mô hình:

Bảng 2: So sánh kết quả với các mô hình tiền nhiệm

Mô hình	Accuracy
DKT	0.7060
DKVMN	0.7079
SAKT	0.7073
SAINT	0.7178
SAINT+ (Ours)	0.7240

Kết quả cho thấy SAINT+ đạt độ chính xác cao nhất (0.7240), vượt trội so với các mô hình tuần tự như DKT và DKVMN cũng như mô hình attention đơn tầng như SAKT. Điều này phản ánh khả năng học sâu hơn từ chuỗi tương tác người học, đặc biệt trong việc xử lý các mối quan hệ phức tạp giữa bài tập và phản hồi.

Mặc dù mô hình SAINT+ trong nghiên cứu này không đạt AUC cao như các phiên bản trước, việc duy trì độ chính xác cao và ổn định vẫn cho thấy tiềm năng ứng dụng trong các hệ thống đánh giá học tập. Độ chính xác là một chỉ số quan trọng trong bối cảnh giáo dục, nơi việc phân loại đúng hành vi người học có thể ảnh hưởng trực tiếp đến lộ trình học tập và phản hồi cá nhân hóa.

4.2.2. Khác biệt giữa SAINT và SAINT+ và hiệu quả từ việc tích hợp động lực thời gian

SAINT+ là phiên bản mở rộng của mô hình SAINT gốc, được thiết kế nhằm cải thiện khả năng truy vết tri thức thông qua việc tích hợp các đặc trưng hành vi và yếu tố thời gian. Tuy nhiên, có nhiều biến thể của SAINT+ tùy theo cách lựa chọn đặc trưng đầu vào. Bảng 3 dưới đây so sánh ba phiên bản: SAINT gốc (Choi et al., 2020), SAINT+ (Shin et al., 2020), và SAINT+ trong nghiên cứu này.

Bảng 3: So sánh mô hình SAINT và SAINT+

Thành phần	SAINT (Choi et al., 2020)	SAINT+ (Shin et al., 2020)	SAINT+ (Ours)
Kiến trúc attention	Self-attention đơn tầng	Encoder-decoder attention	Encoder-decoder attention
Đặc trưng thời gian	Không có	lag time, time-on-task	lag time, time-on-task
Đặc trưng hành vi bổ sung	Không có	elapsed time	prior_question_had_explanation, time, elapsed time
Đầu vào phản hồi	Câu hỏi + kết quả	Câu hỏi + kết quả + thời gian	Câu hỏi + kết quả + thời gian + giải thích trước
Mục tiêu	Dự đoán đúng/sai	Dự đoán đúng/sai	Dự đoán đúng/sai
Accuracy	0.7178	0.7252	0.7240

Phiên bản SAINT+ của Shin et al. tích hợp nhiều đặc trưng thời gian như lag time và time-on-task, giúp cải thiện khả năng mô hình hóa tiến trình học tập. Trong nghiên cứu này, việc thử nghiệm với lag time và prior_question_had_explanation nhằm đánh giá tác động của thời gian phản hồi và việc có lời giải thích trước đó đến khả năng dự đoán.

4.3. Phân tích ảnh hưởng của hành vi người học

Trong nỗ lực cải thiện hiệu suất dự đoán, chúng tôi thử nghiệm bổ sung đặc

trưng prior_question_had_explanation - một biến nhị phân phản ánh việc người học có được giải thích trước khi trả lời câu hỏi hay không. Mục tiêu là kiểm tra xem yếu tố này có giúp mô hình SAINT+ phân biệt tốt hơn giữa các phản hồi đúng và sai.

Tuy nhiên, kết quả (Bảng 4) thực nghiệm cho thấy ảnh hưởng của đặc trưng này là không đáng kể. Khi thêm prior_question_had_explanation vào luồng huấn luyện, các chỉ số kiểm định gần như không thay đổi so với mô hình gốc:

Bảng 4: So sánh kết quả của cùng mô hình với 2 đầu vô khác nhau

Mô hình	ACC	AUC
SAINT+	0.7240	0.7600
SAINT+ và biến prior_question_had_explanation	0.7226	0.7578

Sự khác biệt giữa hai thiết lập là rất nhỏ và không mang ý nghĩa thống kê rõ rệt. Điều này cho thấy rằng trong bối cảnh dữ liệu hiện tại, đặc trưng prior_question_had_explanation không đóng vai trò quan trọng trong việc cải thiện năng lực dự đoán của mô hình. Một lý do có thể là thông tin này không được phản ánh đầy đủ trong hành vi tương tác của người học, hoặc mức độ ảnh hưởng của lời giải thích không đồng nhất giữa các cá nhân.

Kết quả này gợi ý rằng việc lựa chọn đặc trưng đầu vào cần được cân nhắc kỹ

luồng, và không phải mọi biến liên quan đến ngữ cảnh đều mang lại giá trị gia tăng cho mô hình truy vết tri thức.

4.4. Hướng cải tiến trong tương lai

Mô hình SAINT+ đã cho thấy hiệu suất dự đoán ổn định trong bài toán truy vết tri thức, đặc biệt với các phản hồi dạng đúng-sai (nhị phân). Tuy nhiên, một hạn chế đáng chú ý là khả năng mở rộng của mô hình đối với các dạng đánh giá phức tạp hơn. Trong thiết lập hiện tại, đầu ra của mô hình chỉ phản ánh xác suất nhị

phân (0 hoặc 1), điều này phù hợp với các câu hỏi trắc nghiệm nhưng chưa đủ để mô hình hóa các dạng phản hồi mềm như bài luận, bài tập mở, hoặc các câu hỏi đánh giá theo thang điểm.

Trong thực tế giáo dục, phản hồi của người học không phải lúc nào cũng tuyệt đối đúng hoặc sai. Ví dụ, một học sinh có thể đạt 80% độ chính xác trong một bài viết hoặc thể hiện mức hiểu biết ở mức 0.9 trong một bài tập phân tích. Việc mô hình hóa các phản hồi này dưới dạng liên tục (continuous feedback) sẽ giúp hệ thống đánh giá năng lực người học một cách tinh tế và sát thực hơn. Hướng cải tiến tiềm năng bao gồm:

- Mở rộng đầu ra mô hình sang dạng hồi quy (regression) để dự đoán điểm số hoặc mức độ hiểu thay vì chỉ phân loại đúng-sai.

- Tích hợp các dạng dữ liệu mềm như thang điểm, phản hồi văn bản, hoặc đánh giá của giáo viên để huấn luyện mô hình đa nhiệm (multi-task learning).

- Thiết kế kiến trúc phù hợp với phản hồi liên tục, chẳng hạn như sử dụng hàm mất mát MSE (Mean Squared Error) hoặc KL-divergence thay cho BCE.

- Áp dụng mô hình trong các ngữ cảnh đánh giá mở, như hệ thống chấm điểm bài luận hoặc phân tích tiến trình học tập theo thời gian.

Những cải tiến này không chỉ giúp mô hình SAINT+ mở rộng phạm vi ứng dụng mà còn tăng cường khả năng cá nhân hóa, hỗ trợ người học theo cách sâu sắc và linh hoạt hơn.

V. Kết luận

Nghiên cứu này đã triển khai và đánh giá mô hình SAINT+ trong bối cảnh truy vết tri thức cá nhân hóa, sử dụng tập dữ liệu

EDNet với quy mô lớn và độ phức tạp cao. Mô hình được xây dựng dựa trên kiến trúc Transformer phân tách encoder-decoder, kết hợp với các đặc trưng hành vi như thời gian làm bài và phản hồi của người học. Quá trình huấn luyện cho thấy SAINT+ đạt hiệu suất ổn định với $AUC = 0.760$ và độ chính xác $= 0.724$, phản ánh khả năng phân biệt hợp lý giữa các phản hồi đúng và sai.

Tuy nhiên, mô hình hiện tại chỉ xử lý phản hồi nhị phân (0 hoặc 1), điều này giới hạn khả năng ứng dụng trong các dạng đánh giá phức tạp hơn như bài luận, bài tập mở, hoặc các câu hỏi theo thang điểm. Việc mô hình hóa các phản hồi này sẽ giúp hệ thống đánh giá năng lực người học một cách tinh tế và sát thực hơn.

Ngoài ra, thử nghiệm bổ sung đặc trưng `prior_question_had_explanation` không mang lại cải thiện rõ rệt, cho thấy rằng không phải mọi yếu tố ngữ cảnh đều có giá trị gia tăng trong mô hình hóa tri thức. Điều này nhấn mạnh tầm quan trọng của việc lựa chọn đặc trưng đầu vào một cách có cơ sở và phù hợp với mục tiêu sự phạm.

Tổng thể, SAINT+ là một bước tiến đáng kể trong việc ứng dụng mô hình attention vào giáo dục cá nhân hóa. Dù còn một số hạn chế, nghiên cứu này đặt nền móng cho các hệ thống học tập thông minh có khả năng thích ứng tốt hơn với nhu cầu và hành vi của từng người học.

Tài liệu tham khảo

- [1]. Choi, Y., Lee, Y., Shin, D., Cho, J., Park, S., Lee, S., Baek, J., Bae, C., Kim, B., & Heo, J. (2020). *EdNet: A Large-Scale Hierarchical Dataset in Education*. In *Artificial Intelligence in Education* (pp. 69-73). Springer.
- [2]. Min, S., Lee, J., & Yoon, S. (2021). SAINT: Self-Attentive Interactive Model for Knowledge Tracing.

- Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '21)*, 1989-1997.
- [3]. Carpenter, S. K., Cepeda, N. J., Rohrer, D., Kang, S. H. K., & Pashler, H. (2012). Using spacing to enhance diverse forms of learning: Review of recent research and implications for instruction. *Educational Psychology Review*, 24(3), 369-378.
- [4]. Santoro, H. (2021). The neuroscience behind the spacing effect. *BrainFacts*.
- [5]. Teschner, J. (2024). Understanding decay theory: Exploring memory's natural erosion. *International Journal of School and Cognitive Psychology*, 11(6), 1-2.
- [6]. Ebbinghaus, H. (1913). *Memory: A Contribution to Experimental Psychology*. Teachers College, Columbia University. (Original work published 1885)
- [7]. Leppink, J., & Pérez-Fuster, P. (2019). Mental effort, workload, time on task, and certainty: Beyond linear models. *Educational Psychology Review*, 31, 421-438.
- [8]. Lee, J. (2019). Time-on-task as a measure of cognitive load in TBLT. *The Journal of Asia TEFL*, 16(3), 958-969.

INTERPRETABLE TEMPORAL MODELING IN STUDENT KNOWLEDGE TRACING: INSIGHTS FROM SAINT+ AND EDNET

Pham Hoang An²

Abstract: Artificial intelligence is emerging as a powerful tool in shaping personalized learning experiences, particularly through models that trace and predict learners' knowledge states over time. Among these, the SAINT (Separated Attention-based Knowledge Tracing) model achieves high accuracy by leveraging temporal features such as task duration and intervals between responses. However, its opaque decision-making nature poses challenges for educators seeking interpretability and trust in AI systems. This study analyzes the interpretability of SAINT's temporal attention mechanism using the large-scale EdNet dataset, which contains over 130 million student interactions across diverse learning environments. We focus on how temporal dynamics influence prediction quality, while visualizing attention weights to explain how time-based behavioral patterns shape learner modeling. Results show that transparent temporal modeling not only improves predictive performance but also fosters trust and alignment with pedagogical intuition. This work is particularly relevant for educators in resource-constrained regions such as Vietnam, where interpretable and accessible AI tools can help promote equitable and feasible personalized education.

Keywords: SAINT+, EdNet, personalized learning, knowledge tracing

² Ho Chi Minh city Open University