

인지과학적 망각 효과를 반영한 트랜스포머 기반 지식 추적 모델

박주영*¹, 이태권*², 배지훈*³

Transformer-based Knowledge Tracing Model Reflecting Cognitive Science-Informed Forgetting Effects

Ju-Yeong Park*¹, Tae-Gwon Lee*², and Ji-Hoon Bae*³

본 논문은 한국교원대학교 2025년 국립대학 육성사업 교육연구프로그램(ERP)의 지원을 받아 수행되었습니다.

요 약

최근 지식 추적(KT, Knowledge Tracing)은 온라인 학습 환경에서 학습자의 지식 상태를 추정하고 개인화된 학습 경로를 제공하는 핵심 기술로 주목받고 있다. 하지만, 기존의 Self-Attention 기반 KT 모델은 학습 시퀀스의 인지적 망각 과정을 충분히 반영하지 못하는 한계가 있다. 이에 본 연구는 ACT-R 이론의 시간 감쇠 원리를 SAKT 구조에 통합한 새로운 SAKT-MI 모델을 제안한다. 제안 모델은 문제 유사도와 시간 경과에 따른 기억 강도 간의 복잡한 상호작용을 학습하는 어텐션 재구성 메커니즘을 도입함으로써, 학습자의 망각 패턴을 효과적으로 모델링할 수 있도록 한다. 두 공개 데이터셋 실험에서 제안 모델은 기존 SAKT보다 전반적으로 우수한 예측 성능과 안정적인 학습 수렴 특성을 보였다. 본 연구는 인지과학적 망각 이론을 딥러닝 기반 KT 구조에 결합함으로써, 교육 데이터 마이닝 분야에서 지식 추적의 예측 성능 향상과 해석 가능성을 제시한다.

Abstract

Recently, Knowledge Tracing (KT) has attracted significant attention as a core technique for estimating learners' knowledge states and providing personalized learning paths in online learning environments. However, existing self-attention-based KT models have limitations in sufficiently reflecting cognitive forgetting processes in learning sequences. To address these issues, this study proposes a novel model, Self-Attentive Knowledge Tracing with Memory Integration (SAKT-MI), which integrates the temporal decay principle of the ACT-R theory into the original SAKT architecture. The proposed model introduces a reconstructed attention mechanism that learns the complex interaction between item similarity and retention strength over time, thereby enabling effective modeling of learners' forgetting patterns. Experiments on two public datasets demonstrate that the proposed model consistently outperforms the baseline SAKT model in terms of predictive accuracy and stability. This research demonstrates that combining cognitive forgetting theory with deep learning-based KT can improve both predictive accuracy and interpretability in educational data mining.

Keywords

knowledge tracing, ACT-R, SAKT, SAKT-MI, time-decay weight

* 한국교원대학교 컴퓨터교육과(*³ 교신저자)

- ORCID¹: <https://orcid.org/0009-0000-9616-9713>

- ORCID²: <https://orcid.org/0009-0000-2089-9285>

- ORCID³: <http://orcid.org/0000-0002-0035-5261>

• Received: Oct. 31, 2025, Revised: Nov. 13, 2025: Accepted: Nov. 16, 2025

• Corresponding Author: Ji-Hoon Bae

Dept. of Computer Education, 250 Taeseongtabyeon-ro, Gangnae-myeon, Heungdeok-gu, Gyeongju-si, Chungbuk 28173, Korea

Tel.: +82-43-230-3705, Email: baejh@knue.ac.kr

I. 서 론

지난 수십 년간 지능형 튜터링 시스템(ITS, Intelligent Tutoring Systems)과 대규모 공개 온라인 강좌(MOOCs, Massive Open Online Courses)를 포함한 e-러닝 시스템은 학습자의 상호작용 데이터를 수집 및 분석하여 개인화된 학습을 지원해왔으며, COVID-19 이후 온라인 학습의 중요성은 더욱 부각되었다[1]. 이러한 환경에서 학습자의 지식 상태를 정밀하게 진단하고 미래 성취를 예측하는 것은 맞춤형 교육 제공의 핵심 과제이다[2]. 이를 해결하기 위한 대표적인 기술이 지식 추적(KT, Knowledge Tracing)으로, 학습자의 과거 문제 풀이 기록을 활용하여 특정 개념에 대한 숙련도를 추정하고 응답 결과를 예측한다[3]. 정확한 KT는 불필요한 반복 학습을 줄이고 학습자 수준에 적합한 문제 제시 등 맞춤형 학습 경로 최적화에 기여할 수 있다[4].

KT 연구는 초기의 확률적 모형(BKT, Bayesian Knowledge Tracing)에서 딥러닝 기반 순환신경망(RNN, Recurrent Neural Networks)을 활용한 DKT(Deep Knowledge Tracing)[4]와 DKVMN(Dynamic Key-Value Memory Networks)[5]으로 발전하였다. 이러한 방법은 연속적인 지식 상태 추적에 성과를 보였으나, 데이터 희소성 환경에서 일반화 성능이 저하되는 한계가 보고되었다[3]. 최근에는 트랜스포머(Transformer) 기반 SAKT(Self-Attentive Knowledge Tracing)가 현재 문항(Query)과 과거 상호작용(Key/Value)의 유사도를 계산하여 관련성 높은 기록만 선택적으로 고려함으로써 RNN 대비 안정성과 효율을 개선하였다[3]. 하지만 SAKT는 시간 및 위치 정보를 단순 위치 인코딩에 의존하기 때문에, 실제 로그의 불규칙한 시간 간격이 초래하는 망각 효과를 충분히 반영하지 못하는 한계가 있다[6].

이러한 한계를 극복하기 위해 본 연구에서는 ACT-R(Adaptive Control of Thought - Rational) 인지 아키텍처의 기억 및 망각 원리를 기존 SAKT의 트랜스포머 기반 지식 추적 모델에 통합한 SAKT-MI(Self-Attentive Knowledge Tracing with Memory Integration) 모델을 제안한다. ACT-R은 인간의 인지 과정을 모사하는 대표적인 인지 과학 모델로, 학습 정보가 시간 경과에 따라 약화되지만, 빈

도와 최근성(Recency)에 따라 강화된다는 원리를 구조화하며[7], 이는 망각 효과를 충분히 반영하지 못하는 SAKT의 한계를 보완할 수 있는 이론적 근거를 제공한다. 이에 본 연구는 ACT-R 원리를 어텐션 구조에 통합하여, 단순 정오답 패턴을 넘어 최근성의 효과를 반영하는 새로운 지식 추적 모델을 구축하고자 한다. 본 논문의 주요 기여는 다음과 같다.

- 1) 인지 과학 이론(ACT-R)을 트랜스포머 기반 KT 모델에 접목하여 시간 및 망각 효과를 구조적으로 반영하는 새로운 방법을 정의하고 구현하였다.
- 2) 공개 학습 로그 데이터를 활용한 실험을 통해 제안 모델의 성능 개선을 입증하였다.

본 논문의 구성은 다음과 같다. 2장에서는 관련 연구로 트랜스포머 기반 KT 모델과 ACT-R 이론을 소개하고, 3장에서는 본 연구에 활용한 데이터를 설명한다. 4장에서는 제안 모델을 상세히 기술하고, 5장에서는 실험 결과를 제시하고 논의한다. 마지막으로 6장에서는 결론과 향후 연구 방향을 제시한다.

II. 관련 연구

2.1 트랜스포머

트랜스포머는 2017년에 제안된 아키텍처로, 기존의 RNN이나 합성곱 신경망(CNN, Convolutional Neural Network) 대신 셀프 어텐션(Self-attention)에 기반한다. 이 모델은 입력 시퀀스를 병렬적으로 처리하면서도 위치 간 의존성을 효과적으로 학습할 수 있기 때문에, RNN의 병렬화 한계와 장기 의존성 문제를 동시에 해결하였다. 트랜스포머는 주로 인코더(Encoder)와 디코더(Decoder) 구조로 구성되며, 멀티헤드 어텐션(Multi-head attention)과 포지션 인코딩(Positional encoding)을 통해 다양한 관계와 순서 정보를 반영한다[8]. 이러한 특성 덕분에 기계번역(Machine translation)을 포함한 자연어처리(NLP, Natural Language Processing) 전반에서 탁월한 성능을 보여주고 있다[9-10]. 특히 길고 불규칙한 학습자 상호작용 데이터에서 중요한 상호작용 추출에 강점을 발휘하여 RNN 기반 모델의 한계를 보완하는 대안으로 주목받고 있다[11].

2.2 SAKT 지식 추적 모델

SAKT는 2019년에 제안된 모델로, RNN 대신 셀프 어텐션 메커니즘만으로 학습자의 지식 상태를 추적한다[3]. 기존 RNN 기반 모델이 장기 의존성 반영이 어렵고 데이터 희소성에서 성능이 불안정해지는 한계를 극복하기 위해, SAKT는 셀프 어텐션을 통해 학습자가 푸는 현재 문제(Query)와 과거에 풀었던 문제(Key/Value) 간의 관련성을 계산한다. 이로써 현재 문제와 밀접한 상호작용만 선택적으로 고려하여 불필요한 정보는 제거한다. SAKT는 병렬 처리가 가능하고 학습 속도가 빠르며, 기존 RNN 기반 모델 대비 평균 약 4% 이상 높은 AUC 성능을 제공하였다. 특히 데이터가 부족한 환경에서도 안정적인 성능을 유지하여[3], KT 분야에서 트랜스포머 기반 모델의 가능성을 입증한 대표적 연구로 평가된다[2].

2.3 ACT-R 모델

ACT-R은 Anderson이 제안한 대표적인 인지 아키텍처로, 인간의 학습과 기억 과정을 설명하는 이론적 기반을 제공한다. 이 이론은 인간의 인지 과정이 어떻게 지식을 습득, 사용, 망각하는지를 설명하며, 지식의 접근 가능성은 단순히 저장 여부가 아니라, 얼마나 자주, 그리고 얼마나 최근에 사용되었는지에 따라 결정된다는 점을 강조한다. 이러한 특성을 설명하는 핵심 개념이 기저 활성화도(Base-level activation)이며, 이는 식 (1)과 같이 표현된다.

$$B_i = \ln \sum_{j=1}^n t_j^{-d} \quad (1)$$

여기서 B_i 는 특정 지식 단위(Chunk) i 의 활성화 정도를, t_j 는 현재로부터 과거 j 번째 노출까지의 학습 경과 시간을, d 는 망각률(Decay rate)을 각각 의미한다. 이 수식은 세 가지 핵심 원리를 내포한다[7].

- 1) 학습의 멱법칙(Power law of learning): 반복 학습을 통해 수행 능력이 점진적으로 향상된다.
- 2) 망각의 멱법칙(Power law of forgetting): 시간이 지남에 따라 기억이 급격히 감소하다가 이후 완만하게 감소한다.

3) 연습-시간 곱효과(Multiplicative effect of practice and retention): 학습 효과는 단순히 연습량과 시간의 합이 아니라 두 요인의 상호작용에 의해 결정된다.

이러한 특징은 학습자의 지식 변화 추적을 목표로 하는 KT 연구와 연결되며, 실제로 ACT-R은 초기 인지 튜터(Cognitive tutor) 시스템 개발에도 활용되어 학습 효과 향상에 기여하였다[12].

III. 데이터셋 및 전처리

3.1 데이터셋 소개

본 연구는 지식 추적 연구에서 널리 활용되는 두 가지 공개 데이터셋인 ASSISTChall(ASSISTments 2017 Challenge)와 Algebra 2005-2006을 사용하였다. 각 데이터셋의 특징은 다음과 같다.

3.1.1 ASSISTChall

ASSISTments 2017 Challenge 데이터셋은 무료 온라인 개인교습 플랫폼인 ASSISTments에서 수집된 초·중등 수학 문제 풀이 기록으로, 2017년 ASSISTments Data Mining Competition을 통해 공개되었다[2]. 해당 데이터셋은 총 1,709명의 학습자, 3,162개의 문항, 102개의 학습 개념(KC, Knowledge Component), 그리고 942,816건의 상호작용(Interaction)으로 구성된다. 여기서 상호작용은 학습자의 단일 문제 풀이 행위를 의미하며, 본 연구에서는 “studentId”, “skill”, “problemId”, “correct”, “startTime”의 다섯 가지 주요 속성을 활용하였다. 평균적으로 문항당 약 298회의 응답이 기록되어 있어, ASSISTments 계열 데이터셋 가운데 밀집도가 가장 높다는 특징을 가진다.

3.1.2 Algebra 2005-2006

Algebra 2005 - 2006 데이터셋은 KDD Cup 2010 Educational Data Mining Challenge에서 제공된 자료로, 미국 Carnegie Learning Inc.의 지능형 튜터링 시스템을 통해 수집된 학생들의 대수(Algebra) 문제 풀이 기록으로 구성되어 있다[2]. 다른 데이터셋과

달리 각 문제는 여러 개의 하위 문항(Sub-items)으로 세분화되어 있으며, 평균적으로 문제당 약 750회 이상의 응답이 기록되어 밀도가 높은 편이다. 본 연구에서는 두 데이터셋의 인터페이스 일관성을 위해, 원본 Algebra 2005-2006의 속성을 ASSISTChall 데이터셋과 동일한 의미를 갖는 “studentId”, “problemId”, “skill”, “startTime”, “correct”의 다섯 개 컬럼으로 변환하여 사용하였다. 위에서 소개한 두 데이터셋의 특징 비교는 표 1과 같다.

표 1. 데이터셋의 주요 규모 및 구조적 통계 비교
Table 1. Comparison of key scale and structural statistics of datasets

Dataset	Students	Interaction	KC	Question
ASSISTChall	1,709	942,816	102	3,162
Algebra 2005-2006	571	607,014	271	173,113

3.2 데이터셋 항목별 특성

본 연구에서 사용된 데이터셋의 주요 속성(Feature)은 다음과 같이 정의된다.

- 1) studentId: 학습자를 고유하게 식별하는 ID로, 학습자별 문제 풀이 이력을 구분에 활용된다.
- 2) problemId: 각 문항을 구분하는 고유 식별자로, 문항 단위 정오답 분석에 활용된다.
- 3) skill: 문항과 관련된 KC로, 학습자의 개념별 성취 수준 평가에 활용된다.
- 4) startTime: 문항 풀이 시점을 나타내는 타임스탬프로, 시퀀스 데이터 생성에 사용된다.
- 5) correct: 풀이 결과의 정답 여부(정답: 1, 오답: 0)이며, 지식 추적 모델의 예측 대상 변수로 활용된다.

3.3 데이터 전처리

본 연구는 구조가 상이한 ASSISTChall과 Algebra 2005-2006 두 데이터셋을 활용하기 때문에, 비교 및 분석의 일관성을 위해 “studentId”, “problemId”, “skill”, “startTime”, “correct”의 공통 속성 형식으로 정리하고 다음의 주요 전처리 과정을 수행하였다.

- 1) ASSISTChall: “skill”, “correct”, “startTime” 컬럼

에 결측치가 포함된 행을 제거하고, 중복된 “studentId”, “problemId”, “skill” 조합을 삭제하였다. 이후 각 학생별 풀이 기록을 “startTime” 기준으로 정렬하였으며, “startTime”의 가장 이른 시점을 0으로 맞추고 뒤 초 단위를 일(Day) 단위로 변환하여 시간 정보를 정규화하였다.

2) Algebra 2005-2006: ASSISTChall과 동일한 속성 형식으로 통일하기 위해 먼저, “Problem Name” 컬럼과 “Step Name” 컬럼을 결합하여 새로운 “problemId”를 생성하고, 복수로 기록된 스킬 값은 하나의 고유 스킬로 통합하였다. 이어서 “First Transaction Time” 컬럼은 UNIX 초 단위로 변환 후 “startTime”으로 변경하고, 결측치 및 비정상 값을 제거하였다. 마지막으로, ASSIST 데이터셋과 동일한 전처리 과정을 적용하였다.

IV. 제안 방법

본 장에서는 제안 모델의 전체 구조와 계산 과정을 설명한다. 그림 1은 모델의 전체 구조도로, 예측 대상 문항 시퀀스는 Query로, 과거 상호작용은 Key/Value로 임베딩 변환된다. 기존 SAKT가 Query-Key의 점곱 유사도(Scaled dot-product)만으로 어텐션 점수를 계산하는 데 반해, 본 연구는 ACT-R의 기저 활성도를 반영한 시간 감쇠(Time-decay) 채널을 추가한다. 구체적으로, 원래의 점곱 어텐션 맵(Attention map)과 타임스탬프 간격으로부터 계산한 시간 감쇠 가중치 맵(Time-decay weight map)을 결합하고, 이를 소규모 MLP(Multi Layer Perceptron)에 적용하여 재구성 어텐션 맵(Reconstructed attention map)을 생성한다. 이 맵은 causal mask와 softmax 정규화를 거쳐 최종 어텐션 가중치로 변환되고, 이를 Value와의 가중합을 통해 시점별 문맥 벡터(Context vector)를 얻는다. 마지막으로 문맥 벡터는 잔차 연결(Residual connection)을 포함하는 순방향 신경망(FFN, Feed-Forward Network)과 층 정규화(LN, Layer Normalization)로 구성된 비선형 변환을 거친 후, sigmoid를 통해 최종 정답 확률을 출력한다. 이후에서는 문제 정의와 표기를 먼저 정리한 뒤, 제안 모델의 계산 흐름을 단계별로 상세히 기술한다.

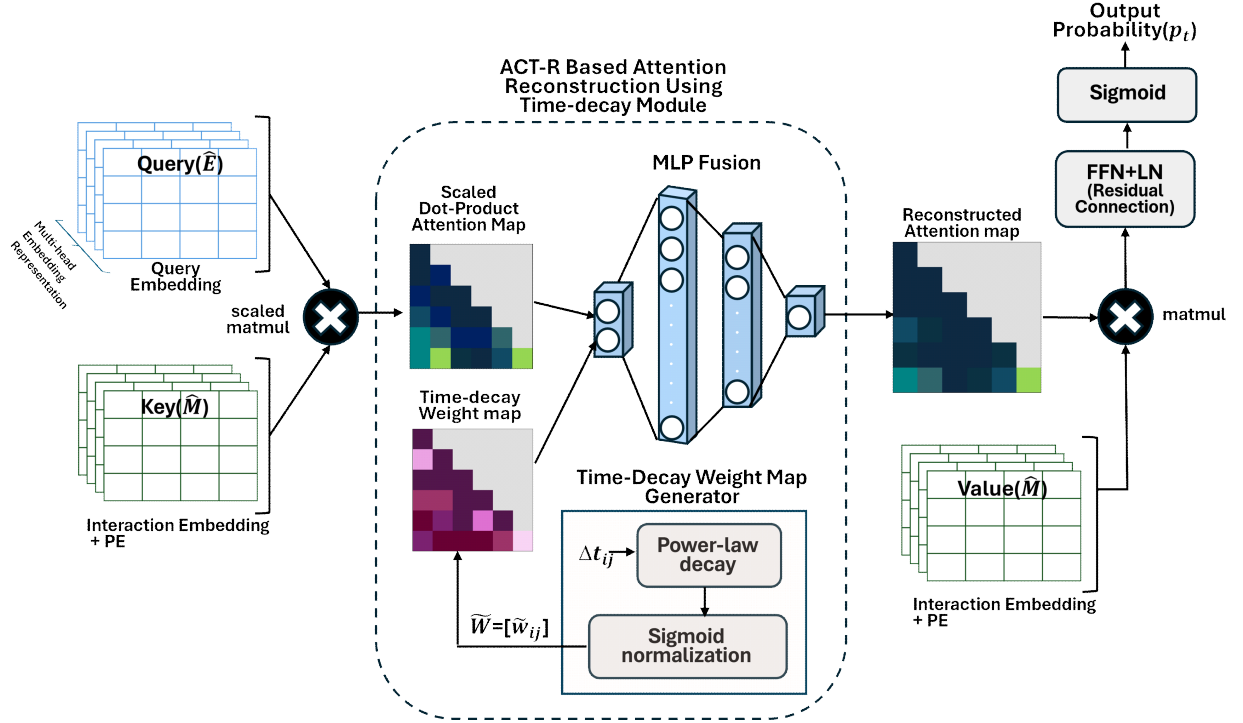


그림 1. 제안된 SAKT-MI 모델의 구조

Fig. 1. Architecture of the proposed SAKT-MI model

4.1 문제 정의 및 기호 표기

본 연구에서 제안하는 모델은 학습자의 이전 상호작용 시퀀스 $X_{<t}$ 를 기반으로, 현재 문항 e_t 에 대한 정답 확률을 예측하고자 한다. 한 학습자의 상호작용 시퀀스 $X = (x_1, x_2, \dots, x_i)$ 는 시점 i 에서의 문항 e_i 와 정오답 여부 $r_i \in \{0, 1\}$ 의 쌍 $x_i = (e_i, r_i)$ 로 구성된다. 모델은 길이 L 의 고정 윈도우로 분할된 $X_{<t} (x_{t-L}, \dots, x_{t-1})$ 를 입력받아 학습자의 과거 상호작용 기록을 학습에 활용한다. 각 x_i 가 문항과 정오답 여부를 동시에 포함하기 때문에, 모델은 학습자가 푼 문제와 그 정답 여부를 모두 참조할 수 있다. 반면, 예측 대상인 현재 문항 e_t 는 풀이 결과가 관측되지 않았으므로, 해당 문항을 포함하는 시퀀스를 $(e_1, \dots, e_n)$ 와 같이 문항 ID로만 표현한다. 즉, Key/Value 시퀀스는 “문항-정답”쌍을 포함하는 $x_{<t}$ 로, Query 시퀀스는 정답 여부가 미지인 문항을 나타내기 때문에 e_t 로 구분된다. 따라서 모델은 각 시점 t 에서 현재 문항 e_t 의 정답 확률 p_t 을 식 (2)와 같이 계산한다.

$$p_t = P(r_t = 1 | x_{<t}, e_t), \quad t = 1, \dots, n, \quad (2)$$

여기서 $x_{<t}$ 는 학습자의 과거 상호작용 시퀀스를, e_t 는 예측 대상 문항을 각각 의미한다. 또한, 제안 모델 학습에 사용되는 배치 크기는 B , 임베딩 차원은 d , 멀티헤드 수는 h , 그리고 각 헤드의 차원은 $d_k = d/h$ 로 표기한다.

4.2 상호작용 임베딩과 문항 임베딩

본 연구는 SAKT와 동일하게 상호작용과 문항을 각각 독립적으로 임베딩한다. 먼저 시점 t 에서의 상호작용은 문항 ID e_t 와 정오답 r_t 를 결합하여 상호작용 인덱스 $y_t = e_t + r_t \times E$ 로 정의한다. 여기서 E 는 전체 문항의 개수를 의미한다. 이 인덱스는 학습자가 특정 문항에 정답 또는 오답으로 반응했는지를 구분하며, 가능한 상호작용의 총 개수는 $2E$ 가 된다. 이 인덱스를 바탕으로 상호작용 임베딩 행렬 $\hat{M} \in \mathbb{R}^{2E \times d}$ 에 저장된 임베딩에 식 (3)과 같이 위치 임베딩 P_t 를 더하여 $\hat{M}_t$ 를 생성한다.

$$\hat{M}_t = M[y_t] + P_t, \quad (3)$$

여기서, $P \in \mathbb{R}^{L \times d}$ 는 시퀀스 내 문항의 순서 정보를 반영하기 위한 학습 가능한 절대 위치 임베딩 행렬을 의미한다.

다음으로, 문항 자체의 의미적 정보를 반영하기 위하여 시점 t 에서의 문항 임베딩 $\hat{E}_t$ 는 식 (4)와 같이 정의되며, 문항 임베딩 행렬 $E \in \mathbb{R}^{E \times d}$ 에서 각 문항 ID에 해당하는 벡터를 추출한다.

$$\hat{E}_t = E[e_t] \quad (4)$$

따라서 시퀀스 단위의 상호작용 임베딩과 문항 임베딩은 각각 $\hat{M}, \hat{E} \in \mathbb{R}^{L \times d}$ 로 표현된다. 이후 $\hat{M}$ 과 $\hat{E}$ 는 멀티헤드 어텐션 구조에 맞게 (B, h, L, d_k) 형태의 텐서로 변환되며, 이때 $\hat{M}$ 은 Key/Value로, $\hat{E}$ 는 Query로 사용된다.

4.3 ACT-R 기반 시간 감쇠 가중치맵 설계

기존 위치 기반 인코딩은 시퀀스 내 인덱스 간 거리가 가까울수록 높은 가중치를 부여하지만, 실제 학습 로그는 문제 풀이 간격이 불일정하고 학습자마다 응답 시간이 크게 다르기 때문에, 단순 위치 정보만으로는 시간 경과에 따른 기억의 감쇠 효과 (Forgetting effect)를 충분히 반영하기 어렵다. 이에 본 연구는 ACT-R 인지 과학 이론의 기저 활성화 원리를 어텐션 구조에 통합하였다. 즉, 시간 경과에 따라 기억이 점차 약화되는 현상을 수학적으로 모델링함으로써, 기존 어텐션이 반영하지 못하는 망각 효과를 보완한다.

구체적으로, 시점 i 의 Query가 과거 시점 $j < i$ 의 Key를 참조할 때, 두 시점 간 경과 시간 $\Delta t_{ij} = t_i - t_j > 0$ 에 대해 식 (5)와 같은 시간 감쇠 가중치(Time-decay weight)를 정의한다.

$$w_{ij} = (\Delta t_{ij})^{-\frac{1}{2}}, \quad w_{ij} = 0 \text{ if } j \geq i \quad (5)$$

여기서 w_{ij} 는 시간이 지남에 따라 점차 감소하며, 최근 상호작용일수록 큰 값을 갖는다. 이는 ACT-R

의 망각의 역법칙(Power law of forgetting)을 반영한 것이다. 하지만 학습 로그의 시간 간격 분포가 매우 불균형하기 때문에, 모델의 안정적인 학습을 위해 w_{ij} 에 로그 스케일로 변환하고 시그모이드(Sigmoid) 함수를 적용하여 정규화된 기저 활성화도 $\tilde{w}_{ij}$ 를 식 (6)과 같이 정의한다.

$$\tilde{w}_{ij} = \sigma(\log(w_{ij} + \epsilon)), \quad \epsilon > 0. \quad (6)$$

정규화된 $\tilde{w}_{ij}$ 는 최근 상호작용일수록 높은 값을 갖는 스칼라로 해석된다. 이는 단순 위치 인코딩과 달리, 동일한 인덱스 차이라 하더라도 풀이 시간 간격이 길 경우 더 강한 망각 효과가 반영되도록 시퀀스 내 각 시점 쌍 (i, j) 간의 시간적 의존성을 반영한다. 최종적으로 $\tilde{w}_{ij}$ 는 모든 어텐션 헤드가 동일한 시간 감쇠 패턴을 공유하도록 브로드캐스트(Broadcast)되어 식 (7)과 같이 4차원 텐서 형태의 기저 활성화도 맵으로 확장된다.

$$\tilde{W} = \text{Broadcast}(\tilde{w}) \in \mathbb{R}^{B \times h \times L \times L} \quad (7)$$

여기서, L 은 시퀀스 길이(Sequence length)를 나타낸다. 4차원 텐서 $\tilde{W}$ 는 그림 1의 “Time-decay Weight Map”에 해당하며, 결과적으로 “Original Attention Map”과 함께 결합되어 다음 절에서 제안하는 “Reconstructed Attention Map” 계산에 사용된다.

4.4 메모리 정보화 기반 재구성 어텐션맵 설계

기존 SAKT는 Query - Key 점곱 유사도 (Dot-product similarity)에 기반하여 어텐션 가중치를 산출한다. 반면 본 연구에서 제안하는 SAKT-MI 모델은 점곱 어텐션 맵(Scaled dot-product attention map)과 ACT-R 이론에 기반한 시간 감쇠 가중치맵을 통합하여, 시간 경과에 따른 망각 효과를 반영하는 메모리 정보화 기반 재구성 어텐션맵 (Memory-informed reconstruction attention map)을 생성한다. 기본 어텐션 맵은 각 헤드 h 에서 기존 SAKT와 동일하게 점곱 유사도 기반으로 어텐션 행렬을 계산하며, 이는 식 (8)과 같이 구성된다.

$$raw_{ij}^{(h)} = \frac{\langle Q_i^{(h)}, K_j^{(h)} \rangle}{\sqrt{d_k}} \quad (8)$$

여기서 Query/Key/Value는 각각 $Q^{(h)} = \hat{E} W_Q^{(h)}$, $K^{(h)} = \hat{M} W_K^{(h)}$, $V^{(h)} = \hat{M} W_V^{(h)}$ 로 표현된다. 이때 시간 감쇠 가중치맵은 4.3절에서 정의된 정규화된 시간 가중치 $\tilde{w}_{ij}$ 를 모든 헤드에 복제한 $\tilde{w}_{ij}^{(h)}$ 로 정의된다. 이 채널은 실제 풀이 시간 간격 Δt_{ij} 을 반영하여 망각 효과를 모델링하는 역할을 수행한다. 다음으로, 상기 두 채널에서 얻은 스칼라 입력 $raw_{ij}^{(h)}$ 와 $\tilde{w}_{ij}^{(h)}$ 은 마지막 차원을 기준으로 스택하여 입력 벡터를 구성하며, 이 벡터는 식 (9)로 표현된다.

$$z_{ij}^{(h)} = [raw_{ij}^{(h)}, \tilde{w}_{ij}^{(h)}] \in \mathbb{R}^2 \quad (9)$$

이를 통해 점곱 어텐션 맵 정보와 시간 감쇠 가중치맵 정보를 통합하여 학습할 수 있으며, 식 (10)과 같이 입력 벡터를 헤드별 소규모 MLP에 통과시켜 최종 어텐션 점수 $\ell_{ij}^{(h)}$ 를 계산한다.

$$\ell_{ij}^{(h)} = f_{\phi}^{(h)}(z_{ij}^{(h)}) \quad (10)$$

여기서 $f_{\phi}^{(h)}$ 는 표 3에서 구성된 소규모 MLP로, 헤드 간 파라미터를 공유하며, 각 $z_{ij}^{(h)}$ 에 대해 두 맵 정보의 비선형적 결합 효과를 학습한다.

4.5 재구성 어텐션 가중치 적용 및 문맥 벡터 산출

소규모 MLP에서 추출된 스칼라 정보 $\ell_{ij}^{(h)}$ 는 미래 시점 $j > i$ 의 Key를 참조하지 않도록 마스킹을 적용한 뒤, softmax 함수를 통해 정규화된 재구성 어텐션 가중치 $a_{ij}^{(h)}$ 로 변환된다. 이 과정은 식 (11)과 같이 주어진다.

$$a_{ij}^{(h)} = \text{softmax}(\ell_{ij}^{(h)}), \quad j < i. \quad (11)$$

계산된 가중치를 이용하여 Value 행렬 $V^{(h)}$ 와의 가중합을 수행함으로써 각 헤드별 문맥 정보 (Context information)를 얻을 수 있다. 이후 모든 헤드의 출력을 연결하여 최종 문맥 벡터 S 를 생성하며, 그 수식은 식 (12)와 같다.

$$S = \text{Concat}_h(\sum_{j < i} a_{ij}^{(h)} V_j^{(h)}) \in \mathbb{R}^{L \times d} \quad (12)$$

여기서 $V_j^{(h)}$ 는 시점 j 의 상호작용(문항 - 정오답 쌍) 임베딩 벡터를 의미하며, 학습자의 과거 풀이 이력에서 의미 정보를 전달하는 역할을 수행한다.

4.6 정답 확률 예측

문맥 벡터 S 는 2개의 층으로 구성된 위치별 순방향 신경망(FFN)을 통과한다. 이 과정에서 층 정규화와 잔차 연결이 적용되며, 그 출력은 식 (13)과 같이 표현된다.

$$H = \text{LN}(S + \text{FFN}(S)) \quad (13)$$

여기서 FFN은 식 (14)와 같은 두 층의 비선형 변환으로 정의된다.

$$\text{FFN}(x) = \text{Dropout}(\text{ReLU}(x W^{(1)} + b^{(1)})) W^{(2)} + b^{(2)} \quad (14)$$

마지막으로, 정규화된 은닉 표현 H 의 각 시점 벡터 h_i 는 선형 변환을 거쳐 시그모이드 함수에 입력되며, 그 결과 현재 문항에 대한 정답 확률 $p_t = \sigma(W_p h_t + b_p)$ 로 계산된다. 이를 통해 모델은 학습자의 과거 상호작용 시퀀스를 기반으로 현재 문항에 대한 정답 확률을 추론하게 된다.

V. 실험 결과

5.1 실험 설정

본 연구에서는 제안한 ACT-R 기반 SAKT-MI 모

델과 기존 SAKT 모델의 성능을 비교·검증하였다. 실험에는 공개 데이터셋인 ASSISTChall과 Algebra 2005 - 2006을 사용하였으며, 입력 속성은 3.2절에서 정의된 “studentId”, “problemId”, “skill”, “startTime”, “correct”으로 구성하였다. 모든 모델은 공정한 비교를 위해 입력 시퀀스 길이 N=100을 포함하여 표 2의 동일한 하이퍼파라미터로 학습되었으며, MLP 관련 설정은 표 3과 같다.

모델 학습에는 Adam(Adaptive Moment Estimation) 옵티마이저를 사용하였고, 4-Fold Group K-Fold 교차 검증(Group K-Fold cross validation)을 적용하여 성능을 평가하였다. 학습 과정에서 검증 손실이 최소가 되는 시점의 가중치를 저장하고 다음의 평가 지표들을 활용하여 최종 성능을 비교하였다.

1) AUC(Area under the ROC curve): 모델의 정오답 예측 능력을 종합적으로 평가

2) 정확도(Accuracy, ACC): 예측 결과의 정답 일치율을 측정

3) 검증 손실(Validation loss): 모델의 예측 오차를 정량적으로 평가

또한 학습 과정에서 재구성 어텐션 히트맵과 시간 감쇠 가중치맵을 시각화하여 어텐션 메커니즘의 변화를 분석하였으며, 에포크별 손실, AUC, 정확도를 그래프로 표현하여 학습 추세를 확인하였다.

5.2 실험 결과 및 분석

본 연구에서는 각 폴드별로 계산된 AUC, 정확도, 검증 손실을 기준으로 모델의 성능을 평가하였다. 각 모델의 폴드별 평균과 표준편차를 산출하여 전반적인 성능을 비교하였다.

표 2. 모델 학습에 사용된 주요 하이퍼파라미터
Table 2. Main hyperparameters used for model training

Hyperparameter	Value
Sequence length (L)	100
Batch size	256
Max epochs	100
Learning rate	1e-3
Dropout	0.2
Embedding dimension (d)	100
Number of Multi-heads (h)	5

표 3. 제안된 MLP 구조

Table 3. Proposed MLP architecture

MLP architecture (layers)	Activation function	Dropout
2→16→8→1	GELU	0.2

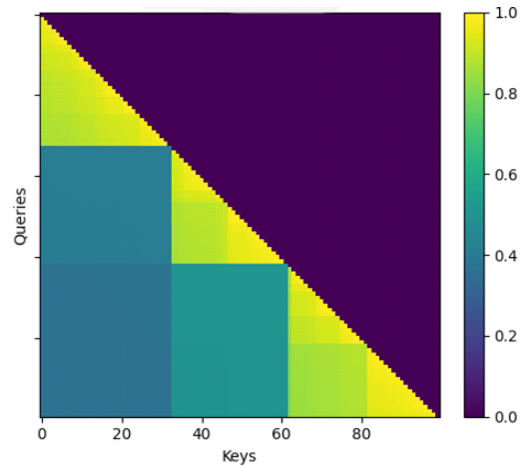


그림 2. SAKT-MI 모델의 시간 가중치 맵

Fig. 2. Time-weight map of the SAKT-MI model

먼저 모델 내 시간 감쇠 가중치맵을 통해 학습자의 기억 감쇠 효과가 어텐션 메커니즘에 어떠한 방식으로 반영되는지를 분석하였다. 그림 2는 시간 감쇠 가중치맵의 시각화 결과로, 문항 간 시간 간격에 따라 기억의 활성화 수준이 달라지는 양상을 보여준다. 이 가중치는 ACT-R 이론의 기억 감쇠 함수를 기반으로 계산되며, 시간이 경과할수록 기억 강도가 점진적으로 감소하는 특성을 가진다. 따라서 최근의 학습 경험에는 높은 가중치가, 오래된 경험에는 낮은 가중치가 부여되어, 모델이 학습자의 기억 지속성과 시간적 맥락을 동시에 고려할 수 있도록 한다.

이와 같은 시간 감쇠 가중치의 효과는 학습 과정에서의 손실 함수 및 예측 성능 변화에서도 확인된다. 그림 3과 그림 4는 각각 검증 손실과 AUC의 에포크별 변화를 나타낸다. 두 모델 모두 학습 초기에 손실이 급격히 감소하였으나, SAKT-MI 모델은 이후 손실이 더 낮은 수준에서 안정적으로 수렴하였다. 이는 시간 감쇠 가중치의 도입이 과적합을 완화하고 모델의 일반화 성능을 향상시킨 결과로 이해할 수 있다. 또한 그림 4와 같이, 제안된 SAKT-MI 모델은 학습 전반에 걸쳐 기존 SAKT보다 높은 AUC를 지속적으로 유지하였다. 특히 학습

초반에서 AUC가 빠르게 상승하는 양상을 보였는데, 이는 시간 감쇠 가중치가 초기 학습 단계에서 효율적인 정보 집증을 유도하여 학습 효율을 높였음을 의미한다. 이후에도 제안 모델은 안정적으로 수렴하며 전 학습 구간에서 기존 모델 대비 높은 예측 성능을 유지하였다.

표 4. ASSISTChall 데이터셋의 성능 비교

Table 4. Performance comparison on the ASSISTChall dataset

Model	AUC	ACC	Loss
SAKT	0.7066 ± 0.0043	0.6575 ± 0.0016	0.6182 ± 0.0016
SAKT-MI	0.7082 ± 0.0040	0.6587 ± 0.0017	0.6167 ± 0.0012

표 5. Algebra 2005-2006 데이터셋의 성능 비교

Table 5. Performance comparison on the Algebra 2005-2006 dataset

Model	AUC	ACC	Loss
SAKT	0.7890 ± 0.0050	0.8045 ± 0.0066	0.4388 ± 0.0092
SAKT-MI	0.7942 ± 0.0058	0.8063 ± 0.0065	0.4354 ± 0.0087

모델의 최종 성능은 각 폴드별 AUC, 정확도, 검증 손실을 기준으로 평가하였다. 두 데이터셋에 대한 4-Fold 교차검증 결과의 평균과 표준편차는 각각 표 4 및 표 5와 같이 산출되었다. 두 데이터셋 모두에서 제안한 SAKT-MI 모델은 기존 SAKT에 비해 AUC, 정확도, 손실 평가 항목 전반에서 향상된 성능을 제공하였다. Algebra 2005-2006 데이터셋의 경우, 제안된 SAKT-MI 모델은 AUC가 0.7942 ± 0.0058 로 기존 SAKT의 0.7890 ± 0.0050 보다 약 0.52%p 향상되었으며, 정확도 또한 0.8063으로 0.18%p의 개선을 보였다. ASSISTChall 데이터셋에서도 AUC가 0.7082 ± 0.0040 로 기존 SAKT의 0.7066 ± 0.0043 보다 0.16%p 향상되었다. 이러한 결과는 시간 감쇠 가중치가 학습자의 기억 감쇠 특성과 시간적 의존성을 효과적으로 반영하여 문항 간 관계를 보다 정교하게 학습했기 때문으로 해석된다. 결과적으로 제안된 SAKT-MI 모델은 학습자의 시간적 인지 맥락을 고려함으로써 예측 안정성과 정확도를 동시에 향상시

켰으며, 이는 지식 추적에서 인지적 요인을 통합하는 새로운 방향성을 제시한다.

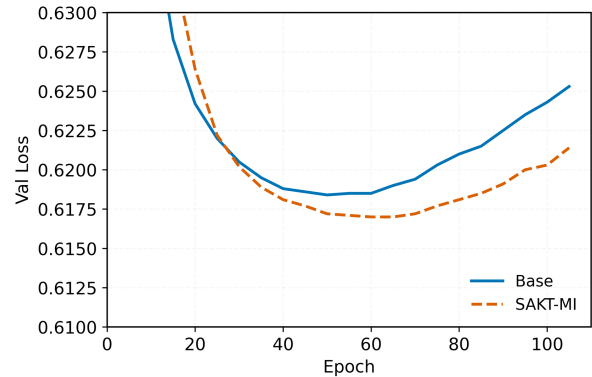


그림 3. 에포크별 검증 손실 비교

Fig. 3. Validation loss comparison across epochs

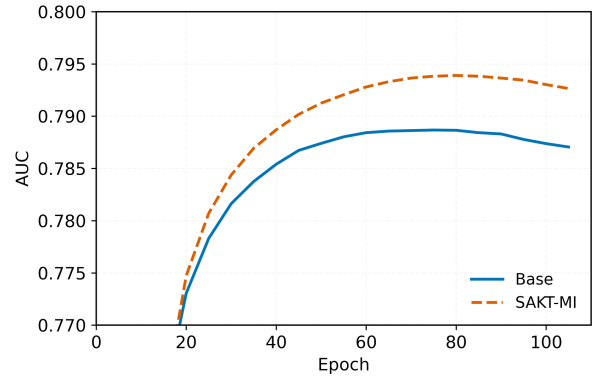


그림 4. 에포크별 AUC 비교

Fig. 4. AUC comparison across epochs

VI. 결 론

본 연구는 기존 SAKT 모델의 한계점인 시간적 망각 효과 반영 부족을 해결하기 위해, ACT-R 인지 아키텍처의 기저 활성화 원리를 도입한 SAKT-MI 모델을 제안하였다. 제안 모델은 학습자의 시간 의존적 기억 감쇠 특성을 어텐션 메커니즘에 통합하여, 인간의 인지 과정을 반영하는 지식 추적을 설계하였다. 실험 결과, 두 개의 공개 데이터셋 (ASSISTChall, Algebra 2005-2006)에서 제안 모델은 기존 SAKT 대비 일관된 성능 향상을 보였다. 특히 시간 감쇠 가중치 맵의 시각적 분석을 통해 모델이 실제로 학습자의 기억 감쇠 패턴을 효과적으로 학습하고 있음을 확인하였다.

본 연구 결과를 통해 지식 추적 분야에서 인지적 원리에 기반한 접근이 성능 향상뿐 아니라 모델 해석력 측면에서도 새로운 가능성을 제시함을 확인하였다. 앞으로는 다양한 최신 기술과의 결합, 그리고 데이터가 부족한 상황에서의 Zero-Shot 추론 등 새로운 기술적 패러다임과 융합을 통해 보다 혁신적인 지식 추적 모델 개발로 이어질 수 있을 것으로 기대한다. 또한 향후 연구에서는 학습자 및 개념별로 상이한 망각 특성을 반영할 수 있는 동적 감쇠율(Learnable decay parameter)의 통합을 통해, 모델의 개인화 및 일반화 능력을 한층 강화하는 방향으로 확장할 계획이다.

References

- [1] X. Zhou, Z. Zhang, X. Xie, and J. Zhang, "Deep learning based knowledge tracing in intelligent tutoring systems", *Scientific Reports*, Vol. 15, No. 21395, pp. 1-15, Jul. 2025. <https://doi.org/10.1038/s41598-025-07422-7>.
- [2] G. Abdelrahman, Q. Wang, and B. Nunes, "Knowledge Tracing: A Survey", *ACM Computing Surveys*, Vol. 55, No. 11, pp. 1-37, Feb. 2023. <https://doi.org/10.1145/3569576>.
- [3] S. Pandey and G. Karypis, "A Self-Attentive Model for Knowledge Tracing", *arXiv preprint arXiv:1907.06837*, Jul. 2019. <https://doi.org/10.48550/arXiv.1907.06837>.
- [4] C. Piech, J. Bassen, J. Huang, S. Ganguli, M. Sahami, L. J. Guibas, and J. Sohl-Dickstein, "Deep Knowledge Tracing", *Advances in neural information processing systems*, Montreal, Canada, Vol. 1, pp. 505-513, Dec. 2015. <https://dl.acm.org/doi/10.5555/2969239.2969296>.
- [5] J. Zhang, X. Shi, I. King, and D. yeung, "Dynamic Key-Value Memory Networks for Knowledge Tracing", *Proc. of the 26th International Conference on World Wide Web*, Perth Australia, pp. 765-774, Apr. 2017. <https://doi.org/10.1145/3038912.3052580>.
- [6] Y. Im, E. Choi, H. Kook, and J. Lee, "Forgetting-aware Linear Bias for Attentive Knowledge Tracing", *Proc. of the 32nd ACM International Conference on Information and Knowledge Management (CIKM)*, Birmingham, United Kingdom, pp. 3958-3962, Oct. 2023. <https://doi.org/10.1145/3583780.3615191>.
- [7] J. R. Anderson and C. D. Schunn, "Implications of the ACT-R Learning Theory: No Magic Bullets", R. Glaser (Ed.), *Advances in Instructional Psychology*, Mahwah, NJ:Erlbaum, Vol. 5, pp. 1-33, 2000.
- [8] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need", *Advances in Neural Information Processing Systems*, California, USA, pp. 6000-6010, Dec. 2017. <https://dl.acm.org/doi/10.5555/3295222.3295349>.
- [9] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, J. Davison, S. Shleifer, P. von Platen, C. Ma, Y. Jernite, J. Plu, C. Xu, T. Le Scao, S. Gugger, M. Drame, Q. Lhoest, and A. M. Rush, "Transformer: State-of-the-Art Natural Language Processing", *Proc. of the 2020 Conference on Empirical Methods in Natural Language Processing: Systems Demonstrations*, Online, pp. 38-45, Oct. 2020. <https://doi.org/10.18653/v1/2020.emnlp-demos.6>.
- [10] S. Lim and S. Youn, "A Study on Efficient Natural Language Processing Method based on Transformer", *The Journal of the Institute of Internet, Broadcasting and Communication (JIIBC)*, Vol. 23, No. 4, pp. 115-119, Aug. 2023. <https://doi.org/10.7236/JIIBC.2023.23.4.115>.
- [11] T. Liu, M. Zhang, C. Zhu, and L. Chang, "Transformer-based convolutional forgetting knowledge tracking", *Scientific Reports*, Vol. 13, No. 19112, pp. 1-8, Nov. 2023. <https://doi.org/10.1038/s41598-023-45936-0>.

- [12] S. Ritter, J. R. Anderson, K. R. Koedinger, and A. T. Corbett, "Cognitive Tutor: Applied research in mathematics education", *Psychonomic Bulletin & Review*, Vol. 14, No. 2, pp. 249-255, Apr. 2007. <https://doi.org/10.3758/bf03194060>.

저자소개

박 주 영 (Ju-Yeong Park)



2023년 3월 ~ 현재 :

한국교원대학교 컴퓨터교육과
학사과정

관심분야 : 인공지능, 추천시스템,
지식추적

이 태 권 (Tae-Gwon Lee)



2023년 3월 ~ 현재 :

한국교원대학교 컴퓨터교육과
학사과정

관심분야 : 데이터과학, 인공지능,
지식추적

배 지 훈 (Ji-Hoon Bae)



2000년 2월 : 경북대학교

전자·전기공학부(공학사)

2002년 2월 : 포항공과대학교

전자컴퓨터공학부(공학석사)

2016년 2월 : 포항공과대학교

전자·전기공학과(공학박사)

2002년 1월 ~ 2019년 8월 :

한국전자통신연구원 책임연구원

2019년 9월 ~ 2024년 8월 : 대구가톨릭대학교

AI빅데이터공학과 조교수

2021년 3월 ~ 2024년 8월 : 대구가톨릭대학교

SW중심대학사업단 SW기초교육센터장

2024년 9월 ~ 현재 : 한국교원대학교 컴퓨터교육과

부교수

관심분야 : 임베디드 AI, 전이학습, 지식추적, 딥러닝,
컴퓨터 교육