Towards LLM-Empowered Knowledge Tracing via LLM-Student Hierarchical Behavior Alignment in Hyperbolic Space

Xingcheng Fu[1,2], Shengpeng Wang[1,2], Yisen Gao[2,3], Xianxian Li[1,2], Chunpei Li[1,2], Qingyun Sun[2,4], Dongran Yu[1,2]

[1]Key Lab of Education Blockchain and Intelligent Technology, Guangxi Normal University, Guilin, China
[2]Guangxi Key Lab of Multi-source Information Mining & Security, Guangxi Normal University, Guilin, China
[3] Computer Science and Engineering , The Hong Kong University of Science and Technology, Hong Kong, China
[4]School of Computer Science and Engineering, Beihang University, Beijing, China
{fuxc, lixx, yudran, SPalmW}@gxnu.edu.cn, sunqy@buaa.edu.cn

Abstract

Knowledge Tracing (KT) diagnoses students' concept mastery through continuous learning state monitoring in education. Existing methods primarily focus on studying behavioral sequences based on ID or textual information. While existing methods rely on ID-based sequences or shallow textual features, they often fail to capture (1) the hierarchical evolution of cognitive states and (2) individualized problem difficulty perception due to limited semantic modeling. Therefore, this paper proposes a **L**arge Language Model **H**yperbolic **A**ligned **K**nowledge **T**racing(L-HAKT). First, the teacher agent deeply parses question semantics and explicitly constructs hierarchical dependencies of knowledge points; the student agent simulates learning behaviors to generate synthetic data. Then, contrastive learning is performed between synthetic and real data in hyperbolic space to reduce distribution differences in key features such as question difficulty and forgetting patterns. Finally, by optimizing hyperbolic curvature, we explicitly model the tree-like hierarchical structure of knowledge points, precisely characterizing differences in learning curve morphology for knowledge points at different levels. Extensive experiments on four real-world educational datasets validate the effectiveness of our Large Language Model Hyperbolic Aligned Knowledge Tracing (L-HAKT) framework.

Introduction

Knowledge Tracing (KT) (Abdelrahman, Wang, and Nunes 2023)is a key technique in educational intelligence for tracking student knowledge states based on historical behaviors. It enables personalized instruction by assessing mastery levels (Yin et al. 2023; Jones and Jo 2004). Existing methods are mainly divided into two categories: Sequence-centric approaches model student interactions as temporal processes (Pandey and Karypis 2019; Piech et al. 2015a), capturing learning dynamics through patterns in response sequences. Graph-centric approaches, in contrast, encode structural relationships among concepts, leveraging graph neural networks to model dependencies.

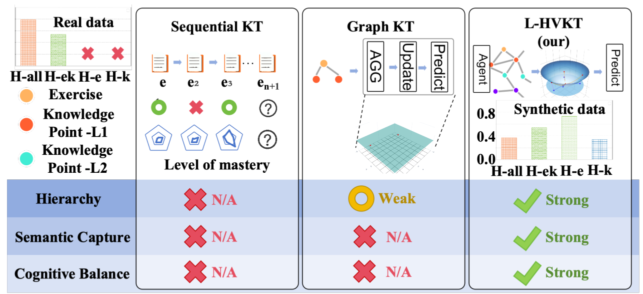


Figure 1: The bar chart labels H-all, H-ek, H-e, and H-k denote hyperbolicity measures for student-question-knowledge, question-knowledge, question-question, and knowledge-knowledge relationships, respectively. Lower values indicate stronger structural alignment with tree-like or hyperbolic space characteristics.

However, the existing methods rely on simple structured information and fail to fully exploit the rich semantics of questions (Mislove et al. 2007)[1], making it difficult to capture the hierarchical dynamics of students' cognitive states. As shown in Figure 1: (1) Traditional methods model in Euclidean space (Nickel and Kiela 2017), where the flat geometric structure cannot express the tree-like hierarchical properties of knowledge systems. (2) The implicit topological relationships between knowledge concepts in the problem semantics have not been effectively captured (Piech et al. 2015b), resulting in the connections between the questions of related concepts not being well utilized. (3) The degree of individual mastery distorts the difficulty of the practice (Lee et al. 2022; Chen et al. 2024). For instance, models trained on data from low-performing regions might wrongly classify moderately complex problems as high-difficulty ones, while high-performing students might label most problems as simple ones, which exposes the flaws of traditional difficulty assessment.

To bridge this gap, we attempt to utilize LLMs to extract knowledge point concepts (Tan et al. 2021; Tong, Zhou, and Wang 2020) from the questions to construct a knowledge

[1]We interchangeably use the terms *question* and *exercise* in this paper.

graph, thereby modeling the potential correlations among different questions. It also enables LLMs to simulate human learning behaviors (Gao et al. 2025; Piech et al. 2015a; Li et al. 2024), provide objective difficulty information of the questions, and construct a more accurate dynamic learning cognitive process.

Despite this, there are still challenges in accurately simulating students' hierarchical cognition from simple to difficult:(1) Absence of hierarchical concept representation: The implicit hierarchical associations among knowledge points need to be modeled. For instance, the basic definition will gradually evolve into different inferences. Therefore, it is necessary to establish the hierarchical relationship of knowledge points to depict the learning trajectory of higher-order knowledge (Fu et al. 2023; Song et al. 2021). (2) Absence of hierarchical dynamic cognition: Although large models can simulate human behavior to assess the difficulty of questions, there are still different levels of understanding biases among individual students regarding the questions (Park et al. 2023). Therefore, it is necessary to align the general cognition of the difficulty level of agents with the dynamic understanding cognition of individual students.

In light of the above analysis, we propose a Large Language Model hyperbolic-aligned knowledge tracing method named L-HAKT. This method first utilizes the Teacher Agent to parse the semantic meaning of question texts, explicitly excavating the contained knowledge point concepts and their hierarchical dependencies, thereby resolving the absence of hierarchical representation for knowledge points. Through the Student Agent simulating individual learning behaviors, it generates synthetic data to supplement the cognitive process information missing in traditional data. Then, we design a hyperbolic contrastive alignment mechanism:

Conducting contrastive learning between synthetic data and real student behavioral data in hyperbolic space to reduce distributional differences in key features such as question difficulty and forgetting patterns; Finally, optimizing hyperbolic curvatures explicitly models the inherent tree-like hierarchy of knowledge points. This calibrates hierarchical-aware alignment between universal question difficulty perception and individual students' dynamic cognitive understanding, effectively resolving hierarchical dynamic cognitive deviation. Consequently, the model precisely characterizes learning curve differences across knowledge levels and leverages hyperbolic space's hierarchical propagation to sensitively capture mastery leaps in higher-order knowledge points. The main contributions are as follows:

- Our work is the first to propose an LLM dual-agent teacher-student collaborative framework, addressing the modeling blind spot of "student thinking paths" in traditional data.
- We innovatively introduce hyperbolic geometric space, achieving explicit modeling of knowledge point hierarchies through curvature optimization, and precisely characterizing learning curve morphology differences across knowledge levels.
- Based on aligned hierarchical data, we design a hyperbolic knowledge state tracker that leverages hierarchical propagation properties to sensitively capture mastery degree in higher-order knowledge points.

Experimental results demonstrate that the proposed L-HAKT possesses superior effectiveness and indeed enhances the capability of KT models in practical inference.

Related Works

Knowlege Tracing

Knowledge Tracing (KT) dynamically infers a student's knowledge state evolution by analyzing historical interactions and test data to predict future performance. Existing approaches bifurcate into two paradigms: (1) Sequence Modeling-based KT: Representative works include Deep Knowledge Tracing (such as DKT (Piech et al. 2015a), ATKT (Guo et al. 2021), DIMKT (Shen et al. 2022), etc.) which pioneered the use of RNNs; the Dynamic Key-Value Memory Network (such as DKVMN (Shen et al. 2022), SKVMN (Zhang et al. 2017), DGMN (Abdelrahman and Wang 2019)), which explicitly updates knowledge point states via an external memory matrix; and Transformer architecture models (such as SAKT (Pandey and Karypis 2019) and SAINT (Choi et al. 2020a) that utilize self-attention mechanisms to capture long-range dependencies. (2) Graph-based KT (Tong et al. 2020; He et al. 2025): Typical works include Graph-based Knowledge Tracing(GKT (Nakagawa, Iwasawa, and Matsuo 2019)), which explicitly constructs knowledge point adjacency graphs, and Structured Knowledge Tracing(SKT (Tong et al. 2020)), which integrates dependency relationships from knowledge graphs. These two categories of methods advance the refinement of knowledge state representation from the perspectives of temporal dynamics and structural relevance, respectively.

Hyperbolic Machine Learning

Hyperbolic geometric space excels in graph representation learning due to its hierarchical representation capability (Fu et al. 2024; Nickel and Kiela 2018; Cheng et al. 2025b), making it ideal for tree-like structures such as knowledge graphs and user-item interaction graphs (Sun et al. 2023). It has been successfully applied in recommender systems, for example with the LKGR model (Chen et al. 2022) capturing user-item hierarchies in Lorentz hyperbolic space. Significant progress has also occurred in knowledge graph embedding (Sun et al. 2025b,a), including models like HHNE++ (Wang, Zhang, and Shi 2019) that achieve low-distortion hierarchical representation of heterogeneous information networks (Cheng et al. 2025a; He et al. 2024; Sun et al. 2024b,c). For recommender systems, hyperbolic space primarily optimizes user-item interactions (Zhang et al. 2019; Wang et al. 2021), while for knowledge graphs it focuses on hierarchical dependencies between entities (Perozzi, Al-Rfou, and Skiena 2014). However, hyperbolic graph learning remains largely unexplored for intelligent education and knowledge tracing, particularly for hierarchical modeling of dynamic student-knowledge point interactions.

Preliminary

Hyperbolic Geometry

Hyperbolic space commonly refers to manifolds with constant negative curvature and is used for modeling complex networks (Gao et al. 2022). Among the common isometric models used to describe hyperbolic spaces, the hyperboloid model has recently been widely used in machine learning. The hyperboloid model is an n-dimensional hyperbolic geometry as a manifold in the $(n+1)$-dimensional Minkowski space. A hyperboloid manifold $\mathbb{H}_\kappa^n = \{x \in \mathbb{R}^{n+1} | \langle x, x \rangle_\mathbb{H} = 1/\kappa, \kappa < 0\}$ in n-dimensional space with curvature κ. $\langle \cdot, \cdot \rangle_\mathbb{H}$ is Lorentzian scalar product and $\langle x, y \rangle_\mathbb{H} := -x_0 y_0 + x_1 y_1 + \cdots + x_n y_n$. The distance $d_\kappa^\mathbb{H}(x, y)$ on the hyperboloid model is defined as:

$$d_\kappa^\mathbb{H}(x, y) = \frac{1}{\sqrt{|\kappa|}} \operatorname{arccosh}(|\kappa| \langle x, y \rangle_\mathbb{H}). \tag{1}$$

To ensure compatibility with standard machine learning methods, the logarithmic map is employed to project data into a local Euclidean tangent space ($\mathcal{T}_x \mathbb{H}_\kappa^n$) for computation (Song et al. 2021), and the exponential map is used to project the results back. The logarithmic map $\log_x^\kappa(\cdot)$ and exponential map $\exp_x^\kappa(\cdot)$ are defined as:

$$\log_x^\kappa(y) = d_\kappa^\mathbb{H}(x, y) \frac{y + |\kappa| \langle x, y \rangle_\mathbb{H} x}{\| y + |\kappa| \langle x, y \rangle_\mathbb{H} x \|_\mathbb{H}}, \tag{2}$$

$$\exp_x^\kappa(v) = \cosh\left(\sqrt{|\kappa|} \|v\|_\mathbb{H}\right) x + v \frac{\sinh\left(|\kappa| \|v\|_\mathbb{H}\right)}{\sqrt{|\kappa|} \|v\|_\mathbb{H}}, \tag{3}$$

where $\|v\|_\mathbb{H} = \sqrt{\langle v, v \rangle_\mathbb{H}}$ is the Lorentzian norm of v.

Problem Statement

Student Interaction Sequence: For each student S, we observe a chronologically ordered interaction sequence, denoted as $\mathcal{S} = \{s_1, s_2, \ldots, s_m\}$. Interaction Content: Each interaction s_j is a structured quadruple $s_j =< q_j, \{c\}, r_j, t_j >$, where: q_j represents the question the student solved in the j-th interaction. $\{c\}$ is the set of knowledge components associated with the question $R_j \in \{0, 1\}$ is a binary response label. $R_j = 1$ indicates that student S answered question q_j correctly in the j-th interaction, while $R_j = 0$ indicates an incorrect answer. t_j represents the timestamp when the $j-th$ interaction occurred, recording the temporal information of the interaction. Knowledge State and Embedding Space:k_t^c represents the latent knowledge state of student S for knowledge concept c at timestep t.

Method

In this section, we propose our L-HAKT framework, aiming to assist knowledge tracing tasks through data generated by large models. It consists of three parts. (1) We deploy two LLM-driven Agents, Teacher Agent: Constructs hierarchical knowledge graphs by parsing exercise semantics and explicit knowledge dependencies. Student Agent:Simulates personalized learning behaviors (e.g., engagement and forgetting) to generate synthetic interaction data. (2) We introduce a relation-aware hyperbolic graph neural network , capturing hierarchical dependencies within knowledge systems by modeling hyperbolic embeddings with curvature optimization. (3) We perform distribution calibration between Teacher-generated data and real-world student interactions via a hyperbolic contrastive alignment mechanism, integrating hierarchical states for knowledge tracing.

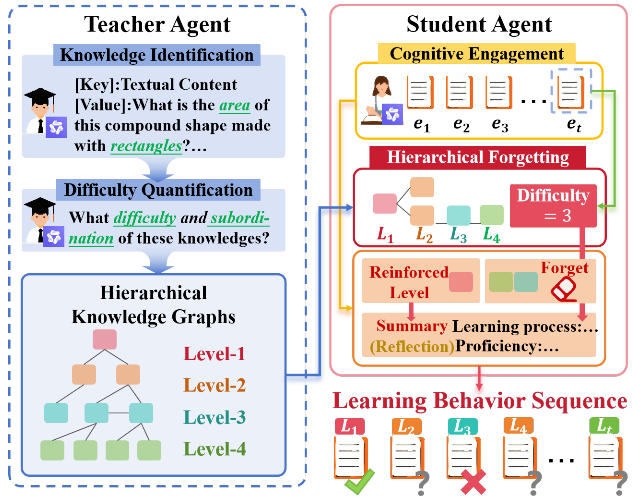


Figure 2: Illustration of LLM-based Teacher-student Behavior Modeling and Data Augmentation.

LLM-Based Teacher-Student Behavior Augmentation

Here, we employ two Agents: (1) Teacher Agent: it extracts rich information from texts, constructing knowledge graphs. (2) Student Agent: it captures learners' engagement states and behaviors to simulate their responses.

Teacher Agent. The Teacher Agent M_t processes problem images q^{image} using specialized prompt templates to achieve three core outputs:

- *Hierarchical Knowledge Identification*: It parses images into semantic text $q_i^{text} = M_t(q_i^{\text{image}})$ then extracts and classifies knowledge points into explicit hierarchy levels $L_j \in \{1, 2, 3, 4\}$ (from basic definitions to comprehensive reasoning);
- *Structured Knowledge Graph Construction:* It builds a fine-grained hierarchical knowledge graph by establishing parent-child dependencies between knowledge points based on their levels, creating a tree-like pedagogical structure;
- *Exercise Difficulty Quantification:* It assigns objective difficulty scores to each exercise by analyzing the hierarchy levels of its associated knowledge points $c_{i_1}^L, c_{i_2}^L, \cdots, c_{i_k}^L$, where higher-level knowledge combinations yield higher difficulty values.

This process provides: 1) Explicit knowledge hierarchy mapping, 2) Pedagogically structured knowledge graphs, and 3) Semantically-grounded exercise difficulty metrics for downstream components.

Student Agent. Utilizes the LLM's knowledge base and reasoning capabilities to simulate the problem-solving cognitive process of a real student. Given a student's historical interaction sequence $l_{j-1} = \{s_1, s_2, \ldots, s_{j-1}\}$, where $s_i = \langle q_i, C_i, R_i \rangle$ contains the question q_i, associated knowledge points C_i, and response result $R_i \in \{0, 1\}$,$R_i = 1$ indicates correct,$R_i = 0$ indicates error . To simulate personalized learning behaviors, our Student Agent incorporates two specialized modules: The Cognitive Engagement Module dynamically assesses the student's concentration level based on current exercise difficulty and mastery of related knowledge points. The Hierarchical Forgetting Module models differential memory decay patterns according to knowledge point difficulty levels.

- *Cognitive engagement module*: A student's concentration during learning is influenced by problem difficulty q_j and mastery of related knowledge points. The learner's current engagement is defined as $\Gamma_j = \sigma(W_q \cdot [X_{q_i}; X_{c_{ij}}; t] + b_q)$. If the interval t for such problems is short, it indicates active learning. The current cognitive engagement state is evaluated based on embeddings of the current problem X_{q_i} and difficulty of related knowledge points $X_{c_{ij}}$.
- *Hierarchical forgetting module*: We divide memory types into three levels based on problem difficulty. Basic knowledge points $L \in \{1, 2\}$ cover core concepts, basic formulas, and simple rules. Intermediate knowledge points $L \in \{3\}$ involve typical problem-solving methods, medium-complexity strategies, and combined applications. The forgetting curve has moderate slope. Difficult knowledge $L \in \{4\}$ includes complex problem decomposition. The forgetting curve is steepest with rapid decay. We model forgetting with $F_j = \exp(-\lambda \cdot L_{\text{avg}} \cdot t_j)$, where L_{avg} is the average level of designed knowledge points.

Using the student's current engagement state and forgetting degree information for the current problem, we update the current knowledge state via LSTM (Graves 2012).

$$\mathbf{h}_j^s = \mathbf{LSTM}([X_{q_j}; \sum_{c \in C_{ij}} w_c X_c] \oplus (\Gamma_j \odot F_j \odot h_{j-1}^s)) \quad (4)$$

where $\mathbf{X}_{q_j}, \mathbf{X}_c$ are the question and knowledge point embeddings, respectively,w_c is the weight of the knowledge point.

Based on the current exercise e_j and historical state h_{j-1}^s, the student Agent predicts the student's response $A_j \in 0, 1$ and constructs/outputs the underlying reasoning path $\mathcal{P}_j = [c^{(1)}, \cdots, c^{(k)} \to A_j]$, where A_j is obtained through analyzing relevant knowledge points c_j of the problem.

Hyperbolic Encoding & Alignment

Since knowledge points are not isolated, we first construct a heterogeneous graph: $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, Then, our connecting exercise-knowledge point edges, and establishing dependency edges between knowledge points based on the knowledge graph. $\mathcal{E}_{\text{hie}} = \{(c_i, c_j) | L_{c_i} < L_{c_j}\}$ constructs directed hierarchical dependency chains (e.g., limit(level 2) $\to$ derivative(level 2) $\to$ differential equations(level 4)).

Relational-aware Hyperbolic Graph Neural Network To model the complex hierarchical dependencies in both real κ_{real} and synthetic κ_{syn} graphs within hyperbolic space, we extend multi-head graph attention with curvature-specific processing. For each graph with its own curvature $\kappa \in \{\kappa_{\text{real}}, \kappa_{\text{syn}}\}$, thereby defining two hyperbolic spaces $\mathbb{H}^{\kappa_1}$ and $\mathbb{H}^{\kappa_2}$ for real and synthetic graphs respectively, we project the initial Euclidean node embeddings $\mathbf{X}^{\mathbb{E}}$ into hyperbolic manifolds via the exponential map, yielding hyperbolic representations $\mathbf{H}_i^{\mathbb{H}_\kappa} = \exp_{\mathbf{0}}^{\kappa}(\mathbf{X}^{\mathbb{E}})$. Finally, we perform relation-aware hyperbolic aggregation. Specifically, the representation for each node i at layer L is updated as follows:

$$\mathbf{h}_i^{(L+1)} = \exp_{\mathbf{0}}^{\kappa}\left(\sigma\left(\sum_{j \in \mathcal{N}i} \alpha_{ij}^{(L)} \cdot \mathbf{W}^{(L)} \log_{\mathbf{0}}^{\kappa}\left(\mathbf{h}_j^{\mathbb{H}_\kappa^{(L)}}\right)\right)\right) \quad (5)$$

where $\alpha_{ij}^{(L)}$ is the attention score at layer L with , and σ is an activation function.

Hyperbolic Constractive Alignment We use Hyperbolic Graph Neural Network (HGNN) (Liu, Nickel, and Kiela 2019) to process $\mathcal{G}$, capturing the hierarchical nature of real and generated data. This embeds foundational knowledge points in the flatter central region, while distributing higher-order knowledge points in the higher-curvature peripheral regions, obtaining the L-th layer embeddings $h_v^{(L)}$,$h_u^{(L)}$, and apply contrastive learning to enhance their representational capability. For the exercise-concept side, exercises and knowledge concepts shared by both embedding spaces are labeled as positive samples, while other entities are treated as negative samples. Based on these positive and negative samples, the following contrastive loss is adopted:

$$\mathcal{L}_{\text{con}} = -\sum_{(u,v) \in \mathcal{P}} \log \frac{pos(h_u^{(L)}, h_v^{(L)})}{pos(h_u^{(L)}, h_v^{(L)}) + neg(h_u)}, \quad (6)$$

where $\mathcal{P}$ is positive pairs, $sim(h_u^{(L)}, h_v^{(L)})$ denotes the cosine similarity between embeddings of $h_u^{(L)}$ and $h_v^{(L)}$, τ is a temperature coefficient.

Hyperbolic Knowledge Tracing

Sequence Propagation We obtain hyperbolic embeddings for exercises and associated knowledge points $X_{qc}^{\mathbb{H}}$ from our relation-aware hyperbolic graph neural network. To incorporate the correctness rate R_t into these embeddings while maintaining computational efficiency, we first project $X_{qc}^{\mathbb{H}}$to the tangent space at the origin via the logarithmic map $X_{qc}^{\mathbb{E}} = \log_0^{\kappa}(X_{qc}^{\mathbb{H}})$, then perform the update $X_i = \phi(X_{qc}^{\mathbb{H}} \oplus R_t)$in Euclidean space. For sequential processing, we similarly map the previous hyperbolic hidden state $H_{t-1}^{\mathbb{H}}$ to its Euclidean counterpart $H_{t-1}^{\mathbb{E}} = \log_0^{\kappa}(H_{t-1}^{\mathbb{H}})$.We then construct joint representations by integrating question difficulty q^{diff} and the average knowledge point level $\overline{c^L}$,applying hierarchical differentiation through $D^t = \phi(W_d \cdot q^{\text{diff}} + W_L \cdot \overline{c^L})$ where higher-level problems receive reinforced processing

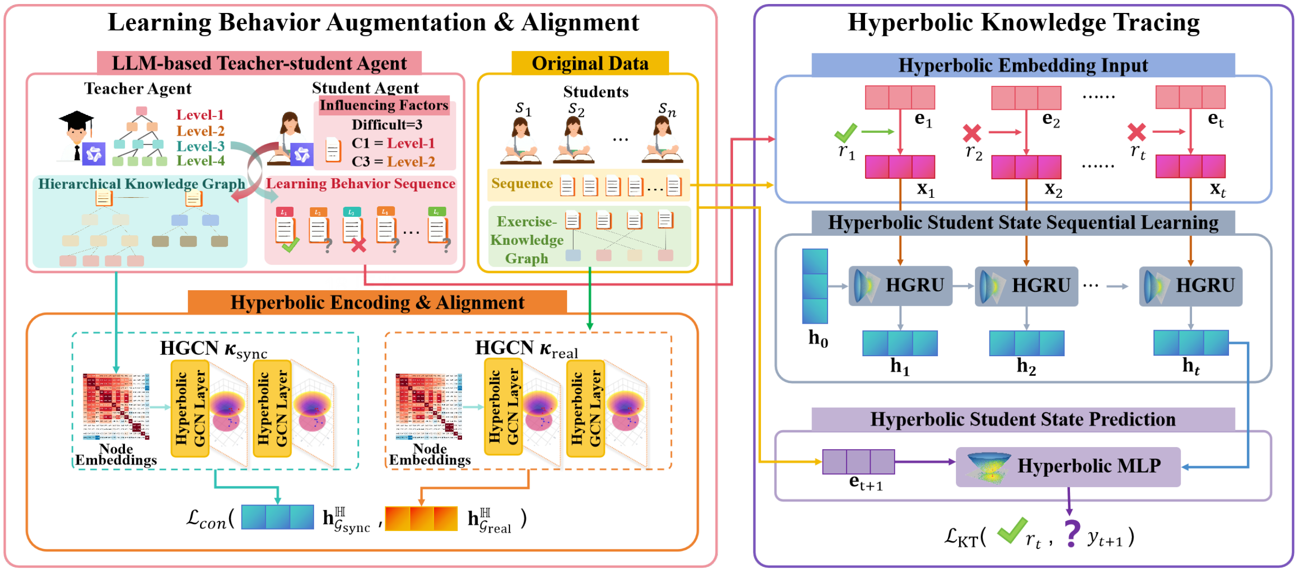


Figure 3: Illustration of the architecture of L-HAKT.

and basic problems undergo standard handling. This dual-space approach leverages hyperbolic geometry's hierarchical representation capabilities while maintaining computational tractability through tangent space operations (Yang et al. 2022).

$$H_t^{\mathbb{E}} = HGRU(X_t^{\mathbb{E}}, H_{t-1}^{\mathbb{E}}, D_t) \tag{7}$$

As the GRU is built upon the tangent space, logarithmic maps are needed . Then, we feed the states into the GRU and map the hidden state back to hyperbolic space $H_t^{\mathbb{H}} = exp_0^{\kappa}(H_t^{\mathbb{E}})$. As we can see, the final $H^{\mathbb{H}}$ is capable of integrating the knowledge point hierarchy structure, content, and question-solving information.

Model Training The KT loss $\mathcal{L}_{kt}$ is defined as the binary cross-entropy loss between the prediction y_t and the correct answer $\hat{y}$, calculated as follows:

$$\mathcal{L}_{\text{KT}} = -\sum_{i=1} (y_t log\hat{y} + (1-y_t) \, log \, (1-\hat{y})) \tag{8}$$

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{KT}} + \alpha \mathcal{L}_{\text{con}} \tag{9}$$

where α is the hyperparameter controlling $\mathcal{L}_{con}$.

Experiment

Dataset Description

We conducted experiments on four public education datasets: ASSIST2009 (Feng, Heffernan, and Koedinger 2009), ASSIST2012 (Feng, Heffernan, and Koedinger 2009), EdNet (Choi et al. 2020b), and Eedi(Wang et al. 2020). These datasets comprehensively record student-question-knowledge point interaction triples, with Eedi additionally containing image information corresponding to questions. To address Eedi's multimodal characteristics, we utilized a vision-Large Language Model (e.g., Qwen-2.5VL) to parse question image content and simulate student problem-solving processes: first converting images to textual descriptions, then combining question text prompts to simulate answer behaviors of students at different cognitive levels via LLM. Detailed dataset statistics are shown in Table 2, which specifically annotates the volume of simulated interaction data newly added to Eedi after LLM enhancement.

Experimental Setup

Similar to ((Sun et al. 2024a; Liu et al. 2023)), we randomly partitioned the dataset into two subsets, with 80% used for training and 20% used for testing. We evaluated model performance using Accuracy (Acc) and AUC. Accuracy measures classification correctness, while AUC assesses the model's ability to distinguish between different classes. All models were trained and tested on a single Nvidia A100 40GB GPU. The vector dimension d was set to match the number of knowledge concepts. The representation learning module consists of 2 layers. A hyperparameter search was conducted for α over the range $[0.0001, 0.001, 0.01, 0.1, 1]$.

Baseline Approaches

To validate the effectiveness of our L-HAKT, we compare it against various baseline methods from two categories.

Sequence Models: This category encompasses temporal knowledge tracing approaches including recurrent-based methods like DKT (Piech et al. 2015a) using LSTM for knowledge state tracking alongside its enhanced variants DKT+(Yeung and Yeung 2018) addressing inconsistent states, DKT+forgetting(Nagatani et al. 2019) incorporating forgetting behavior. Attention mechanisms are represented by ATKT (Guo et al. 2021) employing adversarial training and AT-DKT utilizing auxiliary tasks.DIMKT (Shen et al. 2022) examines the impact of problem difficulty on students' knowledge levels. Memory network approaches feature DKVMN (Zhang et al.

Model	ASSIST09		ASSIST12		EdNet		Eedi	
	AUC(%)	ACC(%)	AUC(%)	ACC(%)	AUC(%)	ACC(%)	AUC(%)	ACC(%)
DKT	75.97±0.25	73.01±0.21	72.90±0.07	73.27±0.16	70.10±0.37	71.11±0.42	76.01±0.03	71.64±3.81
DKT+	77.21±0.03	73.32±0.21	73.64±0.01	73.42±0.01	70.20±0.06	66.21±1.67	74.32±0.01	70.79±0.01
DKT+forget	77.32±0.00	72.13±0.16	73.75±0.03	73.12±0.15	70.46±0.03	71.12±0.17	76.13±0.05	71.66±0.14
DIMKT	72.01±0.36	73.68±0.16	76.23±0.22	74.98±0.02	77.00±0.19	72.93±0.02	79.30±0.02	73.03±0.90
DKVMN	76.10±0.11	72.34±0.07	72.12±0.21	73.32±0.23	69.32±0.14	70.95±0.02	76.01±0.29	71.62±0.09
Deep-IRT	76.21±0.11	72.10±2.17	72.32±0.36	73.64±0.00	69.73±0.08	71.22±0.03	76.00±0.37	71.73±0.20
ATKT	76.53±0.90	73.33±0.01	73.24±0.05	72.98±0.02	70.23±0.06	72.10±0.03	76.21±0.22	72.01±0.36
AT-DKT	76.89±0.02	73.05±0.04	72.74±0.07	72.90±0.02	70.11±0.04	71.03±0.21	75.94±0.04	72.00±0.11
AKT	78.23±0.05	74.45±0.30	78.21±0.11	75.23±0.01	76.78±0.08	73.32±0.10	78.84±0.02	73.01±0.16
SAKT	75.33±0.03	71.77±0.08	73.03±0.58	73.40±0.18	69.27±0.05	71.07±0.21	75.73±0.18	71.23±0.09
SAINT	75.00±0.03	71.21±4.85	75.11±0.09	74.01±0.03	76.04±0.24	73.45±0.12	78.32±0.08	72.93±0.14
CL4KT	76.26±0.14	72.75±0.30	72.36±0.29	73.31±3.11	69.65±0.06	71.18±0.01	75.83±0.03	71.47±0.07
GKT	76.32±0.21	72.44±0.14	72.39±0.04	73.29±0.06	69.21±0.14	71.01±0.01	76.02±2.17	71.68±0.05
GIKT	77.33±0.02	72.69±0.24	76.32±0.05	75.11±0.02	76.02±0.21	73.11±0.10	<u>79.68±0.08</u>	73.02±0.20
simpleKT	77.12±0.42	73.56±0.13	77.21±0.07	75.49±0.14	75.11±0.03	73.57±0.03	78.22±0.33	<u>73.12±0.03</u>
MIKT	<u>79.38±0.03</u>	<u>74.54±0.01</u>	<u>78.65±0.03</u>	<u>76.52±0.07</u>	<u>77.10±0.10</u>	<u>74.22±0.30</u>	79.59±0.07	72.67±0.90
our	**80.22**±0.30	**75.10**±0.04	**80.27**±0.40	**77.10**±0.14	**78.23**±0.05	**75.21**±0.01	**80.29**±0.02	**73.49**±0.03

Table 1: Overall AUC and ACC performance of L-HAKT and all baselines. ACC and AUC should be as large as possible, indicating better model performance; **Bold**:best; <u>Underline</u>: runner-up.

	ASSIST09	**ASSIST12**	**EdNet**	**Eedi**
# Students	4160	5000	5000	5000
# Question	15643	36054	11700	26702
# Concept	167	242	1830	1050
# Interaction	206631	713123	1147423	586234

Table 2: Summary of dataset statistics.

2017) with dynamic key-value memory and Deep-IRT integrating item response theory. Transformer architectures include SAKT (Pandey and Karypis 2019) capturing relationships through self-attention, SAINT (Choi et al. 2020a) implementing full Transformer modeling, AKT (Ghosh, Heffernan, and Lan 2020) simulating forgetting through context attention. CL4KT (Lee et al. 2022) addressing interaction sparsity, and DTransformer tracking stable states. simpleKT (Liu et al. 2023) is based on the simplified AKT model structure, achieving simplicity without sacrificing performance.MIKT (Sun et al. 2024a) tracks the students' state of domain knowledge and conceptual knowledge.

Graph Models: This category focuses on explicit knowledge topology modeling through graph structures, featuring GKT (Nakagawa, Iwasawa, and Matsuo 2019) which propagates conceptual knowledge states via graph neural networks alongside GIKT (Yang et al. 2020) that enhances question representations using graph convolutional networks to aggregate question-concept relationships. These methods directly leverage graph connectivity to capture dependencies between knowledge points.

Performance Comparison

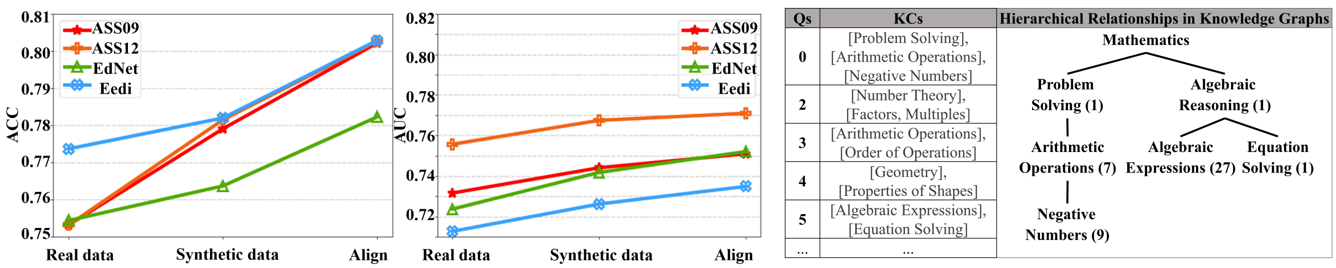
As shown in Table 1, L-HAKT significantly outperforms both sequence models and graph model baselines across four real-world datasets. This advantage verifies that hyperbolic geometric space can effectively capture the latent hierarchical structure among knowledge points, adaptively distinguishing learning dynamics differences between fundamental and advanced knowledge points through hyperbolic curvature. Simultaneously, the data generated by large language models substantially enhances the performance of graph models in knowledge tracing tasks, highlighting the critical role of synthetic data in strengthening model generalization capabilities.

Ablation Study

We conducted ablation studies to evaluate L-HAKT two core mechanisms: hyperbolic representation learning and multi-level contrastive alignment, using Graph-based Knowledge Tracing (GKT) as the backbone network. Experimental results (Table 3) compare three configurations: the full L-HAKT, L-HAKT (w/o con) without contrastive alignment, and L-HAKT (w/o hyp) without hyperbolic representation. Hyperbolic representation learning significantly enhanced hierarchical structure capture, where its adaptive curvature mechanism effectively distinguished learning dynamics between fundamental and advanced knowledge points—particularly crucial for modeling mutation characteristics in complex skills. The performance degradation in L-HAKT (w/o hyp) confirms Euclidean space's limitation in leveraging rich hierarchical relationships. Meanwhile, multi-level contrastive alignment played a vital role in correcting LLM-generated data distribution bias; its removal in L-HAKT (w/o con) demonstrates that synthetic data alone merely provides data augmentation benefits, whereas proper alignment enables authentic distribution calibration and cognitive plausibility enhancement. This module substantially increased sensitivity to mutation states in advanced knowledge points. Collectively, the geometric embedding of knowledge hierarchies in hyperbolic space combined with

Model	ASSIST09				ASSIST12			
	ACC(%)	Δ(%)	AUC(%)	Δ(%)	ACC(%)	Δ(%)	AUC(%)	Δ(%)
GKT	76.32	-	72.44	-	72.39	-	73.29	-
L-HVKT(w/o hyp)	77.55	↑1.61	73.78	↑1.85	78.01	↑7.76	76.84	↑4.84
L-HVKT(w/o con)	76.98	↑0.86	73.21	↑1.06	76.32	↑5.43	75.34	↑2.80
L-HVKT	**80.22**	**↑5.11**	**75.10**	**↑3.67**	**80.27**	**↑10.89**	**77.10**	**↑5.20**
Model	EdNet				Eedi			
	ACC(%)	Δ(%)	AUC(%)	Δ(%)	ACC(%)	Δ(%)	AUC(%)	Δ(%)
GKT	69.21	-	71.01	-	76.02	-	71.68	-
L-HVKT(w/o hyp)	76.54	↑10.59	74.53	↑4.96	79.02	↑3.95	72.48	↑1.12
L-HVKT(w/o con)	75.51	↑9.10	72.67	↑2.34	78.19	↑2.85	71.97	↑0.40
L-HVKT	**78.23**	**↑13.03**	**75.21**	**↑5.91**	**80.29**	**↑5.59**	**73.49**	**↑2.53**

Table 3: The ACC and AUC improvements (%) results of Ablation Study. (**Bold**:best; Underline:runner-up.)



(a) Comparison of the performance with different knowledge graphs.

(b) Validation of knowledge generation of teacher agent.

Figure 4: Validation of Knowledge Graph Effectiveness. (a) Compare the predictive performance of ACC and AUC in four datasets using real data, adding synthetic data, and comparing alignment.(b)Teacher Agent generates a knowledge graph by parsing rich text information.Where (.) denotes the number of related concepts.

precise distribution alignment through contrastive learning enables L-HAKT consistent outperformance of its variants across all four datasets.

Validation of Knowledge Graph Effectiveness

To rigorously validate the effectiveness of our hierarchical knowledge graph, Figure 4.a compares three methodological configurations: the baseline restricted to original question-knowledge point connections, the augmented approach incorporating Student Agent synthetic data, and the full framework integrating Teacher Agent knowledge graphs with hyperbolic contrastive alignment. The results demonstrate that LLM-synthetic data substantially enriches sparse learning trajectories by simulating diverse cognitive pathways. Simultaneously, the contrastive alignment mechanism effectively bridges distributional discrepancies between synthetic and real behavioral patterns while preserving pedagogical hierarchies through geometric constraints in hyperbolic embedding space. Complementing these quantitative insights, Figure 4.b visually substantiates our approach through a representative knowledge graph segment. This topology exhibits pedagogically coherent hierarchical dependencies. Collectively, these outcomes confirm our framework successfully resolves two fundamental limitations of traditional knowledge tracing: the fragmentation of knowledge structures through hierarchical graph construction, and the distortion of difficulty perception via geometrically grounded representation.

Conclusion

We propose **L**arge Language Model **H**yperbolic **A**ligned **K**nowledge **T**racing(L-HAKT), a knowledge tracing framework integrating large language models and hyperbolic geometry to systematically model how hierarchical knowledge structures distinctly influence learning dynamics—progressive mastery of basic concepts versus abrupt evolution of advanced knowledge points. Our dual-agent mechanism employs a Teacher Agent to build logically dependent knowledge graphs and a Student Agent to simulate cognition-driven problem-solving behaviors, generating synthetic data rich in analytical pathways. Through hyperbolic contrastive alignment, we simultaneously calibrate distributions between synthetic and real data while geometrically embedding knowledge hierarchies in curvature-adaptive hyperbolic space. Experiments across four real-world education datasets validate the framework's effectiveness, with hyperbolic embeddings visually demonstrating correlations between knowledge hierarchies and learning curve morphologies. Future work will explore hyperbolic learning generation for instructional optimization and leverage combined model-space for fine-grained diagnostics.

Acknowledgments

Corresponding author is Dongran Yu. This paper is supported by the National Natural Science Foundation of China (No.62462007), The Basic Ability Enhancement Program for Young and Middle-aged Teachers of Guangxi (No.2024KY0073), the Research Fund of Guangxi Key Lab of Multi-source Information Mining & Security (MIMS24-12), and 2024 Guangxi Normal University Joint Training Project of the National Natural Science Foundation (No.2024PY029). We sincerely thank all the authors for their hard work and contributions.

References

Abdelrahman, G.; and Wang, Q. 2019. Knowledge tracing with sequential key-value memory networks. In *Proceedings of the 42nd international ACM SIGIR conference on research and development in information retrieval*, 175–184.

Abdelrahman, G.; Wang, Q.; and Nunes, B. 2023. Knowledge tracing: A survey. *ACM Computing Surveys*, 55(11): 1–37.

Chen, X.; Feng, S.; Yang, M.; Zhao, K.; Xu, R.; Cui, C.; and Chen, M. 2024. Modeling question difficulty for unbiased cognitive diagnosis: A causal perspective. *Knowledge-Based Systems*, 294: 111750.

Chen, Y.; Yang, M.; Zhang, Y.; Zhao, M.; Meng, Z.; Hao, J.; and King, I. 2022. Modeling scale-free graphs with hyperbolic geometry for knowledge-aware recommendation. In *Proceedings of the fifteenth ACM international conference on web search and data mining*, 94–102.

Cheng, J.; Liang, K.; Feng, P.; Liu, W.; Tang, Y.; and He, C. 2025a. Clustering Diffusion Model with Frequency-Signal Modulation for Variational Graph Autoencoders. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1–18.

Cheng, J.; Tang, Y.; He, C.; Feng, P.; Han, K.; and Guan, Q. 2025b. Rethinking Variational Bayes in Community Detection From Graph Signal Perspective. *IEEE Transactions on Knowledge and Data Engineering*, 37(5): 2903–2917.

Choi, Y.; Lee, Y.; Cho, J.; Baek, J.; Kim, B.; Cha, Y.; Shin, D.; Bae, C.; and Heo, J. 2020a. Towards an appropriate query, key, and value computation for knowledge tracing. In *Proceedings of the seventh ACM conference on learning@ scale*, 341–344.

Choi, Y.; Lee, Y.; Shin, D.; Cho, J.; Park, S.; Lee, S.; Baek, J.; Bae, C.; Kim, B.; and Heo, J. 2020b. Ednet: A large-scale hierarchical dataset in education. In *International conference on artificial intelligence in education*, 69–73. Springer.

Feng, M.; Heffernan, N.; and Koedinger, K. 2009. Addressing the assessment challenge with an online system that tutors as it assesses. *User modeling and user-adapted interaction*, 19(3): 243–266.

Fu, X.; Gao, Y.; Wei, Y.; Sun, Q.; Peng, H.; Li, J.; and Li, X. 2024. Hyperbolic geometric latent diffusion model for graph generation. *arXiv preprint arXiv:2405.03188*.

Fu, X.; Wei, Y.; Sun, Q.; Yuan, H.; Wu, J.; Peng, H.; and Li, J. 2023. Hyperbolic geometric graph representation learning for hierarchy-imbalance node classification. In *Proceedings of the ACM Web Conference 2023*, 460–468.

Gao, W.; Liu, Q.; Yue, L.; Yao, F.; Lv, R.; Zhang, Z.; Wang, H.; and Huang, Z. 2025. Agent4edu: Generating learner response data by generative agents for intelligent education systems. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 23923–23932.

Gao, Z.; Wu, Y.; Fan, X.; Harandi, M.; and Jia, Y. 2022. Learning to optimize on Riemannian manifolds. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(5): 5935–5952.

Ghosh, A.; Heffernan, N.; and Lan, A. S. 2020. Context-aware attentive knowledge tracing. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, 2330–2339.

Graves, A. 2012. Long short-term memory. *Supervised sequence labelling with recurrent neural networks*, 37–45.

Guo, X.; Huang, Z.; Gao, J.; Shang, M.; Shu, M.; and Sun, J. 2021. Enhancing knowledge tracing via adversarial training. In *Proceedings of the 29th ACM international conference on multimedia*, 367–375.

He, C.; Cheng, J.; Chen, G.; Guan, Q.; Fei, X.; and Tang, Y. 2024. Detecting communities with multiple topics in attributed networks via self-supervised adaptive graph convolutional network. *Information Fusion*, 105: 102254.

He, C.; Luo, J.; Tang, Y.; Chen, G.; and Guan, Q. 2025. Graph Neural Network Empowers Intelligent Education: A Systematic Review from An Application Perspective. *IEEE Transactions on Learning Technologies*, 1–18.

Jones, V.; and Jo, J. H. 2004. Ubiquitous learning environment: An adaptive teaching system using ubiquitous technology. In *ASCILITE 2004*. Australiasian Society for Computers in Learning in Tertiary Education.

Lee, W.; Chun, J.; Lee, Y.; Park, K.; and Park, S. 2022. Contrastive learning for knowledge tracing. In *Proceedings of the ACM web conference 2022*, 2330–2338.

Li, H.; Xu, T.; Zhang, C.; Chen, E.; Liang, J.; Fan, X.; Li, H.; Tang, J.; and Wen, Q. 2024. Bringing generative AI to adaptive learning in education. *arXiv preprint arXiv:2402.14601*.

Liu, Q.; Nickel, M.; and Kiela, D. 2019. Hyperbolic graph neural networks. *Advances in neural information processing systems*, 32.

Liu, Z.; Liu, Q.; Chen, J.; Huang, S.; and Luo, W. 2023. simpleKT: a simple but tough-to-beat baseline for knowledge tracing. *arXiv preprint arXiv:2302.06881*.

Mislove, A.; Marcon, M.; Gummadi, K. P.; Druschel, P.; and Bhattacharjee, B. 2007. Measurement and analysis of online social networks. In *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*, 29–42.

Nagatani, K.; Zhang, Q.; Sato, M.; Chen, Y.-Y.; Chen, F.; and Ohkuma, T. 2019. Augmenting knowledge tracing by considering forgetting behavior. In *The world wide web conference*, 3101–3107.

Nakagawa, H.; Iwasawa, Y.; and Matsuo, Y. 2019. Graph-based knowledge tracing: modeling student proficiency using graph neural network. In *IEEE/WIC/aCM international conference on web intelligence*, 156–163.

Nickel, M.; and Kiela, D. 2017. Poincaré embeddings for learning hierarchical representations. *Advances in neural information processing systems*, 30.

Nickel, M.; and Kiela, D. 2018. Learning continuous hierarchies in the lorentz model of hyperbolic geometry. In *International conference on machine learning*, 3779–3788. PMLR.

Pandey, S.; and Karypis, G. 2019. A self-attentive model for knowledge tracing. arXiv 2019. *arXiv preprint arXiv:1907.06837*.

Park, J. S.; O'Brien, J.; Cai, C. J.; Morris, M. R.; Liang, P.; and Bernstein, M. S. 2023. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th annual acm symposium on user interface software and technology*, 1–22.

Perozzi, B.; Al-Rfou, R.; and Skiena, S. 2014. Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, 701–710.

Piech, C.; Bassen, J.; Huang, J.; Ganguli, S.; Sahami, M.; Guibas, L. J.; and Sohl-Dickstein, J. 2015a. Deep knowledge tracing. *Advances in neural information processing systems*, 28.

Piech, C.; Spencer, J.; Huang, J.; Ganguli, S.; Sahami, M.; Guibas, L.; and Sohl-Dickstein, J. 2015b. Deep Knowledge Tracing. *arXiv e-prints*.

Shen, S.; Huang, Z.; Liu, Q.; Su, Y.; Wang, S.; and Chen, E. 2022. Assessing student's dynamic knowledge state by exploring the question difficulty effect. In *Proceedings of the 45th international ACM SIGIR conference on research and development in information retrieval*, 427–437.

Song, W.; Chen, H.; Liu, X.; Jiang, H.; and Wang, S. 2021. Hyperbolic node embedding for signed networks. *Neurocomputing*, 421: 329–339.

Sun, J.; Yu, F.; Wan, Q.; Li, Q.; Liu, S.; and Shen, X. 2024a. Interpretable knowledge tracing with multiscale state representation. In *Proceedings of the ACM Web Conference 2024*, 3265–3276.

Sun, L.; Huang, Z.; Peng, H.; Wang, Y.; Liu, C.; and Yu, P. S. 2024b. LSEnet: Lorentz Structural Entropy Neural Network for Deep Graph Clustering. In *Proceedings of the 41st International Conference on Machine Learning (ICML)*.

Sun, L.; Huang, Z.; Wan, Q.; Peng, H.; and Yu, P. S. 2024c. Spiking Graph Neural Network on Riemannian Manifolds. In *Advances in the 38th Neural Information Processing Systems (NeurIPS)*.

Sun, L.; Huang, Z.; Wu, H.; Ye, J.; Peng, H.; Yu, Z.; and Yu, P. S. 2023. DeepRicci: Self-supervised Graph Structure-Feature Co-Refinement for Alleviating Over-squashing. In *Proceedings of the IEEE International Conference on Data Mining (ICDM)*, 558–567. IEEE.

Sun, L.; Huang, Z.; Zhang, M.; and Yu, P. S. 2025a. Deeper with Riemannian Geometry: Overcoming Oversmoothing and Oversquashing for Graph Foundation Models. In *Advances in the 39th Neural Information Processing Systems (NeurIPS)*.

Sun, L.; Huang, Z.; Zhou, S.; Wan, Q.; Peng, H.; and Yu, P. S. 2025b. RiemannGFM: Learning a Graph Foundation Model from Riemannian Geometry. In *Proceedings of the ACM on Web Conference 2025 (WWW)*, 1154–1165. ACM.

Tan, W.; Jin, Y.; Liu, M.; and Zhang, H. 2021. BiDKT: Deep knowledge tracing with BERT. In *International Conference on Ad Hoc Networks*, 260–278. Springer.

Tong, H.; Zhou, Y.; and Wang, Z. 2020. Exercise hierarchical feature enhanced knowledge tracing. In *International conference on artificial intelligence in education*, 324–328. Springer.

Tong, S.; Liu, Q.; Huang, W.; Hunag, Z.; Chen, E.; Liu, C.; Ma, H.; and Wang, S. 2020. Structure-based knowledge tracing: An influence propagation view. In *2020 IEEE international conference on data mining (ICDM)*, 541–550. IEEE.

Wang, F.; Liu, Q.; Chen, E.; Huang, Z.; Chen, Y.; Yin, Y.; Huang, Z.; and Wang, S. 2020. Neural cognitive diagnosis for intelligent education systems. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, 6153–6161.

Wang, H.; Lian, D.; Tong, H.; Liu, Q.; Huang, Z.; and Chen, E. 2021. Hypersorec: Exploiting hyperbolic user and item representations with multiple aspects for social-aware recommendation. *ACM Transactions on Information Systems (TOIS)*, 40(2): 1–28.

Wang, X.; Zhang, Y.; and Shi, C. 2019. Hyperbolic heterogeneous information network embedding. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, 5337–5344.

Yang, M.; Zhou, M.; Xiong, H.; and King, I. 2022. Hyperbolic temporal network embedding. *IEEE Transactions on Knowledge and Data Engineering*, 35(11): 11489–11502.

Yang, Y.; Shen, J.; Qu, Y.; Liu, Y.; Wang, K.; Zhu, Y.; Zhang, W.; and Yu, Y. 2020. GIKT: a graph-based interaction model for knowledge tracing. In *Joint European conference on machine learning and knowledge discovery in databases*, 299–315. Springer.

Yeung, C.-K.; and Yeung, D.-Y. 2018. Addressing two problems in deep knowledge tracing via prediction-consistent regularization. In *Proceedings of the fifth annual ACM conference on learning at scale*, 1–10.

Yin, Y.; Dai, L.; Huang, Z.; Shen, S.; Wang, F.; Liu, Q.; Chen, E.; and Li, X. 2023. Tracing knowledge instead of patterns: Stable knowledge tracing with diagnostic transformer. In *Proceedings of the ACM web conference 2023*, 855–864.

Zhang, J.; Shi, X.; King, I.; and Yeung, D.-Y. 2017. Dynamic key-value memory networks for knowledge tracing. In *Proceedings of the 26th international conference on World Wide Web*, 765–774.

Zhang, J.; Shi, X.; Zhao, S.; and King, I. 2019. Star-gcn: Stacked and reconstructed graph convolutional networks for recommender systems. *arXiv preprint arXiv:1905.13129*.