RESEARCH-ARTICLE

Deep Learning Based Knowledge Tracing: A Review of the Literature

SIYI ZHU, Hubei University, Wuhan, Hubei, China

WENJIE CHEN, Hubei University, Wuhan, Hubei, China

Published: 21 February 2025

ICBDIE 2025: 2025 International Conference on Big Data and Informatization Education
February 21 - 23, 2025
Suzhou, China

ICBDIE '25: Proceedings of the 2025 International Conference on Big Data and Informatization Education (February 2025)
https://doi.org/10.1145/3729605.3729620
ISBN: 9798400714405

Deep Learning Based Knowledge Tracing: A Review of the Literature

Siyi Zhu*
School of Computer Science
Hubei University
Wuhan, Hubei, China
SiyiZhu@stu.hubu.edu.cn

Wenjie Chen†
School of Computer Science
Hubei University
Wuhan, Hubei, China
wenjiechen@stu.hubu.edu.cn

Abstract

This study presents new advances in knowledge tracing modeling with Deep Learning. Knowledge Tracing (KT) refers to assessing learners' mastery of knowledge points by analyzing their problem records. Now with deep learning techniques, DLKT models are well equipped to analyze students' complex learning processes. We divided the existing DLKT models into five categories: RNN-based models, attention-based models, GNN-based models, LLM-based models, and other innovative methods. This study compiles more than thirty DLKT models, compares their performance on seven commonly used datasets, and lists the test results for different metrics. We also discuss the main difficulties facing the knowledge tracing field and also predict future trends in this direction.

CCS Concepts

• **Applied computing** → Education; Computer-assisted instruction.

Keywords

Knowledge Tracing, Deep Learning Models, Educational Data Mining, Model Comparison

ACM Reference Format:
Siyi Zhu and Wenjie Chen. 2025. Deep Learning Based Knowledge Tracing: A Review of the Literature. In *2025 International Conference on Big Data and Informatization Education (ICBDIE 2025), February 21–23, 2025, Suzhou, China.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3729605.3729620

1 INTRODUCTION

Knowledge Tracing (KT) refers to the real-time assessment of learners' mastery of knowledge points by analyzing their practice answer data[1]. As shown in Figure 1, this system tracks students' historical answer records (q_1 to q_4) , from which it predicts their correctness in answering subsequent questions (q_5). This dynamic monitoring can show the learner's cognitive development curve, which provides the basis for personalized tutoring by the intelligent teaching system and assists the teacher in teaching, so as to carry out precise teaching interventions[2].

The value of knowledge tracing is that it can accurately predict the pattern of students' knowledge absorption, retention and forgetting, so as to optimize the teaching effect. Classical models in this field, such as the Bayesian Knowledge Tracing (BKT) model[4], estimate the probability of students' knowledge mastery by analyzing their past records of answering questions. Although this kind of probabilistic model lays a theoretical foundation, it does not fully take into account the intricate nonlinear characteristics of student performance. With the increasing size and complexity of educational data, traditional methods are increasingly overstretched in parsing the ever-changing learning behaviors, which has led to the urgent need for a new generation of intelligent modeling.

Deep learning has broken through the traditional knowledge tracing bottleneck in recent years. RNN-based DKT[6] models question-answering trajectories, which reveals the deep patterns of repeated errors in knowledge points by capturing the temporal correlation of learning behaviors. In current solutions, attention mechanism and forgetting curve modeling are proposed to improve prediction accuracy, but real-world scenarios still face challenges such as data noise and high-dimensional features[3], and these technical bottlenecks have yet to be broken.

This overview unpacks the latest advances in Deep Learning for Knowledge Tracing (DLKT). Beginning with why traditional methods like BKT falter in real-world scenarios—akin to using an abacus for big data—we then detail five DLKT model categories in Chapter 3: RNN-based, attention-based, GNN-based, LLM-based, and other innovative architectures. Chapter 4 evaluates model performance across standard KT datasets using literature-reported metrics, while Chapter 5 outlines future directions including multimodal integration and knowledge graphs, culminating in three core contributions to the field:

1. we fully investigate and categorize DLKT models: RNN-based models, attention-based models, GNN-based models, LLM-based models, and other innovative approaches.

2. we analyze the current challenges of KT, such as data sparsity, model interpretability, and generalization to different environments.

3. we propose possible future research directions such as integrating multimodal data, utilizing knowledge graphs, and exploring transfer learning, which we can use to improve students' ability to share knowledge across different educational settings.

In summary, in this paper, we review the progress of DLKT, we emphasize the limitations of traditional models and illustrate the contribution of deep learning techniques, we compare the datasets

*Corresponding author.
†Co-first authors: Siyi Zhu, Wenjie Chen

ICBDIE 2025, Suzhou, China

ACM ISBN 979-8-4007-1440-5/2025/02
https://doi.org/10.1145/3729605.3729620

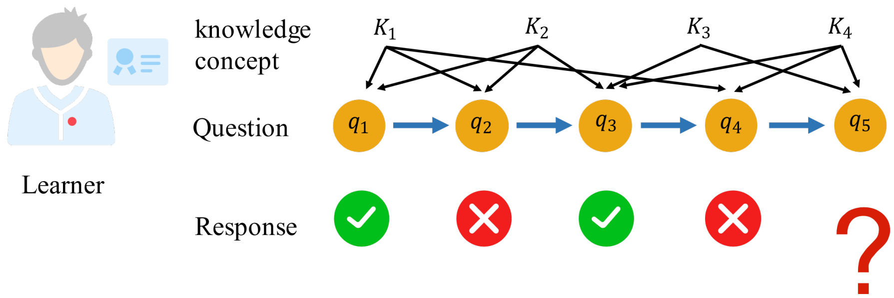

Figure 1: An illustration of knowledge tracing. The student answers a sequence of questions (q_1 to q_4). The task is to predict the student's response to the next question (q_5) based on their interaction history.

and performance of various models, and we propose future research directions for enhanced KT.

2 BACKGROUND

2.1 Traditional Approaches to Knowledge Tracing

Traditionally, Knowledge Tracing (KT) has relied on probabilistic models to estimate student mastery of knowledge. These models, while simple and interpretable, have significant limitations when dealing with complex educational data[1].

Bayesian Knowledge Tracing (BKT)[4], a fundamental knowledge tracing model, employs probabilistic methods to evaluate mastery by analyzing students' past problem-solving performance. Its four parameters include initial knowledge probability, post-mastery error probability, mastery acquisition probability after incorrect attempts, and forgetting probability. While BKT's simplicity and interpretability established it as the standard in early intelligent tutoring systems, its limitations become evident when modeling intricate learning behaviors or non-linear skill progression patterns.

Performance Factor Analysis (PFA)[5] can be seen as an upgraded version of BKT. This approach takes into account variables such as students' historical performance, topic difficulty, and knowledge requirements, and the PFA model can clearly show how learning paths change, but like BKT, it also encounters technical difficulties when dealing with complex correlations between multiple knowledge points.

2.2 Limitations of Traditional Methods

The traditional BKT and PFA models, while giving a foundation for KT, have several old problems. First, these two models are too simple in their assumptions and do not take into account at all the non-linear progression and dynamic laws of change in the learning process. They simply do not handle the complex connections between knowledge points, student characteristics, and practice questions well. When encountering a large amount of data or a complex learning situation, the prediction effect will be greatly reduced. In addition, the structural limitations of the model itself cannot keep up with the continuous changes in student behavior over time. When the model is used in education scenarios, the prediction effect and adaptability will deteriorate when there are many data interferences, incomplete information and multiple dimensions to be processed at the same time.

2.3 The Emergence of Deep Learning in KT

The limitations of traditional KT models have prompted researchers to turn to deep learning.DLKT enhances modeling capabilities by analyzing the temporal patterns of student interactions.The RNN-based architecture automatically extracts features from raw data to simulate real learning behaviors such as varying learning paces, retention of knowledge points, and cycles of forgetting during student learning, thus completely solving KT.

Compared with traditional methods, DLKT models show three major advantages. First, such models use advanced neural network structures such as RNN, LSTM, and recently introduced attention mechanism and Transformer technology, which can accurately capture the dynamic interaction process between students and teaching content. Second, these models perform particularly well when dealing with massive and high-dimensional student data, which is very suitable for the data-intensive nature of modern educational environments. Third, DLKT solves many shortcomings of traditional methods.

Now DLKT is the mainstream method in the knowledge tracing field. This method can give more accurate personalized learning solutions and is more scalable. By analyzing the temporal patterns of students' learning behaviors, the system can provide tailored learning paths. Deep learning's unique advantage of processing time-series data is driving the entire field from probabilistic models to a new data-driven paradigm. It is these technological breakthroughs that allow DLKT not only to predict better, but also to successfully overcome the many shortcomings of traditional models.

3 DEEP LEARNING-BASED KNOWLEDGE TRACING MODELS

3.1 RNN-based Knowledge Tracing Models

Recurrent Neural Networks (RNN) are good at capturing temporal features in learning behaviors, so they have become a common tool in the field of KT. The Deep Knowledge Tracing (DKT)[6] proposed by researchers introduced RNNs into the field for the first time, which is considered as an important breakthrough. The later developed the DKT-Forget[7] improved on this framework by specifically integrating multiple forgotten features. This improvement not only enhances the prediction accuracy of the evolution of knowledge state over time, but also effectively simulates the natural decay process of the memory of knowledge points.

In the research direction of enhancing the interpretability of the model, the Knowledge Query Network (KQN)[8] takes a new path. It encodes learning into knowledge states and skill vectors, using dot products to simulate interactions. Introducing probabilistic skill relatedness quantifies skill connections, boosting prediction accuracy and interpretability.

The ATKT[9] model addresses system stability by integrating disturbances in interaction data through its Attention-LSTM knowledge tracing module. This enables adaptive adjustment of attention focus based on learning contexts, enhancing both prediction accuracy and interference resistance.

The LPKT[10] model uses a more straightforward technical approach. Focusing on the tracking of changes in the learning process, two components, the learning switch and the forgetting switch, were designed. These components monitor changes in students' knowledge level in real time, combining learning effects, practice intervals, and current level to ensure predictive stability while accurately capturing learning dynamics.

The IEKT[11] model has a special design: students' understanding is assessed before doing the problems, and the absorption of knowledge points is checked before updating knowledge. This dual assessment mechanism reflects individual learning differences and improves prediction accuracy.

The DIMKT[12] model focuses on addressing the effect of topic difficulty by directly coding the difficulty parameter into the topic features. This model uses neural networks that can automatically adjust the students' knowledge updating process in real time according to the difficulty level of the questions, so that the progress of knowledge tracing can be more finely tuned.

The AT-DKT[13] model adds two useful features to the traditional framework: topic feature labeling and student base level prediction. These two new modules improve the extraction accuracy of topic features and also effectively capture students' past learning trajectories. This not only improves the prediction accuracy significantly, but also allows historical learning data to play a greater role in predicting future performance.

The QIKT[14] model lies in the topic sensitivity identification mechanism. This design draws on the classic theory of educational measurement, which not only predicts more accurately, but also makes the whole reasoning process as transparent as glass, so that users can see clearly.

The ReKT[15] model innovatively presents the state of knowledge in three dimensions: topic, knowledge point, and knowledge domain. It uses a lightweight FRU framework (Forget-Response-Update) to achieve quality prediction with little computational resources. By optimizing the cooperation of forgetting law and knowledge updating, this design ensures efficiency while not losing prediction accuracy to complex models.

3.2 Attention-based Knowledge Tracing Models

Attention-based KT models are quite popular nowadays, mainly because they can flexibly adjust the learning focus to improve the prediction accuracy. This kind of model can better grasp the continuity of the learning process by analyzing students' historical learning records. However, researchers have found that how to accurately simulate the process of student forgetting is still a challenge. Because the prediction effect will be significantly improved if the forgetting model can be built, so now everyone is focusing on this direction.

The Self-Attentive Knowledge Tracing (SAKT)[16] model first used the self-attention mechanism in this field. Its cleverness lies in its ability to catch the key points from students' past learning data, and it is especially good at dealing with situations with small amounts of data. Doing so not only resulted in more accurate predictions, but also laid the groundwork for later attention models, proving that this mechanism is really good at capturing key learning nodes.

The SAINT[17] model was later upgraded on the structure of the SAKT[16] model by adopting Transformer's bidirectional structure, where practice questions and answers are processed separately, and each is analyzed with a different attention layer. This design digs deeper into the associations between questions and answers, and the fact that Transformer is naturally good at dealing with long-term dependencies makes it the current solution with the top results.

The AKT[18] model plays with a new combination, using the attention mechanism in conjunction with educational measurement theory. Specifically, the attention part is responsible for assigning weights to the history record, while the other part dynamically adjusts the difficulty of the questions and the level of the students. These two tricks work together to improve accuracy and make the results better understood, which is particularly suitable for teaching scenarios that require individualized instruction.

The RKT (Relation-aware Knowledge Tracing)[19] integrates the relation-aware self-attention layer specifically for forgetting problems. This design takes into account the relationship between the practice problem itself and the forgetting law, and by analyzing the characteristics of knowledge decay over time, it breaks through in terms of prediction accuracy and visualization of results. Teachers can directly see the dynamic connection between learning behavior and knowledge acquisition, which is particularly useful.

The FoLiBiKT[20] adds a forgetting decay parameter to the attention mechanism to simulate the process of natural fading of knowledge. By quantifying the effect of forgetting on knowledge tracing, it can more accurately track student level changes. Practical tests show that this method is particularly effective in dealing with long-term learning data, and the prediction effect is significantly improved.

Finally, the sparseKT[21] model, which is innovative in that it engages a filtering mechanism. By filtering irrelevant information and soft thresholding techniques, it can focus on processing key learning events and effectively solve the data noise problem. The design reduces the computational overhead while maintaining the model robustness, and adopts a selective focus mechanism to ensure the generalization ability for the pain point that small-scale datasets are prone to overfitting. Practical applications demonstrate that this focusing strategy on core learning events both optimizes the prediction performance and improves the model operation efficiency.

3.3 GNN-based Knowledge Tracing Models

Researchers have found that graph neural networks (GNNs) are widely used in KT because they can effectively model complex associations between knowledge points and analyze dependencies in the learning process. Unlike traditional sequence models (e.g., RNN), the research team modeled the learning process of the GNN as a graph structure: nodes correspond to knowledge points, and edges denote connections between knowledge points. This graph-based design can more flexibly describe teacher-student interactions and enable more accurate knowledge state tracking. Multiple research teams are currently continuing to explore options for integrating GNN and KT, with each model progressively building on the previous generation's approach.

The GKT (Graph-based Knowledge Tracing)[22] model utilizes Graph Neural Networks (GNN) to reformulate KT as time-series node classification, enhancing accuracy and interpretability without extra data. It models static exercise-concept associations via graph-structured representations.

GIKT (Graph-based Interaction model for Knowledge Tracing)[23] uses Graph Convolutional Networks (GCN) to model exercise-concept-state interactions. Embedded propagation captures long-term dependencies, overcoming traditional methods' limitations in tracking dynamic learning behaviors.

SKT (Structure-based Knowledge Tracing)[24] model integrates both temporal and spatial dimensions for parsing knowledge structure networks. The model achieves a refined modeling of influence propagation between knowledge points through synchronous propagation to deal with undirected associations and stepwise propagation to deal with directed associations, which improves the ability to analyze complex learning patterns.

JKT (Joint Graph Convolutional Network based Deep Knowledge Tracing)[26] model creates a multidimensional modeling framework, which supports personalized learning path design by integrating exercises, concepts, and their associations, comprehensively analyzing interaction patterns, and improving prediction accuracy and interpretability.

DGEKT (Dual Graph Ensemble Learning for Knowledge Tracing)[27] uses hypergraphs and directed graphs to form a dual structure so that connections between different motion concepts can be discovered. The two graphs are combined during online learning, which makes the model predictions more accurate.

SGKT (Session Graph-based Knowledge Tracing)[28] uses session graphs to simulate the student's question answering process, while connecting the session graphs to the relational graphs to make connections between practice questions and knowledge points. This method uses a graph neural network with gating and an attention mechanism to continuously track students' knowledge tracing and help teachers understand the learning progress more clearly.

Bi-CLKT (Bi-Graph Contrastive Learning based Knowledge Tracing)[25] allows the model to learn more discriminative features by comparing the conceptual subgraphs of the practice questions. This dual learning approach captures the specific details of students' answers and knowledge acquisition, allowing for more accurate prediction results.

MVGKT (Multivariate Graph Knowledge Tracing)[29] combines a spatio-temporal graph neural network with a loop unit with gating, which is used to analyze spatio-temporal variations in the learning process. This method also incorporates a topic difficulty analysis function, which not only makes the prediction more accurate, but also solves the difficulties encountered by the old model when dealing with complex relationships.

With the above modeling, we are able to capture richer multi-dimensional interactions, improve prediction accuracy, and better explain the student learning process, progressively enhancing KT.

3.4 LLM-based Knowledge Tracing Models

Large Language Models (LLMs) are evolving particularly fast nowadays, and their ability to understand language and reason about knowledge has directly brought fire to a new approach to knowledge tracing. From 2024 onwards, as LLM technology becomes more and more mature, models that specialize in knowledge tracing with LLMs are starting to appear, bringing new breakthroughs to the field. Compared to previous approaches using Graph Neural Networks (GNN) or Recurrent Neural Networks (RNN), LLM is significantly more powerful in modeling contextual relationships because of its ability to understand semantics and reason better. While there aren't many models of this type yet, the deeper the research goes, the more we can see their great hidden potential. The following describes several LLM-based models, which are variants of the LLM model, that have better predictive accuracy and interpretability in re-education techniques.

DCL4KT[33] used a new trick to combine difficulty prediction with contrastive learning, and also paired it with LLM's difficulty prediction framework. This allows accurate prediction of the difficulty of questions that students have not seen before, which not only makes knowledge tracing more effective, but also more accurate in judging students' answers.

ECKT (Enhanced Code Knowledge Tracing)[32] takes LLM directly and analyzes the code written by students to automatically generate topic descriptions, knowledge points, and difficulty levels. Doing so allows for a more accurate assessment of programming ability as well as helping to plan an exclusive learning path for each student.

LLM-KT[31] has come up with a specially designed decoder that specializes in capturing the long-term correlation between a student's past problem solving and future performance. The prediction accuracy of this LLM-based method rubbed off when it was tested on the large EdNet dataset, which turned out to be really effective.

LKT (Language model-based Knowledge Tracing)[30] takes an off-the-shelf pre-trained language model and puts it to use, utilizing textual information about both topics and knowledge points. Not only are the predictions more accurate this way, even the cold-start problem when new students are just starting out is solved, and by analyzing the attention scores, we can clearly see how the model makes judgments.

SINKT (Structure-aware INductive Knowledge Tracing)[34] had LLM help build a network graph containing multiple relationships to connect knowledge points to topics. This structured approach captures both the connections between knowledge points and the semantics, and performs significantly better than other models in real teaching data tests.

3.5 Other DLKT Models

Although most of the DLKT models use RNN or GNN architectures, there are actually some unusual models that have achieved good results. Although these models are not considered mainstream players, they do have their own mastery in the areas of memory processing, temporal correlation and feature capture.

The SKVMN (Sequential Key-Value Memory Network)[35] mixes recurrent modeling with memory networks to capture hidden associations in the order of questions. Doing so allows for more accurate discovery of intrinsic connections between topics and concepts.

The CKT (Convolutional Knowledge Tracing)[36] has come up with a private, customized version of a learning model that uses a multi-layer convolutional structure to record each student's unique base level and learning speed. This gives more reliable predictions based on each individual's question history.

The HawkesKT[37] goes straight to hardcore temporal analysis, using an event occurrence probability model to quantify the temporal impact between different knowledge points. For example, if you do a related topic when you have just finished learning a certain knowledge point, and if you do it a few months apart, this model can be calculated in a way that is both understandable and fast to train.

The DKVMN (Dynamic Key-Value Memory Network)[38] is even better, which directly plugs in the personality characteristics of students' problem solving habits and learning abilities into the characteristics of the practice problem records. This allows for more accurate prediction of changes in students' knowledge levels and makes knowledge tracing more effective.

4 COMMON DATASETS AND MODEL PERFORMANCE COMPARISON IN KNOWLEDGE TRACING

In this chapter, we focus on a few real datasets that are most commonly used in the field of knowledge tracing (KT), and at the same time compare the actual performance of each model, with the data in its original version. The main reason for picking these datasets is to show you where different models are stronger and weaker in different task scenarios.

4.1 Dataset

Statics2011: This classic physics dataset comes from a college mechanics class and specializes in keeping track of how physics questions are done. They have a special practice - spelling out topic names and step names to form unique topics. In total, more than 190,000 records of problems were saved, covering 333 students and 1,224 problems.

ASSISTments2009: This math question bank grabs data on questions done on the online tutoring platform during the 09-10 school year. It now holds 340,000 records with 4,217 students having done over 26,000 different questions.

ASSISTments2012: This is an upgraded version from the 12-13 school year and is particularly good for predictive analytics. The amount of data is super large, with 2.54 million records of questions done, from over 27,000 students, and the number of questions has risen to over 45,000 questions.

ASSISTments2015: Again, 15 years of data from this platform, with the highlight being the coverage of 100 knowledge points. almost 20,000 students participated in more than 700,000 records, the highest number of participants in the platform.

ASSISTments2017: From 17 years of data mining competitions, with a smaller but more refined question pool. 940,000 records of questions done mainly from 686 students, with a total of 102 questions.

Algebra2005: This algebra dataset is particularly well suited for analyzing step errors, specializing in recording 13-14 year old students as they work through problems. Operating in a similar way to Statics2011, it ended up collecting over 800,000 records from 574 students who did 210,000 questions on 112 topics.

Ednet: The largest publicly available education dataset, specifically designed to test deep learning and knowledge tracing models. Contains 130 million records of questions done, from more than 780,000 students, a volume that currently has no rivals.

4.2 Performance Comparison

After introducing these datasets, we put together a special comparison table (Table 1). The table clearly lists the AUC scores of the five deep knowledge tracing models (DLKT) on different datasets, which is the indicator that best shows the actual performance of the models. These comparative data are all experimental results extracted from published papers, so the side-by-side comparison makes it clear which model is good at handling which type of data and where there are still deficiencies.

5 CONCLUSION

This study provides a comprehensive overview and in-depth analysis of DLKT models. We divided the existing models into five camps: those using recurrent neural networks, those playing with attention mechanisms, those engaging in graph neural networks, those using large language models, and those saber-rattling new approaches. A side-by-side comparison of more than thirty models reveals that - QIKT[14] in the RNN family is the most capable, especially good at capturing the dynamic changes of the knowledge points; AKT[18] in the attention group sits steadily at the top of the list, playing with the traditional methods and the attention mechanism to a new height; MVGKT[29] on the side of graph neural networks is the most colorful, using graph Convolution of the complex relationship between students and knowledge points to understand the door; large language modeling team LKT[30] came out on top, the semantic understanding of the specialties played to the fullest; and those who do not go out of the ordinary model, DKVMN[38] rely on the memory network architecture to kill, the effect is quite capable of playing.

However, if these models are to be applied on the ground, they still have to overcome the hurdles of insufficient data, poor interpretability, and incompatibility in different scenarios. The next research focus is expected to break through in these directions:

Table 1: The AUC results of the five types of DLKT models.

Category	Model	Dataset	AUC
RNN-based Knowledge Tracing Models	DKT 2015[6]	ASSISTments2009	**0.8600**
	DKT-Forget 2019[7]	ASSISTments2012	**0.7309**
	KQN 2019[8]	Statics2011	**0.8320**
	ATKT 2021[9]	Statics2011	**0.8325**
	LPKT 2021[10]	ASSISTments2012	**0.4069**
	IEKT 2021[11]	ASSISTments2009	**0.7720**
	DIMKT 2022[12]	ASSISTments2012	**0.7899**
	AT-DKT 2023[13]	Algebra2005	**0.8246**
	QIKT 2023[14]	Algebra2005	**0.8416**
	reKT 2024[15]	ASSISTments2009	**0.7917**
Attention-based Knowledge Tracing Models	SAKT 2019[16]	ASSISTments2015	**0.8514**
	SAINT 2020[17]	Ednet	**0.7811**
	AKT 2020[18]	ASSISTments2009	**0.8346**
	RKT 2020[19]	ASSISTments2012	**0.7930**
	FoLiBiKT 2023[20]	Algebra2005	**0.7924**
	sparseKT 2023[21]	ASSISTments2015	**0.7501**
GNN-based Knowledge Tracing Models	GKT 2019[22]	ASSISTments2009	**0.7230**
	GIKT 2020[23]	ASSISTments2009	**0.7896**
	SKT 2020[24]	ASSISTments2015	**0.7460**
	Bi-CLKT 2020[25]	ASSISTments2009	**0.8570**
	JKT 2021[26]	Statics2011	**0.8560**
	DGEKT 2022[27]	ASSISTments2017	**0.7754**
	SGKT 2022[28]	ASSISTments2012	**0.8135**
	MVGKT 2023[29]	ASSISTments2017	**0.9470**
LLM-based Knowledge Tracing Models	LKT 2024[30]	XES3G5M	**0.8513**
	LLM-KT 2024[31]	Ednet	**0.7933**
	ECKT 2024[32]	Private Dataset	**0.7653**
	DCL4KT 2024[33]	Algebra2005	**0.8288**
	SINKT 2024[34]	ASSISTments2009	**0.7726**
Other DLKT Models	SKVMN 2019[35]	ASSISTments2009	**0.8363**
	CKT 2023[36]	ASSISTments2015	**0.7350**
	HawkesKT 2021[37]	ASSISTments2012	**0.7676**
	DKVMN 2022[38]	ASSISTments2009	**0.9190**

first, use text, voice, video and other multimodal data to make the model more understanding of the real learning scenarios; second, invite knowledge graphs and other artifacts to help the model clear up the curves between the knowledge points; and third, try these new tricks of transfer learning, so that the model can be used in different teaching environments.

Acknowledgments

This research did not receive any funding.

References

[1] Wang, Y., Zhu, M.X., Yang, S.H., *et al.* Review and Performance Comparison of Deep Knowledge Tracing Models[J]. Journal of Software, 2022, 34(3): 1365-1395.
[2] Abdelrahman G, Wang Q, Nunes B. Knowledge tracing: A survey[J]. ACM Computing Surveys, 2023, 55(11): 1-37.
[3] Shen S, Liu Q, Huang Z, *et al.* A survey of knowledge tracing: Models, variants, and applications[J]. IEEE Transactions on Learning Technologies, 2024.
[4] Corbett A T, Anderson J R. Knowledge tracing: Modeling the acquisition of procedural knowledge[J]. User modeling and user-adapted interaction, 1994, 4: 253-278.
[5] Pavlik P I, Cen H, Koedinger K R. Performance factors analysis–a new alternative to knowledge tracing[C]//Artificial intelligence in education. Ios Press, 2009: 531-538.
[6] Piech C, Bassen J, Huang J, *et al.* Deep knowledge tracing[J]. Advances in neural information processing systems, 2015, 28.
[7] Nagatani K, Zhang Q, Sato M, *et al.* Augmenting knowledge tracing by considering forgetting behavior[C]//The world wide web conference. 2019: 3101-3107.
[8] Lee J, Yeung D Y. Knowledge query network for knowledge tracing: How knowledge interacts with skills[C]//Proceedings of the 9th international conference on learning analytics & knowledge. 2019: 491-500.
[9] Guo X, Huang Z, Gao J, *et al.* Enhancing knowledge tracing via adversarial training[C]//Proceedings of the 29th ACM International Conference on Multimedia. 2021: 367-375.
[10] Shen S, Liu Q, Chen E, *et al.* Learning process-consistent knowledge tracing[C]//Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining. 2021: 1452-1460.
[11] Long T, Liu Y, Shen J, *et al.* Tracing knowledge state with individual cognition and acquisition estimation[C]//Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2021: 173-182.

[12] Shen S, Huang Z, Liu Q, *et al.* Assessing student's dynamic knowledge state by exploring the question difficulty effect[C]//Proceedings of the 45th international ACM SIGIR conference on research and development in information retrieval. 2022: 427-437.
[13] Liu Z, Liu Q, Chen J, *et al.* Enhancing deep knowledge tracing with auxiliary tasks[C]//Proceedings of the ACM Web Conference 2023. 2023: 4178-4187.
[14] Chen J, Liu Z, Huang S, *et al.* Improving interpretability of deep sequential knowledge tracing models with question-centric cognitive representations[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2023, 37(12): 14196-14204.
[15] Shen X, Yu F, Liu Y, *et al.* Revisiting Knowledge Tracing: A Simple and Powerful Model[C]//Proceedings of the 32nd ACM International Conference on Multimedia. 2024: 263-272.
[16] Pandey S, Karypis G. A self-attentive model for knowledge tracing[J]. arXiv preprint arXiv:1907.06837, 2019.
[17] Choi Y, Lee Y, Cho J, *et al.* Towards an appropriate query, key, and value computation for knowledge tracing[C]//Proceedings of the seventh ACM conference on learning@ scale. 2020: 341-344.
[18] Ghosh A, Heffernan N, Lan A S. Context-aware attentive knowledge tracing[C]//Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining. 2020: 2330-2339.
[19] Pandey S, Srivastava J. RKT: relation-aware self-attention for knowledge tracing[C]//Proceedings of the 29th ACM international conference on information & knowledge management. 2020: 1205-1214.
[20] Im Y, Choi E, Kook H, *et al.* Forgetting-aware Linear Bias for Attentive Knowledge Tracing[C]//Proceedings of the 32nd ACM International Conference on Information and Knowledge Management. 2023: 3958-3962.
[21] Huang S, Liu Z, Zhao X, *et al.* Towards robust knowledge tracing models via k-sparse attention[C]//Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2023: 2441-2445.
[22] Nakagawa H, Iwasawa Y, Matsuo Y. Graph-based knowledge tracing: modeling student proficiency using graph neural network[C]//IEEE/WIC/ACM International Conference on Web Intelligence. 2019: 156-163.
[23] Yang Y, Shen J, Qu Y, *et al.* GIKT: a graph-based interaction model for knowledge tracing[C]//Machine learning and knowledge discovery in databases: European conference, ECML PKDD 2020, Ghent, Belgium, September 14–18, 2020, proceedings, part I. Springer International Publishing, 2021: 299-315.
[24] Tong S, Liu Q, Huang W, *et al.* Structure-based knowledge tracing: An influence propagation view[C]//2020 IEEE international conference on data mining (ICDM). IEEE, 2020: 541-550.
[25] Song X, Li J, Lei Q, *et al.* Bi-CLKT: Bi-graph contrastive learning based knowledge tracing[J]. Knowledge-Based Systems, 2022, 241: 108274.
[26] Song X, Li J, Tang Y, *et al.* Jkt: A joint graph convolutional network based deep knowledge tracing[J]. Information Sciences, 2021, 580: 510-523.
[27] Cui C, Yao Y, Zhang C, *et al.* DGEKT: a dual graph ensemble learning method for knowledge tracing[J]. ACM Transactions on Information Systems, 2024, 42(3): 1-24.
[28] Wu Z, Huang L, Huang Q, *et al.* SGKT: Session graph-based knowledge tracing for student performance prediction[J]. Expert Systems with Applications, 2022, 206: 117681.
[29] Xia Z, Dong N, Wu J, *et al.* Multi-Variate Knowledge Tracking Based on Graph Neural Network in ASSISTments[J]. IEEE Transactions on Learning Technologies, 2023.
[30] Lee U, Bae J, Kim D, *et al.* Language Model Can Do Knowledge Tracing: Simple but Effective Method to Integrate Language Model and Knowledge Tracing Task[J]. arXiv preprint arXiv:2406.02893, 2024.
[31] Zhan B, Guo T, Li X, *et al.* Knowledge tracing as language processing: a large-scale autoregressive paradigm[C]//International Conference on Artificial Intelligence in Education. Cham: Springer Nature Switzerland, 2024: 177-191.
[32] Yu Y, Zhou Y, Zhu Y, *et al.* ECKT: Enhancing Code Knowledge Tracing via Large Language Models[C]//Proceedings of the Annual Meeting of the Cognitive Science Society. 2024, 46.
[33] Lee U, Yoon S, Yun J S, *et al.* Difficulty-Focused Contrastive Learning for Knowledge Tracing with a Large Language Model-Based Difficulty Prediction[J]. arXiv preprint arXiv:2312.11890, 2023.
[34] Fu L, Guan H, Du K, *et al.* Sinkt: A structure-aware inductive knowledge tracing model with large language model[C]//Proceedings of the 33rd ACM International Conference on Information and Knowledge Management. 2024: 632-642.
[35] Abdelrahman G, Wang Q. Knowledge tracing with sequential key-value memory networks[C]//Proceedings of the 42nd international ACM SIGIR conference on research and development in information retrieval. 2019: 175-184.
[36] Shen S, Liu Q, Chen E, *et al.* Convolutional knowledge tracing: Modeling individualization in student learning process[C]//Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. 2020: 1857-1860.
[37] Wang C, Ma W, Zhang M, *et al.* Temporal cross-effects in knowledge tracing[C]//Proceedings of the 14th ACM International Conference on Web Search and Data Mining. 2021: 517-525.
[38] Zhang J, Shi X, King I, *et al.* Dynamic key-value memory networks for knowledge tracing[C]//Proceedings of the 26th international conference on World Wide Web. 2017: 765-774.