Student Classroom Knowledge Tracking Based On Deep Semantics-robust Network

Renhua Gao*

Dalian University of Finance and Economics, Dalian, 116600, China

**Corresponding author. E-mail: rhguodufe@163.com*

Received: Dec. 15, 2024; Accepted: Jan. 07, 2025

Classroom knowledge tracking aims to predict future performance given past performance of students in educational applications. Although current classroom knowledge tracking methods achieve great prediction performance, there exist still two issues: (1) They exhibit sharp changes in knowledge state due to varying responses of students, resulting in semantic shifts in modelling knowledge state of students. (2) Transformer-based methods lack temporal information, limiting their ability to capture the gradual and cumulative nature of learning over time. To this end, a deep semantics-robust network is proposed via improving classroom knowledge tracking of students (DSN-KT) from three aspects, i.e., the data, the model, the loss, which significantly boosts the model stability and accuracy. Specifically, DSN-KT conducts the knowledge augmentation to generate data of different views for stable and robust estimation of knowledge states. And then, DSN-KT introduces the deep semantic learning within the transformer architecture with a time-cumulative attention, to capture the temporal dynamic information of students in learning knowledge. Meanwhile, DSN-KT devises the knowledge state prediction to provide the optimization function that optimizes prediction accuracy using semantic contrastive loss and cross-entropy loss. Three components work seamlessly to enable deep exploration and capture of students' knowledge state patterns. Finally, extensive experiments on four datasets show superiority and effectiveness of DSN-KT.

Keywords: Classroom knowledge tracking; knowledge augmentation; deep semantics-robust learning

http://dx.doi.org/10.6180/jase.202510_28(10).0017

---

1. Introduction

China Education Modernization 2035 plan highlights the need to accelerate educational transformation in the digital era, leveraging modern technologies to drive innovation in talent cultivation models and to achieve an organic integration of large-scale education with personalized learning [1–3]. This shift aims to build a personalized learning-based intelligent teaching support environment, advancing the application of artificial intelligence in education. With the widespread adoption of online education platforms such as massive open online courses and intelligent tutoring systems, vast amounts of teaching resources and student learning data are accumulating [4, 5]. By introducing knowledge tracing models in these systems, it is possible to monitor students' knowledge states in real time, apply personalized interventions, and dynamically adjust course progress as needed. Knowledge tracing (KT) has become a key application of artificial intelligence in education, forming an essential part of intelligent tutoring systems and being widely implemented across multiple online learning platforms [6, 7]. Unlike traditional face-to-face teaching, the advantage of online learning systems is that they can comprehensively record students' learning interactions, enabling more effective assessment of their knowledge mastery. However, the large disparity between the number of students and teachers in online education makes it challenging for teachers to accurately grasp each student's knowledge state and provide personalized guidance. By analyzing the extensive data generated during students' practice activities, teachers can gain a more accurate understanding of students' knowledge states, enabling more targeted instructional interventions. This data-driven approach not only enhances the effectiveness of personalized learning but also significantly optimizes the allocation of educational resources.

Deep Knowledge Tracing (DKT) model was introduced via leveraging Recurrent Neural Networks (RNN) or Long Short-Term Memory networks (LSTM) to dynamically model students' knowledge mastery [8]. Due to the high-dimensional, continuous latent state representations provided by RNNs and LSTMs, DKT can more accurately capture changes in students' knowledge states, breaking the assumption of independence between knowledge points and addressing the temporal dependencies in long-term learning. Building on DKT, various extensions have emerged. For example, Lyu et al. first extract spatial features from students' learning history sequences and then capture temporal features to reveal deeper latent information, effectively utilizing sequential patterns to improve prediction accuracy [9]. Sun et al. proposed the dynamic key-value memory network, which stores latent knowledge concepts through memory networks and directly evaluates students' understanding and mastery of each concept based on the relationships between them, thus updating students' knowledge proficiency levels [10]. To further enhance the generalization of knowledge tracing models, Pandey et al. introduced the self-attentive knowledge tracing (SAKT) model by incorporating the self-attention mechanism from the Transformer architecture [11], which effectively improves the model's flexibility and generalization performance in knowledge tracing tasks.

Despite notable advancements in existing knowledge tracing methods, two critical challenges persist in accurately modeling students' knowledge states. First, many current models exhibit excessive sensitivity to individual responses, leading to sharp fluctuations in students' knowledge states that fail to reflect their true mastery. This issue arises because most models treat each student interaction as an isolated event, where a single correct or incorrect response can disproportionately influence the assessment of a student's overall knowledge. In reality, a single response, whether correct or incorrect, should not drastically alter the evaluation of a student's cumulative knowledge. Learning is a gradual process, and student proficiency should be assessed over time, taking into account the broader context of learning patterns rather than isolated instances. Second, while Transformer-based models have shown promise in capturing relationships between learning events, they often neglect the temporal dimension of learning progression. These models, though powerful in modeling attention-based dependencies, lack mechanisms to account for the cumulative nature of knowledge acquisition over time. Students' knowledge states develop incrementally, with each learning interaction building upon the previous one. The inability to effectively model this gradual accumulation restricts the capacity of Transformer-based attention mechanisms to track long-term knowledge development accurately. Consequently, these models struggle to capture the evolving nature of student learning, hindering their ability to make reliable long-term predictions about student performance and mastery.

To this end, a deep semantics-robust network is proposed for the classroom knowledge tracking of students (DSN-KT), containing three components, i.e., knowledge augmentation, deep semantic learning, and knowledge state prediction. Specifically, DSN-KT devises knowledge augmentation to provide stable and robust data for knowledge state estimation, minimizing the influence of individual response fluctuations through techniques like random question dropping and adjacent item swapping, ensuring that the data used for modeling reflects a consistent, gradual accumulation of student knowledge. DSN-KT devises deep semantic learning to provide a model capable of capturing complex, sequential patterns in student knowledge states. By incorporating the Transformer model with a time-cumulative attention mechanism, DSN-KT effectively models the temporal dynamics of learning, allowing for an in-depth assessment of students' knowledge over time. DSN-KT devises knowledge state prediction to provide a loss function that optimizes prediction accuracy using semantic contrastive loss and cross-entropy loss. This approach enhances the model's ability to extract consistent, stable representations of students' knowledge states and predict future performance reliably. Finally, extensive experiments on four datasets show DSN-KT sets a new baseline in terms of ACC, AUC, MAE, and RMSE, verifying superiority and effectiveness of DSN-KT.

2. Multigraph-based deep programming ability tracing method for students

As shown in Fig. 1, a novel deep semantics-robust network architecture is proposed for the classroom knowledge tracking of students (DSN-KT), which greatly boosts stable and accuracy of the knowledge tracking. DSN-KT contains three components, i.e., knowledge augmentation, deep se-

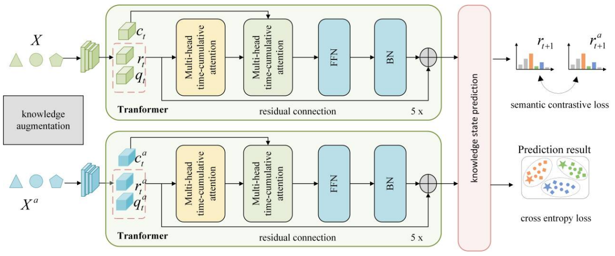

Fig. 1. The illustration of DSN-KT. DSN-KT devises knowledge augmentation to provide stable and robust data for knowledge state estimation, and devises deep semantic learning to provide a model capable of capturing complex. Meanwhile, DSN-KT devises knowledge state prediction to provide a loss function that optimizes prediction accuracy using semantic contrastive loss and cross-entropy loss.

mantic learning, and knowledge state prediction. Next, we provide a detailed description of the composition of the three components.

2.1. Problem definition

Knowledge tracing is based on students' past learning interaction records to track their knowledge mastery status and predict their future learning performance, which is one of the key research areas in online education. Given a student's exercise history record sequence $X = \{x_1, x_2, \ldots, x_i, \ldots, x_t\}$, the task is to predict whether the student can answer the question correctly at time $t+1$. Each x_i in the sequence X is a tuple (q_i, c_i, r_i), where $i \in \{1, 2, \ldots, t\}, q_i \in \{1, 2, \ldots, q\}$ is the question number, $c_i \in \{1, 2, \ldots, c\}$ is the number of the knowledge concept contained in the question, and $r_i \in \{0, 1\}$ is the student's response, with 1 indicating correct and 0 indicating incorrect. The knowledge tracing task can be formalized as:

$$P_t\left(r_{t+1} \mid q_{t+1}, c_{t+1}, X\right) \tag{1}$$

Through knowledge tracing, teachers can understand the students' level of knowledge mastery and learning needs, thereby providing personalized learning support and guidance. This allows teachers to adjust teaching strategies, provide targeted teaching materials, feedback, and supplementary learning resources according to the students' learning status, to maximize the promotion of students' learning outcomes and growth.

2.2. Knowledge augmentation

Students generally follow a continuous process in which their knowledge state gradually accumulates and develops. A single question's response is unlikely to entirely change a student's overall knowledge level. For example, a student may sometimes guess the correct answer by chance (occasional correctness) or make a mistake due to an error (occasional error), and such individual events should not significantly impact the assessment of the student's knowledge state. Therefore, to enhance the stability and robustness of the model's estimation of the student's knowledge state and to minimize the effects of individual response changes, knowledge enhancement strategies are designed, including random question dropping and adjacent item swapping. The purpose of the random question dropping strategy is to reduce the impact of individual question responses on the overall knowledge state assessment, allowing the evaluation to focus more on the student's overall knowledge mastery. This strategy operates by randomly omitting several questions in the answer sequence to lessen the influence of individual questions' chance-based deviations, thereby making the knowledge state estimation more stable and avoiding misjudgments due to occasional correctness or mistakes. The adjacent item swapping strategy addresses potential impacts of question order variations during the learning process by subtly adjusting the question order (randomly selecting and swapping adjacent questions), ensuring that the knowledge state assessment does not change significantly due to minor rearrangements. This strategy improves the evaluation's resistance to question order interference, allowing knowledge state estimation to more accurately reflect the student's true level without being affected by small changes in question order. Both strategies are based on the assumption of the overall continuity of the learning process, allowing the knowledge state assessment to rely less on individual events or question order but rather to reflect the student's performance across the entire sequence of responses.

2.3. Deep semantic learning

Knowledge Tracing (KT) tasks aim to precisely track and predict students' mastery of specific knowledge points. The core characteristics include handling the sequence of students' knowledge states over time and extracting fine-grained features from complex learning interactions. These characteristics require KT models to capture long-term dependencies in sequential data and extract complex features from students' responses that reflect their knowledge mastery status. The Transformer model, with its self-attention mechanism, is capable of handling long-distance dependencies in sequential data and extracting deep-level features from multiple perspectives, making it particularly well-suited to address the challenges of KT tasks. The capabilities of the Transformer enable it to provide more accurate and detailed analysis when dealing with the sequence of students' learning activities and assessing their knowledge states, making it an ideal choice for knowledge tracing tasks.

Specifically, given the knowledge augmentation datasets $D = \{X, X^a\}$, where X^a denotes the knowledge augmentation via random question dropping or adjacent item swapping, the process of the deep semantic learning within the transformer architecture. The original sequence X and the augmentation sequence X^a are converted into embedding vectors, including word embeddings and positional embeddings, so that the model can capture the context of each question as well as their positions in the sequence. Next, DSN-KT obtains question representation sequence q and learning ability representation sequence a based on initial questions and answer scores. Meanwhile, DSN-KT introduces the time-cumulative attention (TC) to model student understanding degrees u_t of questions based on q and a :

$$u_t^j = \sum_{\tau=1}^{t} \operatorname{maxout}\left(\frac{e^{Q_t^T K_\tau/\sqrt{d_k}}}{\sum_{i=1}^{t} e^{Q_t^T K_\tau/\sqrt{d_k}} \cdot e^{-\theta d(t-i)}}\right) \tag{2}$$

where $Q_t = q_t W^Q, K_t = q_t W^K, V_t = a_t W^V, W^Q, W^K$, and W^V are learnable weight matrices of the query, key, and values. τ denotes the considered historical time span. d_k is the dimension of the key vectors. Maxout($\bullet$) scales the largest softmax score to 1 after the softmax operation while maintaining the proportion of other scores. $d(t-\tau)$ denotes the cumulative effect along the time dimension. TC attention considers the temporal effects and cumulative effects in the learning process, allowing the model to better understand the changes in students' knowledge states over time.

Then, DSN-KT uses multi-head time-cumulative attention to learn information in parallel across different subspaces, with each head employing TC to extract features.

$$u_t = \operatorname{Concat}\left(\left\{u_t^j\right\}_{j=1}^{M}\right) W \tag{3}$$

where W is a learnable parameter matrix. Meanwhile, DSN-KT computes attention scores between question understanding and knowledge concepts to learn knowledge states of students, via the multi-head time-cumulative attention:

$$\begin{aligned} z_t^j &= \sum_{\tau=1}^{t} \operatorname{maxout}\left(\frac{e^{Q_n^T K_\tau/\sqrt{d_k}} \cdot e^{-\theta d(t-\tau)}}{\sum_{i=1}^{t} e^{Q_t^T K_\tau/\sqrt{d_k}} \cdot e^{-\theta d(t-i)}}\right) V_\tau \\ z_t &= \operatorname{Concat}\left(\left\{z_t^j\right\}_{j=1}^{M}\right) W \end{aligned} \tag{4}$$

where $Q_n = C_n W^Q, K_t = q_t W^K$, and $V_t = u_t W^V . C_n$ denotes the n-th knowledge concept. Subsequently, the feed-forward network further processes the output z_t, which helps the model learn more complex representations:

$$FFN\left(z_t\right) = \max\left(0, z_t \cdot W_1 + b_1\right) W_2 + b_2 \tag{5}$$

where W_1, W_2, b_1, and b_2 are the network parameters. After each sublayer (including the TC layer and the feed-forward network), we use residual connections and layer normalization to stabilize the training process and help the model better propagate gradients in deep networks.

$$\text{LayerNorm }\left(z_t + \text{Sublayer}\left(z_t\right)\right) \tag{6}$$

Finally, after processing through multiple Transformer blocks, the model outputs the latent representations Z and Z^a, corresponding to the representations of the original dataset and the knowledge augmentation dataset, respectively. These representations capture the students' knowledge states, providing rich information for subsequent knowledge tracing tasks. Subsequently, the feed-forward network further processes the output $z t $, which helps the model learn more complex representations:

$$Z = T\text{ransformer}(X), \quad Z^a = T\text{ransformer}\left(X^a\right) \tag{7}$$

where $Z = \{z_{t,1}, z_{t,2}, \ldots, z_{t,N}\}$ and $Z^a = \left\{z_{t,1}^a, z_{t,2}^a, \ldots, z_{t,N}^a\right\}$ denote proficiency estimations of the N knowledge concepts at the t-th time step. By introducing the time-cumulative attention, Transformer can more accurately capture the temporal dynamics and cumulative learning effects in the learning process, thus providing more accurate and stable prediction of students' knowledge states for knowledge tracing tasks.

2.4. Knowledge state prediction

A prediction head is designed via the multilayer perceptron to obtain answer predictions for the following question q_{t+1} :

$$\begin{aligned} r_{t+1} &= MLP\left(\text{fusion}\left(\bar{z}, q_{t+1}\right)\right) \\ r_{t+1}^a &= MLP\left(\text{fusion}\left(\bar{z}^a, q_{t+1}\right)\right) \end{aligned} \tag{8}$$

where $\bar{z}$ and $\bar{z}^a$ are the weighted sum of the proficiency estimations, given by:

$$\bar{z} = \sum_{i=1}^{N} \alpha_i z_{i,t}, \quad \bar{z}^a = \sum_{i=1}^{N} \beta_i z_{i,t}^a \tag{9}$$

and the attention weights α_i and β_i are calculated using the softmax function:

$$\alpha_i = \frac{e^{q_t^T z_{i,t}}}{\sum_{j=1}^{N} e^{q_t^T z_{j,t}}}, \quad \beta_i = \frac{e^{q_t^T z_{i,t}^a}}{\sum_{j=1}^{N} e^{q_t^T z_{j,t}^a}} \tag{10}$$

In DSN-KT, semantic contrastive loss and cross-entropy loss are selected to optimize the network. The reasons are as follows: semantic contrastive loss is used to ensure that the model can extract stable and consistent knowledge states from similar learning histories. This method helps the model learn more robust feature representations by comparing different but semantically similar learning paths. In the context of knowledge tracing, this means the model can better understand and predict students' knowledge states, even when faced with different question patterns. Cross-entropy loss is used to directly optimize the model's accuracy in predicting whether a student's answer is correct or not. In classification problems, cross-entropy loss can effectively measure the difference between the model's predicted probability distribution and the true label's probability distribution, and train the model by minimizing this difference. In the knowledge tracing task, this helps the model more accurately predict the student's performance on subsequent questions.

For answers r_{t+1} and r_{t+1}^a of the same question, semantic contrastive loss is defined as follows:

$$L_{\text{contrastive}} = -\log \exp\left(\operatorname{sim}\left(r_{t+1}, r_{t+1}^a\right) / \tau\right) \tag{11}$$

where $\operatorname{sim}(\cdot)$ is the similarity function, typically the cosine similarity, τ is the temperature parameter:

$$L_{ce} = -\sum_{i=1}^{C} \left(\left(y_{t+1} \log\left(r_{t+1}\right) + \left(y_{t+1} \log\left(r_{t+1}^a\right)\right)\right) \tag{12}$$

where C is the number of classes, y_t is the one-hot encoding of the true label, and r_t is the probability of the i-th class predicted by the model. The final loss function is a weighted sum of these two losses:

$$L_{\text{total}} = \lambda L_{\text{contriastive}} + (1 - \lambda) L_{ce} \tag{13}$$

where λ is a hyperparameter that balances the importance of the two losses. By adjusting λ, the focus of the model between contrastive learning and cross-entropy loss can be controlled.

Cross-entropy loss optimizes knowledge tracing by predicting the probability of correct or incorrect student responses based on their current knowledge state. Contrastive loss ensures stable and consistent knowledge state representations, distinguishing between different levels of mastery by minimizing the distance between similar states and maximizing the distance between different levels. By combining both loss functions, DSN-KT accurately predicts student responses and learns robust knowledge representations, significantly enhancing its performance in tracking learning progression and predicting future knowledge states.

3. Result and discussion

3.1. Set up

Dataset and Metric: Four knowledge tracing datasets in various learning scenarios are utilized to evaluate the performance between our model and baselines. Assist09 [12], an online tutoring wedpage built in 2004, consists of 4217 students and 26688 questions with 123 concepts where each question contains 346860 responses with the average length 82.25. Assist17 [13], having much longer learning sequences, consists of 1079 students and 3162 questions with 102 concepts where each question contains 942816 responses with the average length 873.79. KDDcup [14] consists of 574 students and 1084 questions with 138 concepts where each question contains 809694 responses with the average length 1410.62. Statics [15], which collects the engineering statics course in the Carnegie Mellon university, consists of 331 students with 138 concepts where each question contains 142124 responses with the average length 429.38. Consistent with current methods, MAE, RMSE, ACC, and AUC are used to evaluate the performance between our model and baselines [16].

Implementation Details: In the experiments, the model is implemented using PyTorch and trained on an NVIDIA GeForce RTX 3090 GPU. The model is optimized using the Adam optimizer with a learning rate of 0.001, which is empirically selected to ensure stable convergence during training. A Dropout rate of 0.2 is applied to prevent overfitting and improve generalization, especially when dealing with limited data. The model architecture consists of 3 layers in the Transformer block, with 8 attention heads per layer to capture various relationships between learning events. To regularize the model, L2 weight decay is applied with a value of 0.00001, which helps prevent overfitting by penalizing excessively large weights. The model is trained for 100 epochs, and a batch size of 256 is used to handle large datasets efficiently while optimizing memory usage. The dataset is split into 80% for training and 20% for testing, ensuring proper validation of the model on unseen data.

3.2. Comparison with baselines

Comparison baselines: Eight deep knowledge tracing methods are compared on four datasets in terms of AUC, ACC, MAE and RMSE, to demonstrate the performance of our model, containing DKT, CAKT, CKT, RKT, EDKT, Bi-CLKT, DGEKT, and TRKT.

Comparison results: DSN-KT consistently outperforms existing models across key evaluation metrics, including AUC, ACC, MAE, and RMSE, as demonstrated in Tables 1 and 2, showcasing its superiority in accurately tracking and predicting students' knowledge states. This performance advantage is driven by three key innovations. DSN-KT improves knowledge state estimation through several key innovations. Data augmentation techniques like random question dropping and adjacent item swapping reduce noise and stabilize knowledge state representations. The model architecture leverages deep semantic learning within an enhanced Transformer framework, incorporating a time-cumulative attention mechanism to capture long-term dependencies and the cumulative nature of knowledge acquisition. This enables DSN-KT to model student learning sequences more effectively, addressing the temporal dynamics often overlooked by traditional models. The dual-loss function, combining semantic contrastive loss and cross-entropy loss, ensures consistent knowledge state representations while optimizing prediction accuracy. Together, these elements enable DSN-KT to outperform traditional models by addressing data noise, long-term dependencies, and improving prediction accuracy for personalized learning.

3.3. Ablation Study

DSN-KT performs three ablation experiments across all datasets, focusing on accuracy (ACC). Variant-1 uses a traditional Transformer with cross-entropy loss, variant-2 adds time-cumulative attention, and variant-3 incorporates both cross-entropy and semantic contrastive loss. Key observations include: (1) Variant-2 improves accuracy by capturing temporal dependencies, enabling a better representation of students' evolving knowledge. (2) Variant-3 further enhances performance by stabilizing knowledge state representations and reducing noise. (3) The full DSN-KT model, combining both time-cumulative attention and semantic contrastive loss, achieves the highest accuracy, demonstrating the synergistic effect of these components in improving model robustness and prediction precision.

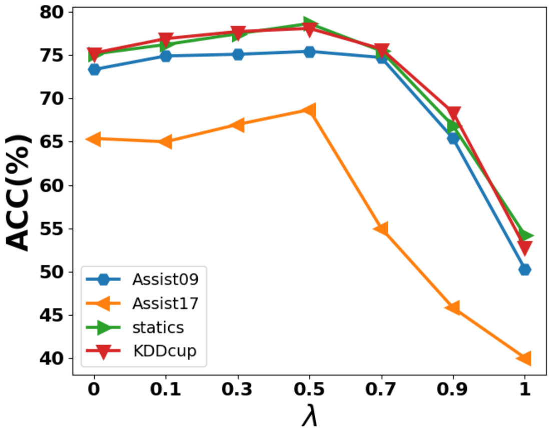

Fig. 2. Parameter analysis of DSN-KT.

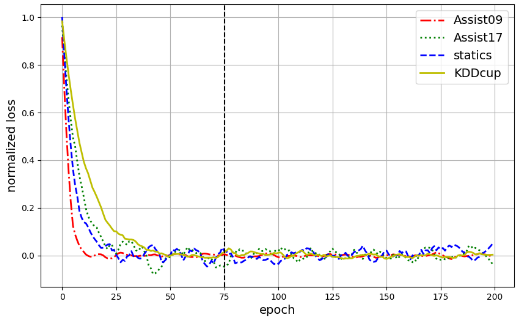

Fig. 3. Convergence analysis of DSN-KT.

3.4. Parameter analysis

The parameter λ plays a crucial role in balancing the cross-entropy loss and semantic contrastive loss in DSN-KT, impacting model accuracy across the Assist09, Assist17, Statics, and KDDcup datasets. As shown in Fig. 2, ACC for different values of λ suggest that performance improves as λ increases from 0 to around 0.5 across all datasets. For example, the highest ACC for Assist09 is achieved at $\lambda = 0.5$ with a score of 75.46%, while Assist17 also peaks around this value at 68.11%. Similarly, Statics and KDDcup exhibit optimal ACC values of 78.68% and 78.12%, respectively, around $\lambda = 0.5$. However, beyond $\lambda = 0.5$, performance starts to decline for each dataset, particularly at $\lambda = 0.9$ and $\lambda = 01$, indicating that over-emphasizing the semantic contrastive loss degrades model performance. The observed performance trend suggests that moderate values of λ, especially around 0.5 , strike a beneficial balance between the cross-entropy loss and the semantic contrastive loss. This balance appears optimal as it allows the model to leverage both classification accuracy and semantic separation without overpowering one aspect. When λ becomes too high, the semantic contrastive loss likely overshadows the cross-entropy component, reducing the model's effectiveness in aligning predictions with ground truth. Therefore, maintaining an intermediate value for λ ensures that both objectives contribute meaningfully to model training, yielding the best performance across datasets.

Table 1. Comparison of DSN-KT on the Assist09 and Assist17 datasets.

	Assist09				Assist17			
Method	ACC	AUC	MAE	RMSE	ACC	AUC	MAE	RMSE
DKT [8]	0.6482	0.6545	0.4015	0.4387	0.5571	0.5409	0.5143	0.5543
CAKT [17]	0.6533	0.6512	0.3862	0.4274	0.6185	0.5875	0.4845	0.5077
CKT [18]	0.6946	0.7026	0.3714	0.4255	0.6314	0.6555	0.4641	0.4812
RKT [3]	0.8946	0.7811	0.8635	0.4195	0.6566	0.6492	0.4314	0.4755
EDKT [12]	0.7414	0.7978	0.3251	0.4199	0.6496	0.6367	0.4194	0.4555
Bi-CLKT [2]	0.7366	0.7700	0.3333	0.4158	0.6612	0.6662	0.4214	0.4655
TRKT [19]	0.7441	0.7887	0.3234	0.4215	0.6312	0.6576	0.4211	0.4651
DGEKT [20]	0.7533	0.7900	0.3200	0.4115	0.6802	0.6867	0.4157	0.4602
DSN-KT	**0.7546**	**0.8016**	**0.3194**	**0.4101**	**0.6811**	**0.6886**	**0.4155**	**0.4517**

Table 2. Comparison of DSN-KT on the statics and KDDcup datasets.

	statics				KDDcup			
Method	ACC	AUC	MAE	RMSE	ACC	AUC	MAE	RMSE
DKT [8]	0.7355	0.8182	0.3215	0.3956	0.8263	0.7558	0.3498	0.3798
CAKT [17]	0.7483	0.8112	0.3162	0.3974	0.8288	0.7577	0.3245	0.3777
CKT [18]	0.7014	0.8222	0.3323	0.3855	0.8311	0.7665	0.3241	0.3792
RKT [3]	0.7311	0.8219	0.2851	0.3846	0.8266	0.7779	0.2549	0.3697
EDKT [12]	0.7649	0.8266	0.2851	0.3797	0.8300	0.6367	0.4194	0.4555
Bi-CLKT [2]	0.7546	0.7700	0.2933	0.3815	0.8312	0.7767	0.2611	0.3651
TRKT [19]	0.7554	0.7809	0.2838	0.3810	0.8321	0.7706	0.2601	0.3671
DGEKT [20]	0.7666	0.8200	0.2883	0.3801	0.8333	0.7799	0.2414	0.3615
DSN-KT	**0.7868**	**0.8282**	**0.2794**	**0.3721**	**0.8354**	**0.7793**	**0.2386**	**0.3517**

3.5. Convergence analysis

The convergence analysis across four datasets—Assist09, Assist17, statics, and KDDcup—reveals distinct loss patterns over training epochs. As shown in Fig. 3, all datasets show rapid loss reduction, indicating effective parameter adjustments. However, as training progresses, loss curves plateau, signaling convergence. The Assist09 dataset converges smoothly with a lower loss, suggesting a better fit or simpler data. In contrast, the Assist17 dataset experiences erratic convergence, likely due to noise or high data variability. The statics and KDDcup datasets stabilize after 75 epochs with minor final loss differences. These varied convergence behaviors highlight the need for dataset-specific tuning, adjusting hyperparameters and training strategies to address dataset complexity and noise. Further investigation into fluctuating loss could improve model robustness and generalization across datasets.

4. Conclusions

DSN-KT represents a significant advancement in classroom knowledge tracking by addressing challenges in data stability, temporal modeling, and prediction accuracy. Through knowledge augmentation, deep semantic learning, and a dual-loss function, DSN-KT offers a robust framework for accurately capturing and predicting students' knowledge states. Extensive experiments show that it outperforms existing models across multiple metrics, setting a new benchmark in knowledge tracing. This model enhances stability and precision in knowledge state estimation, enabling more personalized and adaptive learning interventions. For future development, scalability for large-scale, real-time applications is a key focus. Techniques like model compression or distributed learning could be explored to improve efficiency. Additionally, incorporating multi-modal learning, such as student behavior and engagement data, could further personalize and enhance the model's predictions. These improvements would strengthen the model's applicability and contribute to more personalized, data-driven educational interventions.

References

[1] Y. Jiang, B. Zhang, Y. Zhao, and C. Zheng, (2022) *"China's preschool education toward 2035: Views of key policy experts"* **Ecnu review of education** **5**(2): 345–367. DOI: 10.1177/20965311211012705.

[2] Y. Lu, (2024) *"Application of AI in the Field of Documentary Heritage: A Review of the Literature"* **Journal of Artificial Intelligence Research** ***1***(2): 22–36. DOI: 10.70891/JAIR.2024.110005.

[3] J. Gao, M. Liu, P. Li, A. A. Laghari, A. R. Javed, N. Victor, and T. R. Gadekallu, (2023) *"Deep incomplete multi-*

Table 3. Ablation experiments of DSN-KT on four datasets in terms of ACC.

	Assist09	Assist17	statics	KDDcup
variant-1	0.6946	0.6323	0.7212	0.7956
variant-2	0.7127	0.6412	0.7546	0.8198
variant-3	0.7318	0.6537	0.7532	0.8211
DSN-KT	**0.7546**	**0.6811**	**0.7868**	**0.8354**

view clustering via information bottleneck for pattern mining of data in extreme-environment IoT" **IEEE Internet of Things Journal**: DOI: 10.1109/JIOT.2023.3325272.

[4] S. Shen, Q. Liu, Z. Huang, Y. Zheng, M. Yin, M. Wang, and E. Chen, (2024) *"A survey of knowledge tracing: Models, variants, and applications"* **IEEE Transactions on Learning Technologies**: DOI: 10.1145/3569576.

[5] P. Li, J. Gao, J. Zhang, S. Jin, and Z. Chen, (2022) *"Deep Reinforcement Clustering"* **IEEE Transactions on Multimedia**: DOI: 10.1109/TMM.2022.3233249.

[6] J. Gao, M. Liu, P. Li, J. Zhang, and Z. Chen, (2023) *"Deep Multiview Adaptive Clustering With Semantic Invariance"* **IEEE Transactions on Neural Networks and Learning Systems**: DOI: 10.1109/TNNLS.2023.3265699.

[7] B. Xu, Z. Huang, J. Liu, S. Shen, Q. Liu, E. Chen, J. Wu, and S. Wang. "Learning behavior-oriented knowledge tracing". In: *Proceedings of the 29th ACM SIGKDD conference on knowledge discovery and data mining*. 2023, 2789–2800. DOI: 10.1145/3580305.3599407.

[8] C. Piech, J. Bassen, J. Huang, S. Ganguli, M. Sahami, L. J. Guibas, and J. Sohl-Dickstein, (2015) *"Deep knowledge tracing"* **Advances in neural information processing systems** ***28***: DOI: 10.1016/j.knosys.2022.108274.

[9] L. Lyu, Z. Wang, H. Yun, Z. Yang, and Y. Li, (2022) *"Deep knowledge tracing based on spatial and temporal representation learning for learning performance prediction"* **Applied Sciences** ***12***(14): 7188. DOI: 10.3390/app12147188.

[10] J. Sun, M. Wei, J. Feng, F. Yu, Q. Li, and R. Zou, (2024) *"Progressive knowledge tracing: Modeling learning process from abstract to concrete"* **Expert Systems with Applications** ***238***: 122280. DOI: 10.1016/j.eswa.2023.122280.

[11] S. Pandey and G. Karypis, (2019) *"A self-attentive model for knowledge tracing"* **arXiv preprint arXiv:1907.06837**: DOI: 10.48550/arXiv.1907.06837.

[12] Z. Liu, Q. Liu, J. Chen, S. Huang, B. Gao, W. Luo, and J. Weng. "Enhancing deep knowledge tracing with auxiliary tasks". In: *Proceedings of the ACM Web Conference 2023*. 2023, 4178–4187. DOI: 10.1145/3543507.3583866.

[13] M. Zhang, X. Zhu, and Y. Ji. "Input-aware neural knowledge tracing machine". In: *Pattern Recognition. ICPR International Workshops and Challenges: Virtual Event, January 10–15, 2021, Proceedings, Part IV*. 2021, 345–360. DOI: 10.1007/978-3-030-68799-1_25.

[14] H. Ma, Y. Yang, C. Qin, X. Yu, S. Yang, X. Zhang, and H. Zhu. "HD-KT: Advancing Robust Knowledge Tracing via Anomalous Learning Interaction Detection". In: *Proceedings of the ACM on Web Conference 2024*. 2024, 4479–4488. DOI: 10.1145/3589334.3645718.

[15] K. Zhang, T. Ji, and H. Zhang, (2024) *"Knowledge tracing via multiple-state diffusion representation"* **Expert Systems with Applications**: 124797. DOI: 10.1016/j.eswa.2024.124797.

[16] S. Minn, Y. Yu, M. C. Desmarais, F. Zhu, and J.-J. Vie. "Deep knowledge tracing and dynamic student classification for knowledge tracing". In: *2018 IEEE International conference on data mining (ICDM)*. 2018, 1182–1187. DOI: 0.1109/ICDM.2018.00156.