

Article

Research on Knowledge Tracing-Based Classroom Network Characteristic Learning Engagement and Temporal-Spatial Feature Fusion

Zhaoyu Shou ^{1,2} , Yihong Li ¹, Dongxu Li ^{1,*}, Jianwen Mo ¹ and Huibing Zhang ³

¹ School of Information and Communication, Guilin University of Electronic Science Technology, Guilin 541004, China; guilinshou@guet.edu.cn (Z.S.); 21022201032@mails.guet.edu.cn (Y.L.); jwmo@guet.edu.cn (J.M.)

² Guangxi Wireless Broadband Communication and Signal Processing Key Laboratory, Guilin University of Electronic Technology, Guilin 541004, China

³ School of Computer and Information Security, Guilin University of Electronic Technology, Guilin 541004, China; zhanghuibing@guet.edu.cn

* Correspondence: ldx@mails.guet.edu.cn

Abstract: To accurately assess students' cognitive state of knowledge points in the learning process within the smart classroom, a knowledge tracing (KT) model based on classroom network characteristic learning engagement and temporal-spatial feature fusion (CL-TSKT) is proposed. First, a classroom network is constructed based on the information of the student ID, seating relationship, student–student interaction, head-up or head-down state, and classroom network characteristics obtained from a smart classroom video. Second, a learning engagement model is established by utilizing the student–student interactions, head-up or head-down state, and classroom network characteristics. Finally, according to the learning engagement model and the knowledge point test data, a parallel temporal attention GRU network is proposed. It is utilized to extract the temporal features of the knowledge points and learning engagement. They are fused to obtain the knowledge point–learning engagement temporal characteristics and their associated attributes. Meanwhile, a CNN is used to extract the knowledge point–knowledge point spatial features. We consider the associative properties of knowledge point–knowledge points from a spatial perspective and fuse the knowledge point–knowledge point spatial features with the knowledge point–learning engagement temporal features. To accurately characterize the cognitive state of the knowledge points and provide effective support for teachers' accurate and sustainable interventions for learners in the teaching and learning process, this paper conducts extensive experiments on four real datasets. The CL-TSKT model in this paper shows superior performance in all four evaluation metrics, compared with the state-of-the-art KT models.

Keywords: knowledge tracing; classroom network characteristics; learning engagement; temporal-spatial feature fusion; sustainable interventions



Citation: Shou, Z.; Li, Y.; Li, D.; Mo, J.; Zhang, H. Research on Knowledge Tracing-Based Classroom Network Characteristic Learning Engagement and Temporal-Spatial Feature Fusion.

Electronics **2024**, *13*, 1454. <https://doi.org/10.3390/electronics13081454>

Received: 14 March 2024

Revised: 4 April 2024

Accepted: 9 April 2024

Published: 11 April 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In today's competitive educational environment, it is critical to improve students' learning outcomes. The cognitive state, on the other hand, is one of the most important indicators of a student's learning effectiveness. Teachers analyze students' cognitive states. This helps to understand how students respond to the content and format of the classroom and optimize the classroom format for improved student learning outcomes. It also provides individualized intervention approaches for different students. Meanwhile, a large number of researchers have analyzed students' cognitive states in terms of the KT models. Janssen et al. [1] proposed an item response theory (IRT) that describes a student's cognitive state as a one-dimensional competence value. The cognitive states of students

were also modeled in combination with the difficulty and differentiation of the test questions. Li et al. [2] proposed a knowledge relation rank enhanced heterogeneous learning interaction model. Graph convolutional networks (GCNs) are used to capture the complex interactions between exercises and knowledge points.

However, existing knowledge tracing (KT) models were limited to tracking and predicting student performance based on test data from students on online education platforms. At the same time, offline teaching methods are still the mainstay of current education methods, and there is a correlation between students' cognitive states and multiple factors in smart classrooms. These factors include learning engagement, interactions between students, and the environment in which they are located [3,4]. Therefore, existing knowledge tracing models, when applied to complex smart classroom environments, will expose the limitations of single-dimensional data analysis. Characterizing students' cognitive states by comprehensively considering their learning engagement characteristics in smart classroom environments can lead to better understanding and facilitate their learning process.

Learning engagement reflects the degree of students' attention and participation in a learning task. To some extent, it reflects the state of student learning [5]. Hu et al. [6] constructed a classroom video database based on noninvasive classroom video. They used YOLOV5 to analyze and assess student behavior and gauge students' learning engagement in the classroom. At the same time, students are easily influenced by the students around them in the classroom learning process, and this implicit impact is often reflected in students' classroom learning statuses and academic performance. Lu et al. [7] found that there was a significant effect from students' seating choices on classroom learning engagement. Putnik et al. [8] found that there was also a significant correlation between student learning outcomes and social network structure characteristics. Xiang et al. [9] proposed a method for evaluating students' attention based on emotional evolution and viral transmission. It showed that students' attention in the spatial dimension of the classroom was affected by a different area of negative and positive emotions. Rijsewijk et al. [10] analyzed the helping relationships among students. It was found that there was an association between academic achievement and the unbalanced distribution of classroom helping relationships. Gutierrez et al. [11] found that effective seating arrangements promoted social interaction and classroom participation. Li et al. [12] proposed a one-of-a-kind model of emotion transmission determined by students' facial expressions. It showed that there was student-to-student propagation of learning emotions in the classroom network, thereby affecting individual learning.

In summary, students' cognitive states in smart classrooms are influenced by multiple factors and difficult to accurately characterize. This paper is based on classroom network characteristic learning engagement and knowledge point test data. A knowledge tracing model based on classroom network characteristic learning engagement and temporal-spatial feature fusion (CL-TSKT) is proposed. A parallel temporal attention GRU network is proposed. It is utilized to extract the temporal features of knowledge points and learning engagement. They are fused to obtain the knowledge point-learning engagement temporal characteristics and their associated attributes. Meanwhile, a CNN is used to extract the knowledge point-knowledge point spatial features. We consider the associative properties of knowledge point-knowledge points from a spatial perspective and fuse the knowledge point-knowledge point spatial feature with the knowledge point-learning engagement temporal feature, achieving accurate characterization of learners' cognitive states in smart classrooms.

This paper consists of the following main contributions:

1. For accurately assessing student learning engagement in smart classrooms, a learning engagement model utilizing student–student interactions, student head-up states, and classroom network characteristics is proposed.
2. A temporal-spatial feature fusion algorithm is proposed. A parallel temporal attention GRU network is designed which is utilized to extract the temporal features of knowledge points and learning engagement. They are fused to obtain the knowledge

point-learning engagement temporal characteristics and their associated attributes. Meanwhile, a CNN is used to extract the knowledge point-knowledge point spatial features. We consider the associative properties of knowledge point-knowledge points from a spatial perspective and fuse the knowledge point-knowledge point spatial features with the knowledge point-learning engagement temporal features. To maintain the integrity of the characterization information, the model incorporates classroom network characteristic learning engagement and knowledge point test data to analyze the cognitive states. It avoids the limitations of single-dimensional data analysis and can more accurately characterize learners' cognitive states.

3. Extensive experiments are conducted on four real datasets. They show that the CL-TSKT model proposed in this paper has advantages over the state-of-the-art knowledge tracing algorithms.

The remainder of the paper is organized as follows. Section 2, "Related Works", describes the research on knowledge tracing. Section 3, "Problem Definitions", describes the algorithm-related definitions and calculations. Section 4, "CL-TSKT Model", describes the CL-TSKT model's architecture in detail. Section 5, "Analysis of Experiments and Experimental Results", provides a comparative analysis of the experimental results to evaluate the performance of the algorithm. Finally, Section 6, "Conclusions and Future Work" is a summary and outlook of the work.

2. Related Work

The KT model is designed to measure a student's level of knowledge based on the student's historical interactions. Existing KT models are mainly categorized into statistics-based and deep learning-based models. De et al. [13–16] proposed different DINA models based on students' test situations and knowledge point correlation matrices, and they modeled the students' cognitive states as state vectors of multidimensional knowledge points. Statistics-based KT models could make predictions interpretable. However, students' cognitive abilities could not be assessed dynamically, and the accuracy of the predictions was low. With the rapid development of deep learning, and to address the low predictive accuracy of traditional KT models and the inability to adaptively predict students' cognitive states, a seminal model, the KT model (DKT), was proposed by Piech et al. [17], based on recurrent neural networks (RNNs). As the first demonstration of the effectiveness and potential of deep learning for KT tasks, it solved the problem that traditional models could not learn the characteristics of knowledge relationships autonomously. Ghosh et al. [18] used exponential decay and the context-aware relative distance to compute the attention weights and enhance model feature extraction. Better prediction performance than that of the DKT model was obtained. To address the problem of weak modeling of knowledge point relationships, Yang et al. [19] proposed a novel quantitative relationship neural network for the explainable cognitive diagnosis model (QRCDM). It used explicit and implicit correlations between exercises and corresponding knowledge concepts to calculate student errors and guesses. Sequence modeling KT models took the lead in analyzing students' knowledge states from test data and have achieved some success in the KT field. However, the limitations of single-dimensional data analysis have led a large number of researchers to consider adding data features outside of the test data so as to improve the model prediction performance.

Therefore, Xiao et al. [20] proposed a knowledge tracing model based on multi-feature fusion (KTMFF). It combined features such as the practice texts and test consumption time, and it fused multiple features using a multi-head self-attention mechanism. Liu et al. [21], by incorporating the relationship between students' knowledge states, knowledge concepts, and exercises, modeled students' cognitive responses using a log-linear model. Huang et al. [22] proposed a knowledge tracing model based on temporal and causal enhancement. Meanwhile, a causal self-attention mechanism based on the theory of front-door adjustment was introduced to enhance the interaction representation. It effectively incorporated interval and response features into the model. Text-aware KT models took

into account textual content and analyzed the tiers of difficulty, which was effective in increasing the accuracy of predicting student performance. However, because of the lack of uniformity between datasets and the difficulty of marking up text, it was not widely available for use.

To address the interpretability of KT models and the presence of student forgetfulness, Jiang et al. [23] applied Markov blankets to the KT model. The method used Markov blankets of the target variable as a subset of the features, and interpretable machine learning techniques were applied to the KT model to improve the interpretability of the models. Lee et al. [24] proposed a KT model where contrastive learning reveals semantic similarity. To address the issue of their under-consideration, Im et al. [25] proposed the Forgetting-Aware Linear Bias (FoLiBi) and applied it to the KT model based on a contrastive learning framework (CL4KT). For the heterogeneity of course knowledge structures and the sparseness of interaction records, Ni et al. [26] proposed the learner-question interaction-based heterogeneous graph neural network (HHSKT) model and enhanced modeling by using hierarchical heterogeneous knowledge structures and short-term memory. It obtained the effects of different sequences of practice interactions on learners' knowledge states. Forgetting-aware KT models took into account the presence of learning forgetting behaviors in students. However, it made inferences based primarily on students' practiced responses to the questions, and the reality of the knowledge state is often complex and ambiguous. Therefore, considering forgetting behavior in a single-answer situation would lack reliability.

With the rapid development of graph neural networks and their progress in spatial information feature extraction, Wu et al. [27] proposed a KT model based on session graphs, and the gated graph neural network was utilized to obtain the students' knowledge state from the session graphs. To address the problem that existing KT models ignore the correlation between multiple knowledge concepts in exercise, Huang et al. [28] proposed neural Turing machine-based skill-aware knowledge tracing (NSKT). It modeled students' states of knowledge more accurately by capturing potential correlations between knowledge concepts in the exercises. For the inability of most knowledge tracing methods to capture coarse-grained inter-type associations, Zhao et al. [29] proposed a graph-enhanced multi-activity KT (GMKT) model. It modeled student cognition by jointly learning a fine-grained recurrent memory-enhanced model of student knowledge and a coarse-grained graph neural network. Graph-based KT models could be effective in learning knowledge point-knowledge points and knowledge point-exercise representations, and it has achieved some success in the extraction of spatial feature information. However, graph networks are constructed on the premise that connections between data need to be made in advance and preconstructed at the time of input. Thus, the scope of use was limited.

Based on the above analysis of KT models, this paper provides an overview of the relevant research models, as shown in Table 1.

In summary, to address the difficulty of existing KT models in characterizing the cognitive states of students in smart classrooms, this paper is based on classroom network characteristic learning engagement and knowledge point test data. For comprehensive consideration of students' learning engagement as influenced by the students around them while learning in the smart classroom, a parallel temporal attention GRU network is proposed. It is utilized to extract the temporal features of knowledge points and learning engagement. They are fused to obtain the knowledge point-learning engagement temporal characteristics and their associated attributes. Meanwhile, a CNN is used to extract the knowledge point-knowledge point spatial features. We consider the associative properties of knowledge point-knowledge points from a spatial perspective and fuse the knowledge point-knowledge point spatial features with the knowledge point-learning engagement temporal features, achieving accurate knowledge tracing.

Table 1. Summary of relevant research models.

Models	Paper Numbers	Advantages	Limitations
Traditional KT models	[1,13–16]	Highly interpretable	Static diagnostics, multiple knowledge points difficult to interpret
Sequence modeling KT models	[17–19]	Improved prediction accuracy by fully utilizing test data	Single-dimensional data analysis, lack of consideration of other factors affecting students' cognitive states
Text-aware KT models	[20–22]	Full consideration of exercise data, with different levels of difficulty associated with text content	Poor harmonization of datasets and difficulties in text tagging
Forgetting-aware KT models	[23–26]	Considering students' presence of learning forgetting behaviors based on practice answer situations	Knowledge states in reality are often complex and ambiguous, and single-answer situation analysis of forgetting lacks reliability
Graph-based KT models	[2,27–29]	Considering the spatial relationship between knowledge point-knowledge points, automatically acquiring side rights, and updating cognitive abilities	The existence of data correlation needs to be pre-assumed, and limited scope of use

3. Problem Definitions

3.1. Symbol Definition

Based on the course data, the set of learners of a course is represents by $U = \{u_1, u_2, \dots, u_n\}$. Here, n is the number of learners. The set of knowledge points of the course is represents by $KP = \{kp_1, kp_2, \dots, kp_i, \dots, kp_k\}$, where kp_i represents the i th knowledge points (KPs), with a total of k knowledge points. The test set for the course is represents by $E = \{e_1, e_2, \dots, e_m\}$, with a total of m test questions. The inclusion relationship between the test questions and the knowledge points is represented by the matrix $Q = (q_{jk}) \in R^{m \times k}$, in which $q_{jk} \in \{0, 1\}$ indicates whether test question e^j contains the knowledge point kp_k . The test record for the learner u_i is recorded as $X_i = \{x^1, x^2, \dots, x^m\} \in R^m$.

3.2. Modeling Learning Engagement Based on Classroom Network Characteristics

3.2.1. Learning Engagement

In smart classrooms, students' head-up learning attention states usually correspond to their engagement while learning. A higher head-up rate would be more reflective of students being attentive and engaged in the classroom. Therefore, in a smart classroom, the students' heads-up learning states are monitored through the video surveillance system, and through correlation analysis with learning engagement, teachers can better observe and assess students' learning states. Based on the analytical model proposed by Shou et al. [30] for assessing learners' head posture in smart classroom videos, the head-up rate of learner u_i during the duration of learning k knowledge points is obtained by

$$L_{u_i k} = \frac{t}{m} \quad (1)$$

Here, m represents the number of image frames extracted within the knowledge point, and t represents the number of head-up frames. In this paper, the learner head-up rate obtained from the head posture assessment model is used as the learning engagement in smart classrooms.

3.2.2. Learning Engagement Based on Classroom Network Characteristics

In a smart classroom, it is easier for students with seats near each other to interact socially and influence each other's learning [31]. Therefore, the first step is to construct an empowered and undirected network $G_{UR} = (U, E_{UR})$ of seating relationships in a single

classroom based on the students' classroom seating relationships. Here, U represents the set of student nodes, E_{UR} represents the set of edges, the purple circle represents a student, and the black circles represent the empty positions in the class. This is shown in Figure 1.

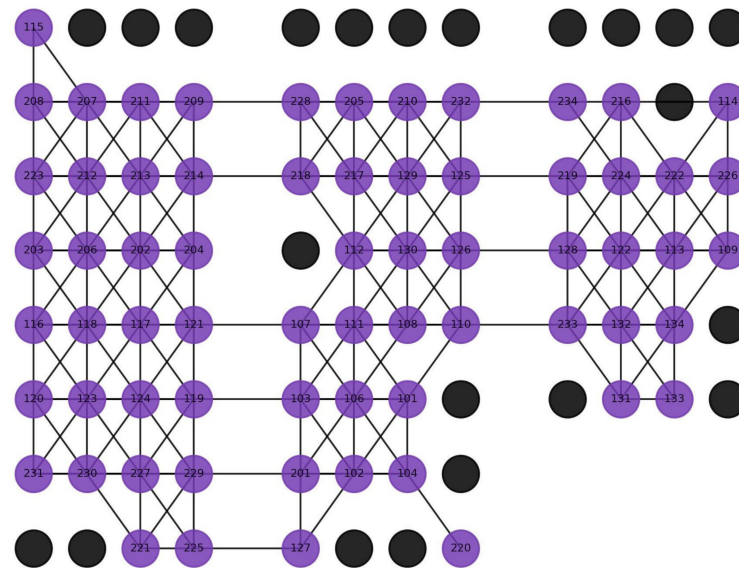


Figure 1. Single classroom seating relationship network.

Second, in social networks, degree refers to the sum of the connected edges of the nodes in the network. The greater the degree of a node, the higher the degree centrality of the node, the more important the node is in the network, and at the same time, the greater the influence on the nodes with which the node has existing connections [32]. The degree of centrality DC_{u_i} of nodes based on the constructed single-classroom empowered undirected network of seating relationships is obtained as shown in the following equation:

$$DC_{u_i} = \frac{k_{u_i}}{N - 1} \quad (2)$$

Meanwhile, in social networks, the frequency of interaction between nodes is also an important factor which reflects the mutual influence between their nodes. The higher the frequency of interaction, the higher the similarity between nodes [33]. Therefore, the frequency of inter-node interactions is high. Their learning engagement during classroom learning will be similar and more influenced by it. This paper obtains the frequency of inter-node interactions Ins_{u_i, u_j} based on classroom video monitoring, and the interaction frequency is normalized as shown in Equation (3):

$$Inter_{u_i, u_j} = \frac{1}{m} Ins_{u_i, u_j} \quad (3)$$

Then, the nearest neighbor nodes around a node are selected as learning influence factors based on the Nearest Neighbor Effective Distance Criterion (NEDC) [12]. The solid box nodes in Figure 2 indicate the NEDC range's nearest neighbor nodes that have a learning influence on the node. Here, the arrow in Figure 2 represents the direction of the student facing the blackboard.

Finally, considering the degree of centrality of students in the classroom network, the Euclidean distance between students, and the frequency of interactions between students, we calculate the side weights between students in a single classroom network. Thus, the specific formulation of the network $G_{UR} = (U, E_{UR})$ of empowered and undirected seating relationships for a single classroom is given in the following equation:

$$\begin{aligned} Edge_{u_i, u_j} &= DC_{u_i} + Inter_{u_i, u_j} + d_{u_i, u_j} \\ U &= \{u_0, u_1, \dots, u_{nt}\}, nt = \text{student on attendance} \\ E_{UR} &= \left\{ (u_i, u_j), Edge_{u_i, u_j} | u_i, u_j \in \text{NEDC range} \right\} \end{aligned} \quad (4)$$

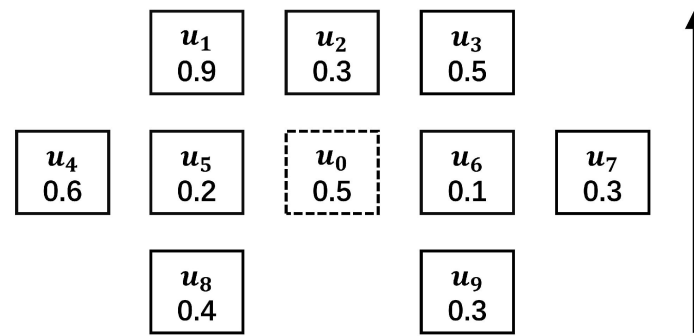


Figure 2. Effective distance of nearest neighbor relationship.

Here, d_{u_i, u_j} is the weight corresponding to the edge (u_i, u_j) . This is specified as the inverse $\frac{1}{\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}}$ of the Euclidean distance between the u_i and u_j seat coordinates (x_i, y_i) , (x_j, y_j) . The Euclidean distance directly calculates the straight-line distance between two points in a two-dimensional space. This straight-line distance is in line with the public's most intuitive understanding of distance, and this calculation is quite intuitive and easy to understand.

Based on a single-classroom network $G_{UR} = (U, E_{UR})$, the learner u_i combines the knowledge points k , learning engagement $L_{u_j, k}$ of the nearest neighbor influence nodes, and the inter-node edge weights $Edge_{u_i, u_j}$ within the lecture time. For the learners u_i , the learning engagement $Net_{u_i}^k$, based on the classroom network characteristics, is calculated as shown in the following equation:

$$Net_{u_i}^k = \frac{1}{nt+1} \left[\sum_{j=1}^{nt} \left(L_{u_j, k} \bullet Edge_{u_i, u_j} \right) + L_{u_i, k} \right] \quad (5)$$

where $E_{u_i, R} = \left\{ (u_i, u_j), E_{u_i, u_j} \mid u_j \in u_{0-NEDC}, nt = \text{number of NEDC ranges}, j \in nt \right\}$

Here, $E_{u_i, R}$ represents the set of nodes and edge weights within the nearest neighbor influence of the learner u_i , nt represents the number of nearest neighbor influencing nodes, and $L_{u_i, k}$ represents its own learning involvement.

4. CL-TSKT Model

The CL-TSKT model architecture is shown in Figure 3. There are two main sections: a knowledge point and learning engagement temporal feature extraction module based on parallel temporal attention GRU networks and a CNN-based knowledge point-knowledge point spatial feature extraction module.

The knowledge point test sequences and the corresponding classroom network characteristics of learning engagement are feature coded and fed into a parallel temporal attention GRU network. This is utilized to extract the temporal features of knowledge points and learning engagement. They are fused to obtain the knowledge point-learning engagement temporal characteristics and their associated attributes. Meanwhile, the CNN is used to extract the knowledge point-knowledge point spatial features. Considering the associative properties of knowledge point-knowledge points from a spatial perspective, the knowledge point-knowledge point spatial feature is fused with the knowledge point-learning engagement temporal feature. Finally, the fused temporal-spatial feature tensor is input into the fully connected layer for nonlinear mapping, and the knowledge tracing results (KTRs) of students' temporal knowledge points are obtained.

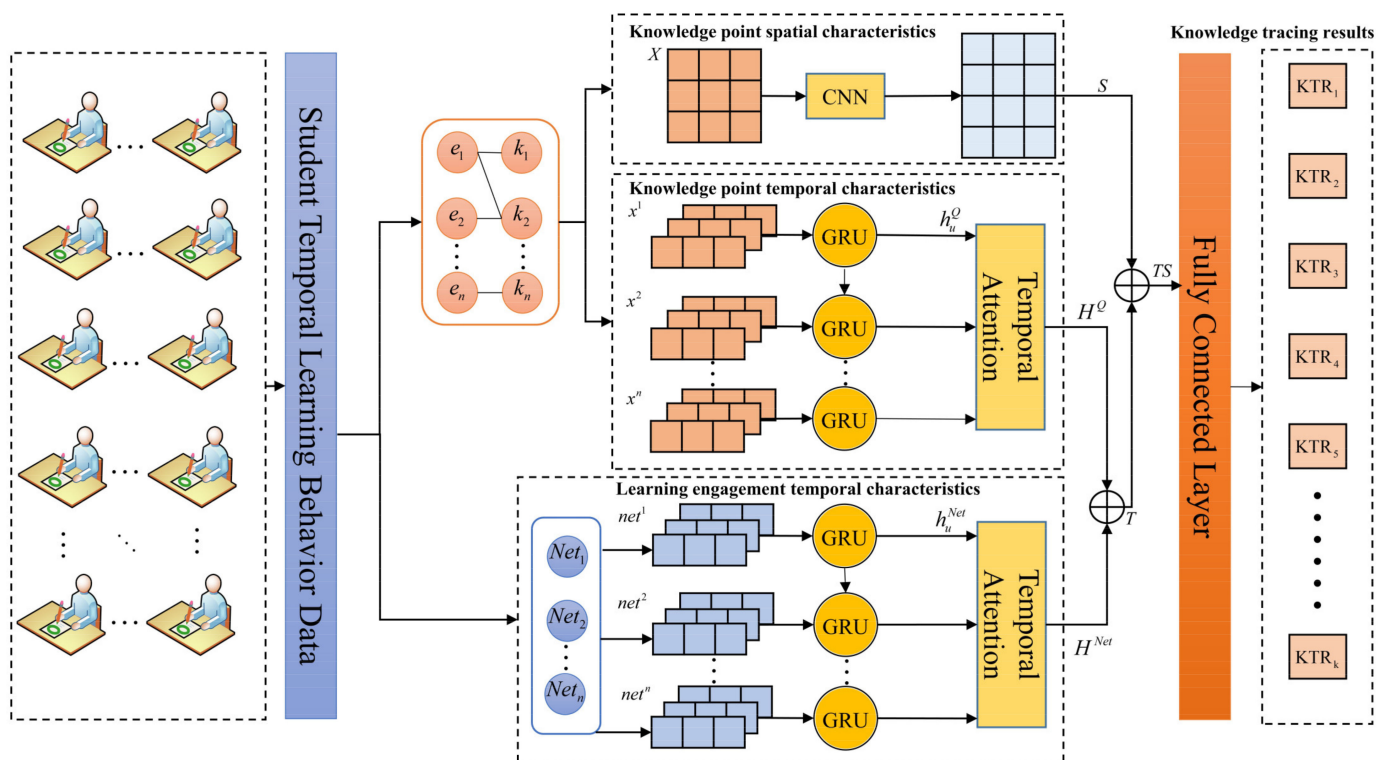


Figure 3. CL-TSKT model framework diagram.

4.1. Temporal Attention-Based GRU Feature Tracking

4.1.1. GRU Feature Tracking

Gated recurrent unit structures (GRUs) are variants of RNNs. They are mainly used for modeling time series data. They can effectively mitigate gradient disappearance and explosion during the training process compared with RNNs, and the effect is better than that of an RNN [34]. The GRU structure is shown in Figure 4.

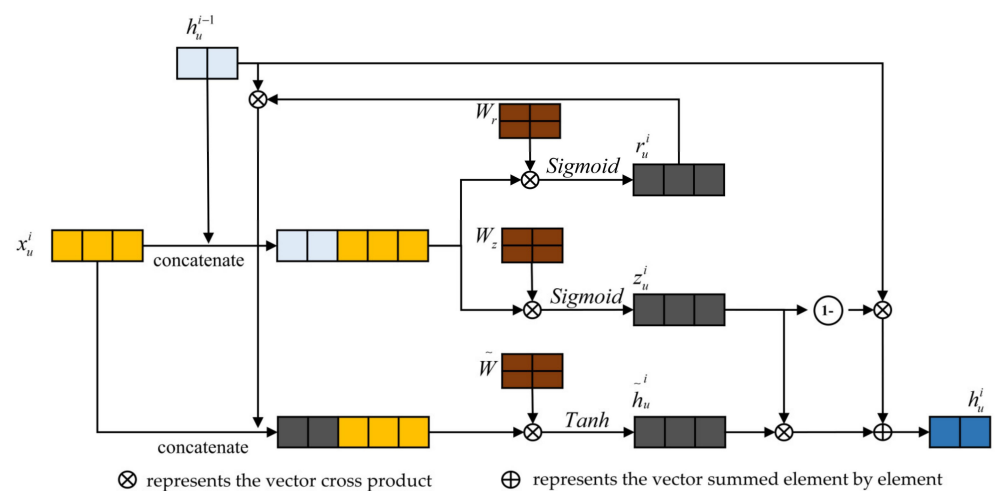


Figure 4. Structure of GRU.

In this paper, the student's record of questions is divided into sequences at intervals of a fixed length. If the student's record of questions is less than the length l , then the made up value is zero. At the same time, the student test sequences are one-hot coded before being entered into the GRUs. For example, the one-hot coding of the data for a particular sequence of knowledge test questions is x . The specific expression is shown in Equation (6):

$$x = \begin{pmatrix} e^1 \\ e^2 \\ \vdots \\ e^i \\ \vdots \\ e^l \end{pmatrix} = \begin{bmatrix} 0 & 1 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 1 & 1 & \ddots & 1 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 1 & \cdots & 1 & 0 \end{bmatrix}_{l \times (k \times 2)} \quad (6)$$

Here, e^i represents the vector of test questions with a serial number i taken by the learner in relation to the knowledge point, 0 represents that the test question does not relate to the knowledge point corresponding to this subscript, and 1 represents that it relates to the knowledge point corresponding to this subscript.

Meanwhile, the classroom network characteristic learning engagement input code Net corresponding to the sequence of knowledge point test questions is

$$Net = \begin{pmatrix} net^1 \\ net^2 \\ \vdots \\ net^i \\ \vdots \\ net^l \end{pmatrix} = \begin{bmatrix} 0.6 \\ 0.7 \\ \vdots \\ 0.5 \\ \vdots \\ 0.8 \end{bmatrix}_{l \times 1} \quad (7)$$

Here, net^i represents the classroom network characteristic learning engagement within the learning duration of the learner engaging with the test question with the serial number i for the knowledge point involved.

According to the characteristic coding of the above data, the input data are subjected to temporal feature extraction using a GRU. Here, the reset gate r_u^i and update gate z_u^i realize the selective forgetting of the previous moment information and selectively memorize the current moment information as the current moment output. The calculation is shown in Equation (8):

$$\begin{aligned} r_u^i &= \text{Sigmoid}(W_r \times [h_u^{i-1}, x_u^i]) \\ z_u^i &= \text{Sigmoid}(W_z \times [h_u^{i-1}, x_u^i]) \end{aligned} \quad (8)$$

The information r_u^i after the reset gate, the input information x_u^i , and the information h_u^{i-1} from the previous moment are concatenated, and the output is updated to obtain the updated feature vector \tilde{h}_u^i as shown in Equation (9):

$$\tilde{h}_u^i = \text{Tanh}(\tilde{W} \times [r_u^i \times h_u^{i-1}, x_u^i]) \quad (9)$$

Finally, the feature information after passing through forgetting and memorizing is selectively summed and fused to obtain the output state vector h_u^i as shown in Equation (10):

$$h_u^i = (1 - z_u^i) \times h_u^{i-1} + z_u^i \times \tilde{h}_u^i \quad (10)$$

The feature tensor h_u^Q and h_u^{Net} are obtained by tracking both the knowledge point test data features and the classroom network characteristics of the learning engagement features using the parallel temporal attention GRU network.

4.1.2. Temporal Attention Mechanism

In time series data prediction tasks, adding an attention mechanism helps the neural network model to focus on the information in the input that is more critical to the task at hand. It can effectively alleviate the problem of the GRU's loss of capture of long sequence information and insufficient extraction of short sequence feature information, and it improves the model's performance and robustness [35]. Therefore, a temporal attention mechanism is proposed in this paper to reassign weights to the temporal knowledge

points and learning engagement features output from the GRUs. This can better extract the feature information related to the current knowledge point and improve the model prediction accuracy.

The temporal knowledge point feature vector h_u^Q output from the GRU is calculated by the temporal attention mechanism as shown in Equation (11):

$$H^Q = \text{relu}\left(h_u^Q \bullet \text{Softmax}\left(\text{elu}\left(h_u^Q \times W_Q\right), \text{dim} = 1\right)\right) \quad (11)$$

In this case, the visualization of the temporal attention mechanism to reassign weights to the temporal knowledge point test data feature information is shown in Figure 5.

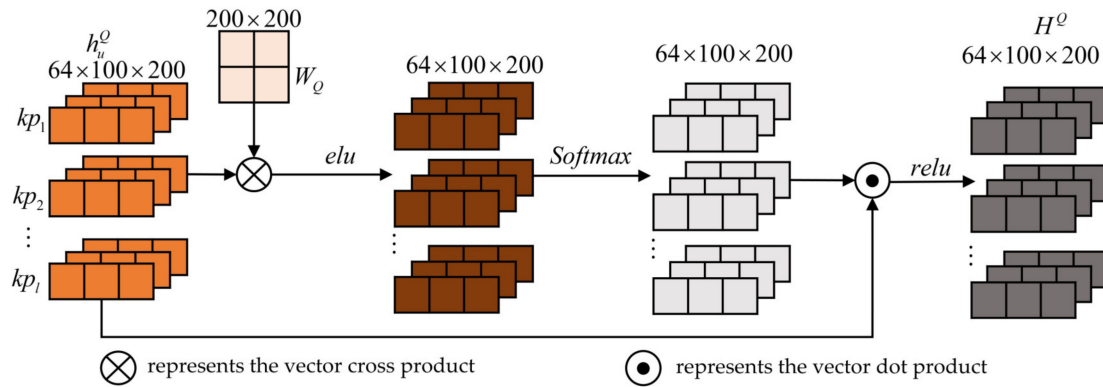


Figure 5. Visualization of the temporal attention mechanism for the weight reassignment of the knowledge point test feature information.

The learning engagement feature vector h_u^{Net} output from the GRU is calculated by the temporal attention mechanism as shown in Equation (12):

$$H^{Net} = \text{relu}\left(h_u^{Net} \bullet \text{Softmax}\left(\text{elu}\left(h_u^{Net} \times W_{Net}\right), \text{dim} = 1\right)\right) \quad (12)$$

Similarly, the visualization of the temporal attention mechanism to reassign weights to the temporal classroom network characteristic learning engagement feature information is shown in Figure 6.

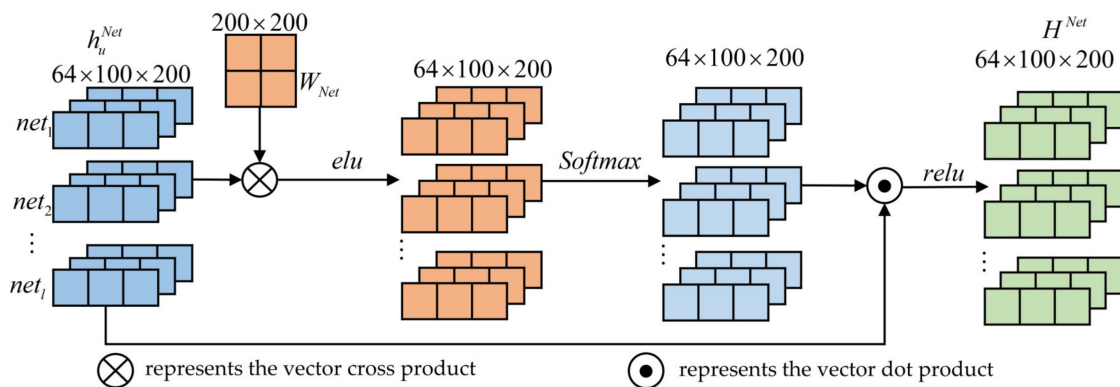


Figure 6. Visualization of temporal attention mechanism for weight reassignment of classroom network characteristic learning engagement feature information.

Finally, the temporal feature H^Q of the knowledge points is fused with the temporal feature H^{Net} of learning engagement to improve the information richness of the temporal characteristics. Thus, the knowledge point-learning engagement time feature tensor T is obtained. The calculation is shown in Equation (13):

$$T = H^Q + H^{Net} \quad (13)$$

4.2. CNN-Based Spatial Feature Extraction

Temporal attention-based GRU networks have been effective at extracting features from temporal data. However, when facing large datasets and time series data that are too long, the ability of the model to extract features decreases. This results in poor prediction performance. This paper is based on one-hot encoded student knowledge point test sequences x , using a CNN for spatial feature extraction of knowledge point time series test data. The connections and variations between successive knowledge points can be accessed through the CNN. To further enhance the model feature extraction capability and improve the prediction accuracy, the operation is expressed in the following equation:

$$\begin{aligned} S_1 &= \text{Conv1d}(x, \text{kernel_size} = 3) \\ S_2 &= \text{MaxPool1d}(S_1, \text{stride} = 2) \\ S_3 &= \text{Upsample}(S_2, \text{scale_factor} = 2) \\ S &= \text{Linear}(\text{transpose}(S_3, (1, 2))) \end{aligned} \quad (14)$$

In this paper, based on the one-hot coding matrix x of the knowledge point test data, spatial feature extraction of the knowledge point-knowledge points is performed using a CNN. The CNN-based spatial feature extraction visualization process is shown in Figure 7.

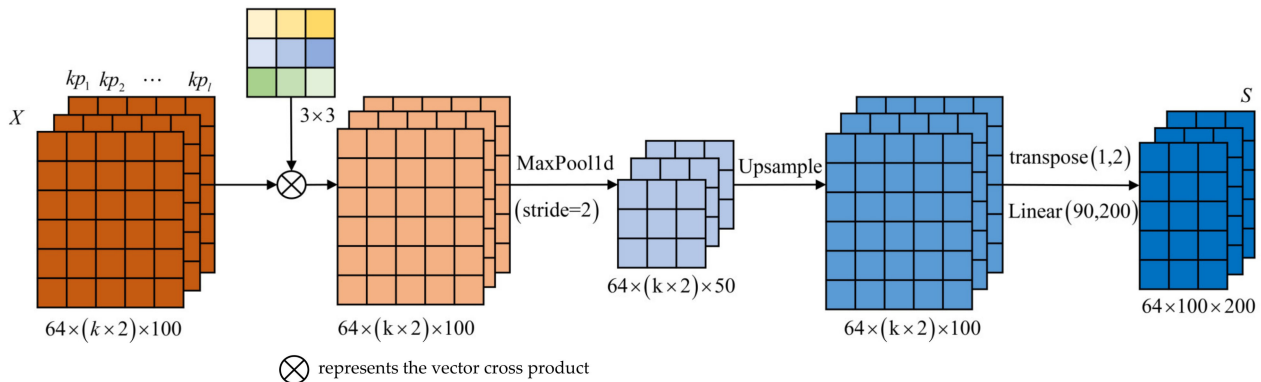


Figure 7. Visualization of CNN-based spatial feature extraction process.

Finally, this paper fully considers the attributes that link knowledge points to learning engagement, fusing the knowledge point-learning engagement temporal feature and knowledge point-knowledge point spatial feature S to enhance the correlation feature information between neighboring knowledge points, and thereby obtaining the final temporal-spatial feature vector TS :

$$TS = T + S \quad (15)$$

4.3. Nonlinear Mapping Based on Fully Connected Layers

The fully connected layer mainly takes the input data for feature extraction by linear transformation of the weight matrix. It converts the raw data into a higher-level representation, leading to better fitting of the training data by learning the complex nonlinear relationships between the input data. This enables the network to accurately classify and regress predictions for the input data [36]. This paper uses a single fully connected layer for nonlinear mapping, which is formulated as follows:

$$p = \text{Sigmoid}(W_1 \times TS + b_1) \quad (16)$$

Here, p represents the model prediction, W_1 represents the trainable weight matrix of the fully connected layer, b_1 represents the trainable bias of the fully connected layer, and TS represents the temporal-spatial eigenvectors.

The output layer targets the real answer situation of the learner. The model parameters are iteratively updated using the cross-entropy loss function, which is shown in Equation (17):

$$Loss = - \sum_{\substack{x_{lj} \in X \\ p_{lj} \in P}} \left[x_{lj} \log(p_{lj}) + (1 - x_{lj}) \log(1 - p_{lj}) \right] \quad (17)$$

Here, x_{ij} represents the actual answers to the j knowledge points involved in the exercise l , and p_{ij} represents the model's predicted value of the j knowledge points involved in the exercise question l .

Meanwhile, this paper takes the answer time as an interval. We synthesize the mastery of knowledge points in each answering moment and calculate the cognitive ability of the students as shown in Equation (18):

$$y = \frac{1}{k} \sum_{i=1}^k p_{li} \quad (18)$$

Here, y represents the student's overall cognitive ability at the time of the test l , k represents the number of knowledge points included in the course, and p_{ij} represents the diagnostic result for the knowledge point i at the exercise l in the model's prediction of the value p .

Algorithm 1 illustrates the basic steps of CL-TSKT:

Algorithm 1 CL-TSKT algorithm

Input: $G_{UR} = (U, E_{UR})$ = seat relationship network in smart classroom; U = set of students; E_{UR} = the set of edges with influential nodes; x = a matrix of student answer records with fixed length per row, which has the shape $l \times (k \times 2)$; L_{uk} = student learning engagement during knowledge point k learning time; T = max epoch.

Output: p = the set of predicted answer records; y = the cognitive diagnosis results.

- 1: Initialize learning rate and hyperparameter randomly;
 - 2: **For** $u \leftarrow 1$ to U **do**;
 - 3: According to the student's learning engagement L_{uk} and weights obtained by calculating the Euclidean distance, student–student interaction, and degree centrality to the nearest neighboring influence nodes E_{uR} is defined in Equations (2)–(4);
 - 4: Combine the network structure E_{uR} to obtain the classroom network characteristic learning engagement, which is defined in Equation (5);
 - 5: According to the matrix of test questions x , obtain the corresponding sequence of the classroom network characteristic learning engagement Net , which has the shape $l \times 1$;
 - 6: **end**;
 - 7: **for** $epoch \leftarrow 1$ to T **do** // The dataset contains the number of students;
 - 8: $h_u^Q \leftarrow GRU(x[n])$ // Track sequence state using GRU;
 - 9: $h_u^{Net} \leftarrow GRU(Net[n])$;
 - 10: // Reassign weights to feature information using temporal attention;
 - 11: $H^Q \leftarrow \text{relu}(h_u^Q \bullet \text{Softmax}(\text{elu}(h_u^Q \times W_Q), \text{dim} = 1))$;
 - 12: $H^{Net} \leftarrow \text{relu}(h_u^{Net} \bullet \text{Softmax}(\text{elu}(h_u^{Net} \times W_{Net}), \text{dim} = 1))$;
 - 13: $T \leftarrow H^Q + H^{Net}$ // Sum the feature tensor;
 - 14: $S \leftarrow \text{CNN}(x[n])$;
 - 15: $TS \leftarrow T + S$;
 - 16: $p \leftarrow \text{Sigmoid}(W_1 \times TS + b_1)$ // Nonlinear mapping output uses fully connected layer;
 - 17: $y \leftarrow \frac{1}{k} \sum_{i=1}^k p_{li}$;
-

Algorithm 1 *Cont.*

```

18: // Update hyperparameters using cross-entropy loss function;
19:  $Loss_i \leftarrow Loss_i - \sum_{\substack{x_{ij} \in X \\ p_{ij} \in P}} [x_{ij} \log(p_{ij}) + (1 - x_{ij}) \log(1 - p_{ij})]$ ;
20: end;
21: Return  $p, y$ .

```

5. Analysis of Experiments and Experimental Results**5.1. Datasets**

To validate the performance of the proposed CL-TSKT model, four sets of experiments on real datasets were conducted, namely the Assistment0910, ASSISTChall [37], and Eedi [38] datasets and the Smart Classroom Dataset (SCD). The Assistment0910 and ASSISTChall datasets were collected by the ASSISTments online tutoring system. The Eedi dataset was released by the NeuralPS2020 Education Challenge and contains a total of four sections. The SCD is an offline smart classroom dataset consisting of historical behavioral data of students who took the C programming course in 2022.

The Assistment0910, ASSISTChall, and Eedi datasets do not contain information about learners' classroom networks. In order to ensure the wholeness of the model, the dataset was preprocessed in this paper. To minimize the impact on the datasets, the learning engagement information component was supplemented with a value of zero in all three datasets. The statistics for the four datasets are shown in Table 2.

Table 2. Dataset statistics.

Dataset	Assistment0910	ASSISTChall	Eedi	SCD
Number of learners	4049	1709	4918	58
Number of concepts	110	102	948	37
Number of exercises	16,000	3000	948	45
Number of interactions	325,000	942,000	104,000	11,000
Average length	80	551	212	195
Engage data	No	No	No	Yes

5.2. Evaluation Metrics and Baseline Modeling

To evaluate the performance of the CL-TSKT model proposed in this paper, the AUC, ACC, MAE, and RMSE were used as evaluation metrics. Among them, the MAE and RMSE are a measure of the difference between the predicted probabilities and the actual labels, and they can reveal more subtle differences in the performance of the model, especially under different thresholds, which can effectively assess the model's ability to handle uncertainty.

The five baseline models for the comparison are shown in Table 3.

Table 3. Baseline models.

Baseline	Description
DKT+ [17]	A recurrent neural network (RNN) is used to track the state of students' knowledge, and two regularization terms are added to solve the DKT model of the reconstruction and the consistency problem.
AKT + Forgetting [18]	Uses attention mechanisms with exponential decay and context-aware relative distance metrics and embeds forgetting linear deviations.
QRCDM [19]	A quantitatively interpretable cognitive diagnostic model based on explicit correlations between test questions and knowledge concepts, with implicit correlations between test questions and irrelevant knowledge concepts.

Table 3. Cont.

Baseline	Description
CL4KT [24]	A KT model for a contrastive learning framework that reveals semantically similar or dissimilar variables.
CL4KT-FoLiBi [25]	A KT model incorporating forgotten linear bias (FoLiBi) into CL4KT.

5.3. Experimental Environment and Model Parameters

The experimental environment of this paper is shown in Table 4.

Table 4. Experimental environments.

Experimental Environment	Environment Configuration
Operating Systems	Linux
CPU	Intel(R) Xeon(R) Gold 6330H
Video Cards	GeForce RTX 3090
RAM	32 GB
ROM	1T SSD
Programming Languages	Python 3.7
Framework	Pytorch

In the CL-TSKT cognitive diagnostic model, the learning rate was set to 0.001, the input feature sequence length l was set to 100, the number of one-sided GRU layers was set to 1, the hidden layer embedding dimensionality was set to 200, the CNN convolutional kernel size was set to 3, and the batch size of the model was set to 64. The gradient descent optimization was carried out using the Adam optimizer.

5.4. Results

The ROC curves for each model on the Assistment0910, ASSISTChall, Eedi, and SCD datasets are shown in Figure 8. The ROC curves for each model facing the online dataset using only the knowledge test question features are shown in Figure 8a–c. The analysis yielded that the AUC values of the CL-TSKT model using the knowledge point test features in the public dataset were all better than those of the baseline model. Figure 8d represents the diagnostic performance of the model on the SCD. The experimental results show that the CL-TSKT model incorporating learners' classroom network characteristic learning engagement demonstrated better performance.

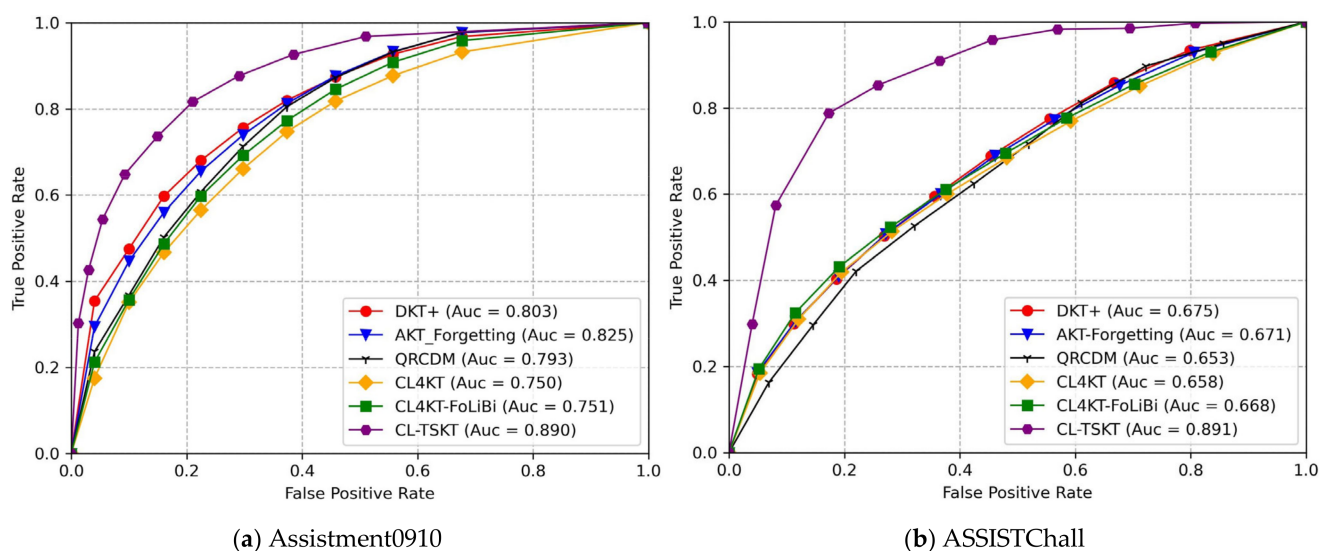


Figure 8. Cont.

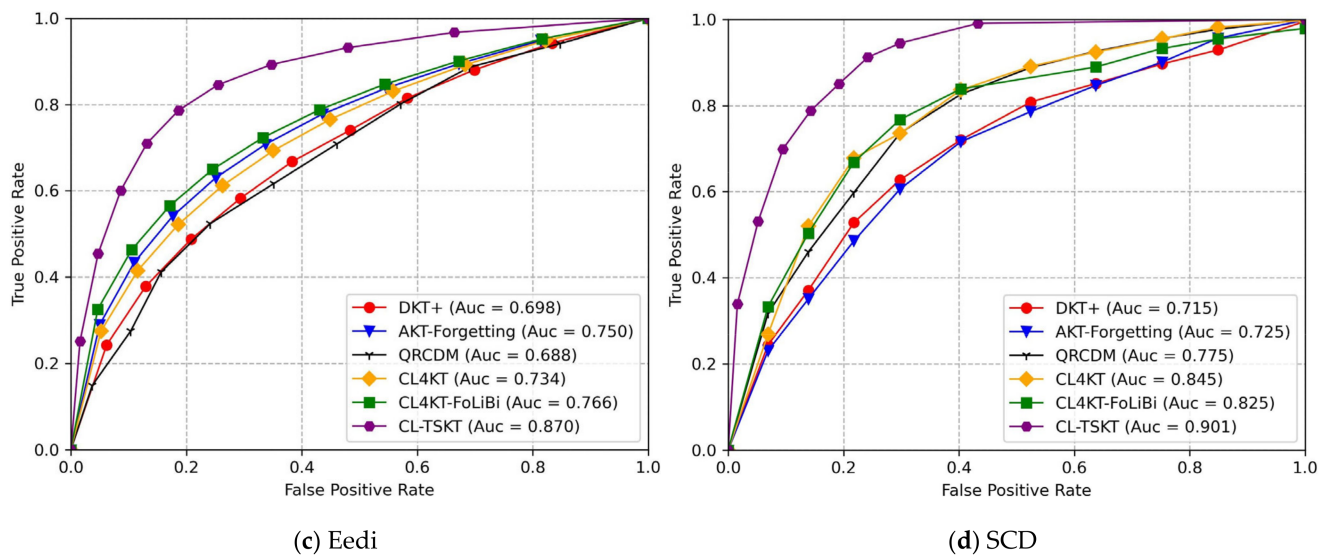


Figure 8. ROC curves for the CL-TSKT model and the baseline model on the dataset. (a–d) The Assistment0910, ASSISTChall, Eedi, and SCD datasets, respectively.

The performance metrics of the CL-TSKT model on the Assistment0910, ASSISTChall, Eedi, and SCD datasets versus the other baseline models are shown in Table 5 and analyzed below.

Table 5. Performance metrics of all models on the dataset.

Dataset	Model	AUC	ACC	MAE	RMSE
Assistment0910	DKT+	0.803	0.772	0.227	0.477
	AKT + Forgetting	0.825	0.773	0.226	0.476
	QRCDM	0.793	0.748	0.252	0.502
	CL4KT	0.750	0.715	0.285	0.441
	CL4KT-FoLiBi	0.751	0.712	0.287	0.437
	CL-TSKT	0.890	0.805	0.194	0.437
ASSISTChall	DKT+	0.675	0.665	0.334	0.578
	AKT + Forgetting	0.671	0.673	0.326	0.571
	QRCDM	0.653	0.619	0.381	0.617
	CL4KT	0.658	0.641	0.358	0.476
	CL4KT-FoLiBi	0.668	0.659	0.341	0.468
	CL-TSKT	0.891	0.809	0.190	0.436
Eedi	DKT+	0.698	0.648	0.351	0.593
	AKT + Forgetting	0.750	0.688	0.312	0.450
	QRCDM	0.688	0.635	0.364	0.603
	CL4KT	0.734	0.673	0.326	0.457
	CL4KT-FoLiBi	0.766	0.698	0.301	0.443
	CL-TSKT	0.870	0.799	0.200	0.447
SCD	DKT+	0.715	0.736	0.264	0.514
	AKT + Forgetting	0.725	0.753	0.246	0.496
	QRCDM	0.776	0.780	0.220	0.469
	CL4KT	0.845	0.845	0.154	0.340
	CL4KT-FoLiBi	0.825	0.834	0.165	0.354
	CL-TSKT	0.901	0.896	0.103	0.321

- (1) The method proposed in this paper achieved the best performance on all four datasets. The evaluation metrics were all better than those of the baseline model on the SCD. It is shown that integrating classroom network characteristic learning engagement in smart classrooms can characterize students' cognitive states more accurately.
- (2) Compared with the RNN-based DKT model and AKT + Forgetting based on contextual attention mechanism, CL-TSKT showed significant improvement. It proved

the superiority of computing accumulation and forgetting based on GRU dynamic gate control.

- (3) CL4KT mainly uses a contrastive learning framework for knowledge tracing, and QRCDM mainly utilizes the α cross-validation idea for feature extraction. CL-TSKT enhances feature extraction by using the temporal attention mechanism, and its obtained results outperformed CL4KT and QRCDM. This proves that the temporal attention mechanism can pay better attention to the important feature information and enhance the model prediction accuracy.
- (4) CL4KT-FoLiBi embedded with the forgetting linear deviation mechanism simulated students' forgetting behavior, but it ignored that students' knowledge processes are ambiguous and complex. In contrast, CL-TSKT started from the aspect of enhanced feature extraction, adopted a CNN to realize spatial feature extraction, strengthened the model feature extraction ability, and achieved better results.
- (5) In Table 5, the CL-TSKT model shows superior performance on all three online platform datasets and one smart classroom dataset. The distributions and quantities of the four datasets are different, as a demonstration of the greater robustness of the CL-TSKT model.
- (6) The poor performance of the RMSE value for CL-TSKT on the Eedi dataset was due to the high number of knowledge points contained in this dataset.

5.5. Ablation Experiment

5.5.1. Ablation Experiments Based on the CNN's Spatial Features and Temporal Attention Mechanisms

To verify that the CNN-based spatial features and temporal attention mechanism can effectively enhance the model to capture the learner's feature information and improve the model's prediction ability, in this paper, ablation experiments were performed on four datasets, where CL-TSKT-TA represents the model without the addition of the temporal attention mechanism and CL-TSKT-CNN represents the model without adding CNN-based spatial features. The experimental results are shown in Figure 9.

From Figure 9 and the analysis of the experimental results, it can be seen that CL-TSKT outperformed CL-TSKT-TA and CL-TSKT-CNN on all four datasets. This proves that CL-TSKT has better stability and robustness.

Therefore, a temporal attention mechanism was added to the GRU. Meanwhile, adding CNN-based spatial features can enhance the model feature extraction ability. This ensures that the integrity of feature information can more effectively improve the stability and prediction ability of the model.

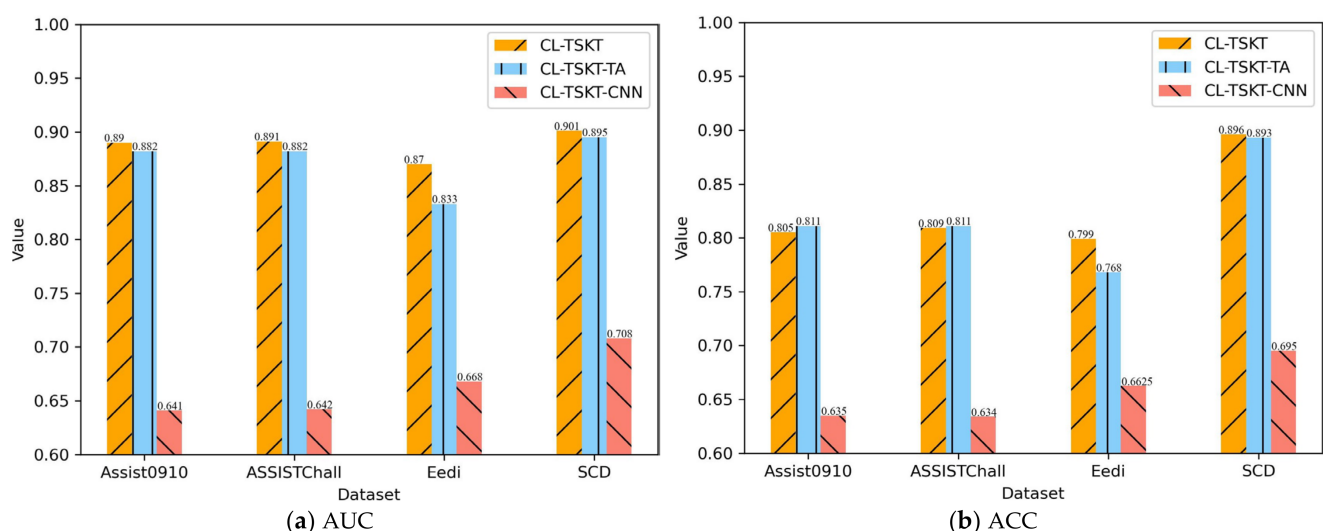


Figure 9. Cont.

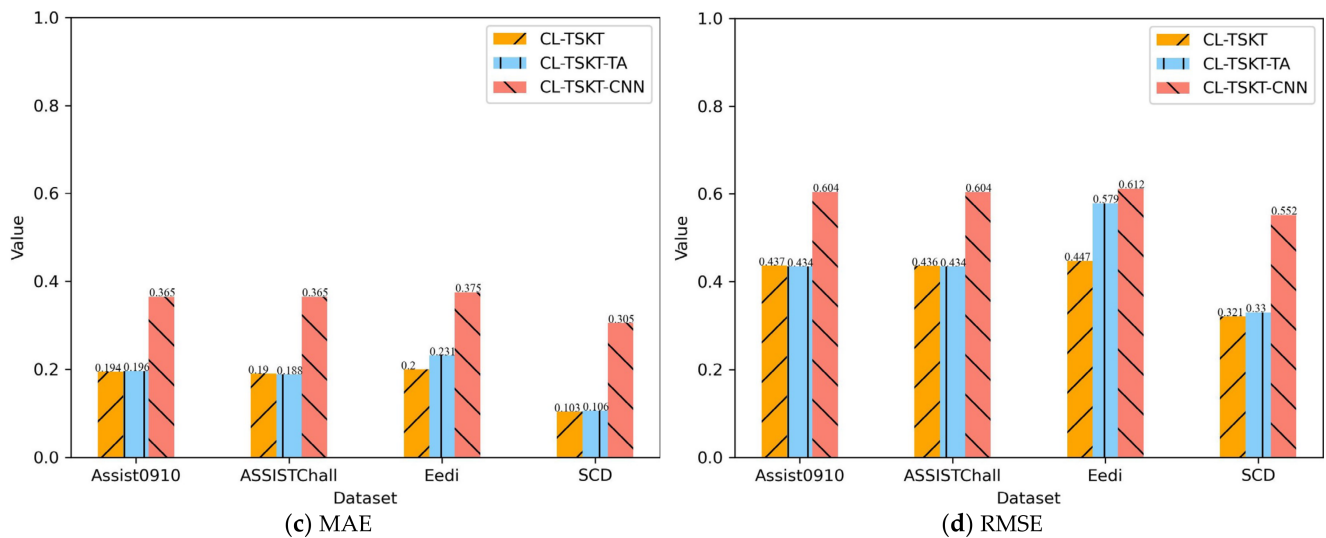


Figure 9. Performance metrics of CL-TSKT, CL-TSKT-TA, and CL-TSKT-CNN models on four datasets. (a–d) The AUC, ACC, MAE, and RMSE values of the models with or without the temporal attention mechanism, respectively.

5.5.2. Ablation Experiments of Classroom Network Characteristic Learning Engagement

To illustrate that incorporating learners' classroom network characteristic-based learning engagement (CL-TSKT) can lead to better prediction performance in smart classrooms with complex environments, in this paper, ablation experiments were performed on the SCD. The experimental results are shown in Figure 10. TSKT represents the model that did not incorporate the learner's learning engagement, and the Net part of the input was zero. L-TSKT represents the model that incorporated the individual's learning engagement and which did not take into account the influence of the surrounding students, and the Net part of the data was obtained directly from the calculation of Equation (1). L-TSKT-ED represents the model that incorporated classroom network Euclidean distance-based learning engagement. This learning engagement only took into account the influence generated by the distance between students, ($Edge_{u_i, u_j} = d_{u_i, u_j}$) in Equation (3) without considering student–student interactions and degree centrality.

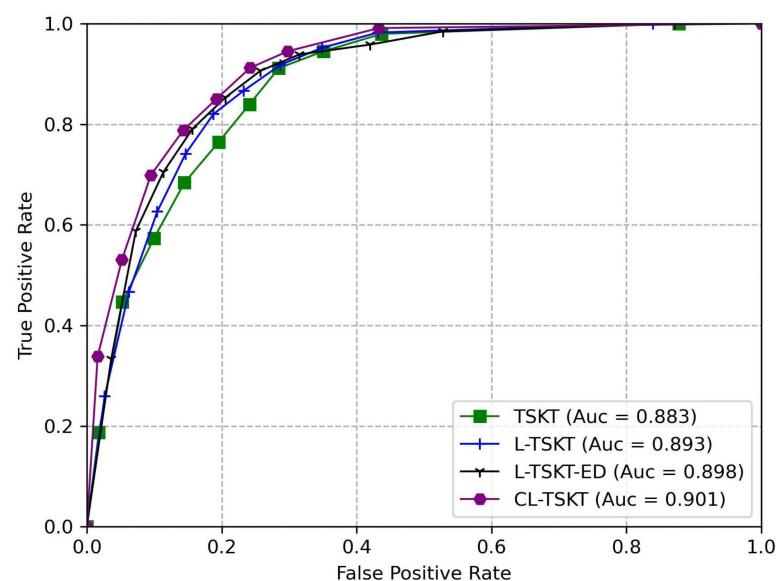


Figure 10. Comparison of ROC curves of ablation experiments for the CL-TSKT model on the SCD.

The performance metrics of the models on the SCD using different learning engagement are shown in Table 6.

Table 6. Comparison of model performance metrics based on different learning engagement.

Dataset	Model	AUC	ACC	MAE	RMSE
SCD	TSKT	0.883	0.883	0.116	0.341
	L-TSKT	0.893	0.886	0.113	0.337
	L-TSKT-ED	0.898	0.878	0.121	0.348
	CL-TSKT	0.901	0.896	0.103	0.321

Meanwhile, this paper provides experimental validation of the test data of some students. The results of the comparison of the accuracy of the KT results output from the TSKT, L-TSKT, and L-TSKT-ED models are shown in Table 7.

Table 7. Comparison of the accuracies of the TSKT, L-TSKT, and CL-TSKT-ED models.

ID	TSKT	L-TSKT	L-TSKT-ED	CL-TSKT
123	89.0%	89.1%	89.0%	89.9%
127	92.3%	92.4%	92.4%	93.9%
134	88.2%	88.3%	88.4%	92.5%
214	89.8%	89.7%	89.9%	92.3%
216	77.9%	78.3%	78.4%	84.3%
224	86.7%	86.7%	86.8%	92.0%

From the analysis of Figure 10 and Tables 6 and 7, it can be seen that the CL-TSKT model incorporating classroom network features outperformed the performance results of the models that did not incorporate the learning engagement features or the learner's learning engagement features and the model based on the classroom network structure of Euclidean distance learning engagement. The overall results suggest that in a smart classroom with multiple learners, a single dimension of exercise characteristics cannot accurately predict student performance. In addition, students' learning engagement not only depends on themselves but is also influenced by the learning engagement of their immediate neighbors based on the network of seating relationships, the distance between students, and the interactions between students. These influences play a direct role in students' mastery of knowledge points as they learn them and are ultimately reflected in the results of the knowledge point tests.

5.6. Analysis of Results

To validate the KT results of the knowledge points predicted by the model, this paper used learners' classroom test scores as the cognitive results for classification validation on the SCD. The KT results of some students' knowledge points predicted by the CL-TSKT model are shown in Table 8.

Table 8. Cognitive abilities of learners within each knowledge point's lecture time.

ID	kp ₁	kp ₂	kp ₃	...	kp _{k-1}	kp _k
123	0.856	0.743	0.603	...	0.718	0.766
127	0.801	0.801	0.519	...	0.631	0.931
134	0.811	0.867	0.829	...	0.939	0.818
214	0.856	0.816	0.596	...	0.681	0.749
216	0.788	0.365	0.565	...	0.794	0.736
224	0.855	0.850	0.832	...	0.754	0.773

To assess the reliability of the KT results predicted by the model, we used a pyramidal hierarchy of cognitive states (remembering, comprehending, applying, analyzing, evaluat-

ing, and innovating) based on Bloom's [39] proposed method. According to the correlation between test scores and cognitive levels proposed in [40–42], the learners were graded according to the KT results at intervals of course learning knowledge points of memory: $F \in [0, 40)$, comprehension: $E \in [40, 65)$, application: $D \in [65, 80)$, analysis: $C \in [80, 90)$, evaluation: $B \in [90, 97)$, and innovation: $A \in [97, 100]$, here, '[]' represents a left-closed right-open interval. '[]' represents left-closed right-closed interval. The result is shown in Table 9.

Table 9. Learners' cognitive states within the lecture time of each knowledge point.

ID	kp ₁	kp ₂	kp ₃	...	kp _{k-1}	kp _k
123	C	D	E	...	D	D
127	C	C	E	...	E	B
134	C	C	C	...	B	C
214	C	C	E	...	D	D
216	D	F	E	...	D	D
224	C	C	C	...	D	D

Finally, the KT results for these six students were combined with the results in Table 9 and compared to each student's final test scores for the course, as shown in Table 10.

Table 10. Students' comprehensive cognitive states and assessment scores.

ID	Cognitive Grade	Test Performance	Test Grade
123	D	79	D
127	D	78	D
134	C	82	C
214	D	65	D
216	E	43	E
224	D	80	D

Classroom tests are effective in assessing students' cognitive states, which are correlated with student learning outcomes. While students are in an environment of active learning engagement, they are more cognizant of their knowledge, and their assessment scores are relatively higher. According to Table 10, the results predicted by the KT model proposed in this paper corresponded to their test scores. The students with higher assessment scores were basically at a higher level of cognition. This shows that the proposed KT model of cognitive states has good reliability.

5.7. Discussion

Our proposed knowledge tracing model can be used as a complete teaching module of the smart classroom teaching platform to help teachers improve the design of the teaching process and realize personalized education. By observing the students' knowledge tracking results, teachers can gain a deeper understanding of the students' current states of knowledge and then adjust their teaching strategies. The students can understand what they have not mastered well based on the knowledge tracking results to clarify their learning goals and improve their learning results.

The above experimental results show that classroom network characteristic learning engagement is of great significance for realizing knowledge tracking in the smart classroom scenario. Similarly, in the field of online education, student learning engagement can be obtained by collecting information such as the frequency and length of interaction between students and knowledge points, which can effectively enhance the model's tracing of students' knowledge statuses. Therefore, the model can also be considered for application in online education platforms to track students' knowledge and provide stronger interpretability for the teaching recommendation system based on the knowledge tracing results.

6. Conclusions and Future Work

In this paper, a KT model based on classroom network characteristic learning engagement and temporal-spatial feature fusion (CL-TSKT) was proposed. It aims to address the existence of student–student learning engagement impacts in smart classrooms with complex environments, as well as the existing KT models that are mainly based on knowledge point test data for cognitive state analysis. The experimental results showed that the prediction accuracy of the converged classroom network characteristic learning engagement model (CL-TSKT) was better than that of the counterpart models in the ablation experiment (TSKT, L-TSKT, and CL-TSKT-ED). It was proven that students' classroom learning engagement can be more accurately captured by fully considering the effects of students' network characteristics, inter-student distance, and inter-student interactions in the classroom. More importantly, parallel temporal attention GRU networks were proposed for knowledge point-learning engagement temporal feature extraction and CNN-based knowledge point-knowledge point spatial feature extraction. The temporal-spatial feature fusion helped to maintain the integrity of the feature information, which further improved the accuracy of the model prediction. The experimental results in this paper on four real datasets show that the proposed CL-TSKT model exhibited better diagnostic performance compared with the five baseline models (DKT+, AKT + Forgetting, QRCDM, CL4KT, and CL4KT-FoLiBi) under different conditions.

The CL-TSKT model can accurately characterize the cognitive states of students in smart classrooms with complex environments. Meanwhile, the design of the educational platform system can be enhanced. In this paper, students' learning engagement was accurately characterized by fully considering student–student interaction, inter-student distance, and students' degree of centrality, combining student knowledge point test data to characterize their cognitive states. It can support teachers in making timely adjustments to student seating during the instructional process and sustainable interventions to improve students' attitudes and learning engagement, thereby contributing to the effectiveness and enhancement of student learning. This research was applied to offline smart classrooms. Existing image processing techniques applied to a smart classroom in a large scene have problems such as difficulty recognizing students in the back row. Therefore, the classroom network structure data in this study cannot be fully automated, and it is necessary to add a manual proofreading link, which increases the workload by a certain amount.

In future research, the effects of students' different classroom social relationships on student cognition will be considered so as to achieve a more accurate characterization of students' cognitive states and to improve the predictive performance of the model. In addition, this paper plans to deploy the algorithm into embedded devices for smart classrooms to provide effective support for teachers' sustainable interventions for learners in the teaching and learning process and thus sustainably enhance student learning outcomes.

Author Contributions: Conceptualization, Z.S. and Y.L.; methodology, Y.L.; software, Y.L.; validation, D.L., J.M. and H.Z.; formal analysis, D.L.; investigation, D.L.; resources, Z.S.; data curation, Y.L.; writing—original draft preparation, Y.L.; writing—review and editing, Z.S.; visualization, Y.L.; supervision, Z.S.; project administration, Z.S.; funding acquisition, Z.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by The National Natural Science Foundation of China (62177012, 61967005, and 62267003) and the Project of Guangxi Wireless Broadband Communication and Signal Processing Key Laboratory (GXKL06240107).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets are available at the following links: Assistentment0910 (<https://edudata.readthedocs.io/en/latest/build/blitz/ASSISTments/ASSISTments2009-2010.html> (accessed on 1 March 2024)); ASSISTChall (<https://edudata.readthedocs.io/en/latest/build/blitz/ASSISTments/ASSISTments2017.html> (accessed on 1 March 2024)); and Eedi (<https://eedi.com/>

[projects/neurips-education-challenge](#) (accessed on 1 March 2024)). The SCD dataset will be made available upon reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Janssen, R.; Tuerlinckx, F.; Meulders, M.; De Boeck, P. A hierarchical IRT model for criterion-referenced measurement. *J. Educ. Behav. Stat.* **2000**, *25*, 285–306. [CrossRef]
- Li, L.; Wang, Z. Knowledge Relation Rank Enhanced Heterogeneous Learning Interaction Modeling for Neural Graph Forgetting Knowledge Tracing. *PLoS ONE* **2023**, *18*, e0295808. [CrossRef]
- Shi, Y.; Chen, L.; Qu, Z.; Xu, J.; Yang, H.H. Study on the Influencing Factors of Junior High School Students' Learning Engagement Under the Smart Classroom Environment. In Proceedings of the International Conference on Blended Learning, Hong Kong, China, 17–20 July 2023; Springer Nature: Cham, Switzerland, 2023; pp. 47–58.
- Jiaming, N. A Study on the Learning Engagement Status and Influencing Factors of University Students in a Smart Classroom Environment: A Case Study of a Smart Classroom at Baise University. 2023. Available online: http://dspace.bu.ac.th/jspui/bitstream/123456789/5178/1/nong_jiam.pdf (accessed on 29 February 2024).
- You, W. Research on the relationship between learning engagement and learning completion of online learning students. *Int. J. Emerg. Technol. Learn.* **2022**, *17*, 102–117. [CrossRef]
- Hu, M.; Wei, Y.; Li, M.; Yao, H.; Deng, W.; Tong, M.; Liu, Q. Bimodal learning engagement recognition from videos in the classroom. *Sensors* **2022**, *22*, 5932. [CrossRef] [PubMed]
- Lu, G.; Liu, Q.; Xie, K.; Zhang, C.; He, X.; Shi, Y. Does the Seat Matter? The Influence of Seating Factors and Motivational Factors on Situational Engagement and Satisfaction in the Smart Classroom. *Sustainability* **2023**, *15*, 16393. [CrossRef]
- Putnik, G.; Costa, E.; Alves, C.; Castro, H.; Varela, L.; Shah, V. Analysing the correlation between social network analysis measures and performance of students in social network-based engineering education. *Int. J. Technol. Des. Educ.* **2016**, *26*, 413–437. [CrossRef]
- Xiang, T.; Ji, H.; Sheng, J. Analysis of Spatiotemporal Characteristics of Student Concentration Based on Emotion Evolution. *Adv. Comput. Signals Syst.* **2023**, *7*, 89–102.
- Van Rijsewijk, L.G.M.; Oldenburg, B.; Snijders, T.A.B.; Dijkstra, J.K.; Veenstra, R. A description of classroom help networks, individual network position, and their associations with academic achievement. *PLoS ONE* **2018**, *13*, e0208173. [CrossRef] [PubMed]
- Gutierrez, A. The Effects of Various Classroom Seating Arrangements on English Learners' Academic Achievement. 2022. Available online: <https://neiudc.neiu.edu/uhp-projects/31> (accessed on 12 May 2022).
- Li, J.; Shi, D.; Tummark, P.; Xu, H. A system for real-time intervention in negative emotional contagion in a smart classroom deployed under edge computing service infrastructure. *Peer Peer Netw. Appl.* **2020**, *13*, 1706–1719. [CrossRef]
- De La Torre, J. DINA model and parameter estimation: A didactic. *J. Educ. Behav. Stat.* **2009**, *34*, 115–130. [CrossRef]
- Chiu, C.Y. Statistical refinement of the Q-matrix in cognitive diagnosis. *Appl. Psychol. Meas.* **2013**, *37*, 598–618. [CrossRef]
- Gu, Y.; Xu, G. The sufficient and necessary condition for the identifiability and estimability of the DINA model. *Psychometrika* **2019**, *84*, 468–483. [CrossRef]
- Ma, W.; de la Torre, J. An empirical Q-matrix validation method for the sequential generalized DINA model. *Br. J. Math. Stat. Psychol.* **2020**, *73*, 142–163. [CrossRef]
- Piech, C.; Bassen, J.; Huang, J.; Ganguli, S.; Sahami, M.; Guibas, L.J.; Sohl-Dickstein, J. Deep knowledge tracing. In Proceedings of the 28th International Conference on Neural Information Processing Systems, Montreal, QC, Canada, 7–12 December 2015; Volume 1, pp. 505–513.
- Ghosh, A.; Heffernan, N.; Lan, A.S. Context-aware attentive knowledge tracing. In Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Online, 6–10 July 2020; pp. 2330–2339.
- Yang, H.; Qi, T.; Li, J.; Guo, L.; Ren, M.; Zhang, L.; Wang, X. A novel quantitative relationship neural network for explainable cognitive diagnosis model. *Knowl. Based Syst.* **2022**, *250*, 109156. [CrossRef]
- Xiao, Y.; Xiao, R.; Huang, N.; Hu, Y.; Li, H.; Sun, B. Knowledge tracing based on multi-feature fusion. *Neural Comput. Appl.* **2023**, *35*, 1819–1833. [CrossRef]
- Liu, H.; Zhang, T.; Li, F.; Yu, M.; Yu, G. A probabilistic generative model for tracking multi-knowledge concept mastery probability. *Front. Comput. Sci.* **2024**, *18*, 183602. [CrossRef]
- Huang, C.; Wei, H.; Huang, Q.; Jiang, F.; Han, Z.; Huang, X. Learning consistent representations with temporal and causal enhancement for knowledge tracing. *Expert Syst. Appl.* **2024**, *245*, 123128. [CrossRef]
- Jiang, B.; Wei, Y.; Zhang, T.; Zhang, W. Improving the performance and explainability of knowledge tracing via Markov blanket. *Inf. Process. Manag.* **2024**, *61*, 103620. [CrossRef]
- Lee, W.; Chun, J.; Lee, Y.; Park, K.; Park, S. Contrastive learning for knowledge tracing. In Proceedings of the ACM Web Conference, Online, 25–29 April 2022; pp. 2330–2338.
- Im, Y.; Choi, E.; Kook, H.; Lee, J. Forgetting-aware Linear Bias for Attentive Knowledge Tracing. In Proceedings of the 32nd ACM International Conference on Information and Knowledge Management, Birmingham, UK, 21–25 October 2023; pp. 3958–3962.

26. Ni, Q.; Wei, T.; Zhao, J.; He, L.; Zheng, C. HHSKT: A learner–question interactions based heterogeneous graph neural network model for knowledge tracing. *Expert Syst. Appl.* **2023**, *215*, 119334. [\[CrossRef\]](#)
27. Wu, Z.; Huang, L.; Huang, Q.; Huang, C.; Tang, Y. SGKT: Session graph-based knowledge tracing for student performance prediction. *Expert Syst. Appl.* **2022**, *206*, 117681. [\[CrossRef\]](#)
28. Huang, Q.; Su, W.; Sun, Y.; Huang, T.; Shi, J. NTM-based skill-aware knowledge tracing for conjunctive skills. *Comput. Intell. Neurosci.* **2022**, *2022*, 9153697. [\[CrossRef\]](#) [\[PubMed\]](#)
29. Zhao, S.; Sahebi, S. Graph-Enhanced Multi-Activity Knowledge Tracing. In Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Turin, Italy, 18–22 September; Springer Nature: Cham, Switzerland, 2023; pp. 529–546.
30. Shou, Z.; Tang, M.; Wen, H.; Liu, J.; Mo, J.; Zhang, H. Key Student Nodes Mining in the In-Class Social Network Based on Combined Weighted GRA-TOPSIS Method. *Int. J. Inf. Commun. Technol. Educ. (IJICTE)* **2023**, *19*, 1–19. [\[CrossRef\]](#)
31. Rani, S.; Kumar, M. Influential Node Detection and Ranking with Fusion of Heterogeneous Social Media Information. *IEEE Trans. Comput. Soc. Syst.* **2022**, *10*, 1852–1874. [\[CrossRef\]](#)
32. Bloch, F.; Jackson, M.O.; Tebaldi, P. Centrality measures in networks. *Soc. Choice Welf.* **2023**, *61*, 413–453. [\[CrossRef\]](#)
33. Shang, Q.; Zhang, B.; Li, H.; Deng, Y. Identifying influential nodes: A new method based on network efficiency of edge weight updating. *Chaos Interdiscip. J. Nonlinear Sci.* **2021**, *31*, 033120. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Yamak, P.T.; Yujian, L.; Gadosey, P.K. A comparison between arima, lstm, and gru for time series forecasting. In Proceedings of the 2019 2nd International Conference on Algorithms, Computing and Artificial Intelligence, Sanya, China, 20–22 December 2019; pp. 49–55.
35. Brauwiers, G.; Frasincar, F. A general survey on attention mechanisms in deep learning. *IEEE Trans. Knowl. Data Eng.* **2021**, *35*, 3279–3298. [\[CrossRef\]](#)
36. Basha, S.S.; Dubey, S.R.; Pulabaigari, V.; Mukherjee, S. Impact of fully connected layers on performance of convolutional neural networks for image classification. *Neurocomputing* **2020**, *378*, 112–119. [\[CrossRef\]](#)
37. Feng, M.; Heffernan, N.T. Informing teachers live about student learning: Reporting in the assistent system. *Technol. Instr. Cogn. Learn.* **2006**, *3*, 63.
38. Wang, Z.; Lamb, A.; Saveliev, E.; Cameron, P.; Zaykov, J.; Hernandez-Lobato, J.M.; Turner, R.E.; Baraniuk, R.S.; Barton, C.; Peyton, S.; et al. Results and insights from diagnostic questions: The NeurIPS 2020 education challenge. In Proceedings of the NeurIPS 2020 Competition and Demonstration Track, PMLR, Online, 9–12 December 2020; pp. 191–205.
39. Rahayu, A. The analysis of students’ cognitive ability based on assesments of the revised Bloom’s Taxonomy on statistic materials. *Eur. J. Multidiscip. Stud.* **2018**, *3*, 80–85. [\[CrossRef\]](#)
40. Nesayan, A.; Amani, M.; Gandomani, R.A. Cognitive profile of children and its relationship with academic performance. *Basic Clin. Neurosci.* **2019**, *10*, 165. [\[CrossRef\]](#)
41. Tikhomirova, T.; Malykh, A.; Malykh, S. Predicting academic achievement with cognitive abilities: Cross-sectional study across school education. *Behav. Sci.* **2020**, *10*, 158. [\[CrossRef\]](#)
42. Peng, P.; Kievit, R.A. The development of academic achievement and cognitive abilities: A bidirectional perspective. *Child Dev. Perspect.* **2020**, *14*, 15–20. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.