

Student Course Recommendation System Using Semantic Aware Temporal Gating With Time Transformer

Joshi Vinay Kumar¹, Dahlia Sam² and N. Kanya²

¹Department of Computer Science and Engineering, Dr. M.G.R Educational and Research Institute, Chennai, India

²Department of Information Technology, Dr. M.G.R Educational and Research Institute, Chennai, India

Article history

Received: 18-03-2026

Revised: 09-06-2026

Accepted: 22-09-2026

Corresponding Author:

Joshi Vinay Kumar

Department of Computer
Science and Engineering, Dr.
M.G.R Educational and
Research Institute, Chennai,
India

Email:

joshivinay0506@gmail.com

Abstract: Course recommendation system works by analyzing users' interests and recognizes their historical behavior such as course enrollment and previously watched videos to recommend relevant courses. However, the existing course recommendation systems in Massive Open Online Courses (MOOCs) face challenges in capturing complex semantic information, which impacts student requirements with an unsatisfied recommendation. By combining historical behaviors without temporal awareness, the model does not provide a proper recommendation and it lacks sequential dependencies. This research proposes Semantic Aware Temporal Gating (SATG) with Time Transformer (TT), which dynamically weights historical learning interactions by cross-referencing elapsed time with content-based Knowledge Graph (KG) semantics. The SATG helps capture sequential dependencies between historical user behaviors; hence, the relevant course is recommended by the model. Additionally, SATG preserves long-term, semantically relevant connections while suppressing irrelevant past behaviors. The proposed SATG-TT achieves a Normalized Discounted Cumulative Gain (NDCG) of 54.82 and 51.02% at @5 as well as @10 on the MOOCube and XuetangX datasets, respectively. These results demonstrate that the proposed model exhibits competitive performance compared to existing baseline models such as Knowledge Sequence Course Recommendation (KSCR).

Keywords: Course Recommendation, Historical Behaviors, Normalized Discounted Cumulative Gain, Semantic Aware Temporal Gating, Time Transformer, Users Interests

Introduction

E-learning (EL) is an online learning platform, which is utilized by various individuals and educational organizations for deeper understanding of knowledge. Nowadays, the idea of online learning has increased, as it enables various users to learn effectively with personalized options to develop individual skills and knowledge (Mrhar and Abik, 2025). EL education is preferred with the help of the internet, making it an ideal option for efficient data transmission, which allows the learners to improve their knowledge without the need to attend physical classes in educational institutions (Shuiping, 2026). The rise of digitalization has allowed MOOC platforms to expand and deliver high-quality educational materials. The capability levels vary among users and the presence of different educational guidelines

is difficult for users to recognize proper content (Feixiang, 2024). In addition, the users struggle to recognize a suitable course due to the massive volume of information, this necessitates the need for a recommendation system, which helps users to select relevant course (Siswipraptini et al., 2024). Furthermore, it attracts a huge learning audience compared to conventional EL platforms (Kukkar et al., 2024). In online education, student progress is enhanced through continuous performance estimation by observing various kinds of assessments. EL eventually became a necessary component for the education sector, which contributes to the development of quality education that transfers knowledge between students and all educational platforms (Bhanuse and Mal, 2023).

Semantic Similarity measures the conceptual relatedness and content alignment between different courses based on their metadata features such as

instructors, institutions, and core knowledge concepts (Chen et al., 2024). It evaluates how closely the domain knowledge of a past course aligns with a future candidate course. The recommendation system is a necessary component in EL as it suggests relevant courses to the learner various methods such as social-networking, collaborative filtering, and group-based approaches (Chanaa and El Faddouli, 2024). The MOOC platform offers educational content for heterogeneous learners with various levels, different experiences, knowledge learning styles, and expectations. The learners are provided with various assistances based on their particular ambitions and learning purposes (Mansouri et al., 2024). The disconnection between formal and non-formal learning environments such as MOOCs creates a gap between learner's need and contributions of MOOC platforms (Haque et al., 2025). The knowledge gap between different competitive environments such as team-based and digital, making course design difficult for educators (Boufaida et al., 2025). The MOOC platform provides filters, which obtain relevant materials based on individual preferences and solve knowledge fragmentation by recommending materials based on logical sequence (Jiang et al., 2023). EL is viewed as a vital resource for all rather than just an alternative option and is used by professionals and adult learners to enhance social skills and workplace competitions (Ahmad et al., 2023). MOOC platforms act as digital hubs because they offer variety of courses from top-tier university such as Harvard. The reputation of these platforms increases the recognition of acquired skills and supports career development (Ma, 2025). Furthermore, the availability of courses across diverse disciplines, enabling learners to pursue educational and professional goals based on their interests (George and Lal, 2024). Temporal intervals refer to an actual elapsed time between consecutive course enrollments or learning behaviors. Standard sequential models assume that a uniform or constant speed between actions and temporal intervals capture irregular real-world schedules and the effects of knowledge decay over time.

The major contributions of this work include:

- A Semantic-Aware Temporal Gating (SATG) mechanism is proposed, which mathematically fuses semantic similarity and continuous temporal distance directly within the self-attention layer. This represents a structural shift from the traditional frameworks by creating an explicit temporal-semantic filter
- Unlike existing sequential models that treat historical actions uniformly, the SATG mechanism produces adaptive gating coefficients. These coefficients selectively preserve long-term, temporally distant but semantically relevant learning pathways while actively suppressing irrelevant historical interactions

- A Time Transformer (TT) is utilized in this research that replaces rigid positional encodings with continuous, time-aware positional values. This allows the system to model real-world student schedules, accounting for knowledge decay, and capture global learning patterns simultaneously without being limited by local sequential constraints
- Experimental results on the MOOCCube and XuetangX datasets show that SATG-TT reported 54.82% and 51.02% NDCG a competitive performance than existing graph-based and deep sequential baselines

Related Work

This section denotes the advantages and limitations of existing approaches for student course recommendation based on MOOCCube and XuetangX datasets.

Deng et al. (2023) implemented Knowledge Sequence Course Recommendation (KSCR) with deep learning, which was used to gather sequential and learning structures for online CR. The KSCR approach analyzed learning sequences and a point wise sequential dependency for CR based on MOOC platforms. The KG and course information were represented by TransD, and course KG in learning sequence modeling enhanced the KSCR performance. However, the KSCR model primarily relied on enrolment sequence order and lacked explicit temporal interval modeling, which reduced its effectiveness in handling long-range learning behaviors and temporally sparse interactions.

Hao et al. (2023) developed a Meta Relationship CR (MRCR) to enhance expression of course relational data, which focused in complex semantic data of multi entity relationships. The semantic information and entity integration were produced by multi-entity relational Self-Symmetric Meta-Path (MSMP) and Associative relational Self-symmetric Meta-Graph (ASMG) was called as MR. The MR Correlation (MRCor) was used to obtain semantic correlational information and integrated the information into the graph embedding to measure heterogeneous information. Nonetheless, MRCR effectively captured semantic relationships between heterogeneous entities, it failed to adapt to the temporal evolution of learner preferences, limiting its ability to adapt to changing learning interests over time.

Wang et al. (2023) introduced a Multi aspect Heterogeneous Information Network (Multi-HIN) for MOOC knowledge concept recommendation. In the primary stage, HIN was used to connect the concepts with the multiple types of entities in the network. In order to learn a precise node representation, the context was dynamically constructed for each node by representing student interest dimensions. The item representation was learned by an aspect selection procedure, which was performed by a gumbel softmax approach and extended Matrix Factorization (MF) approach was applied at the

end for concept recommendations. The Multi-HIN highlighted heterogeneous semantic information modeling; however, the absence of explicit temporal sequence modeling restricted its ability to represent dynamic learning trajectories, evolving student interests.

Li et al. (2024) developed a learning path-based hypergraph neural network called as EduGraph for MOOC CR. EduGraph was considered by the combination of learning hypergraph path, where the learners were represented as hyper edges and courses were defined as vertices. The EduGraph framework encompassed a framelet-based hypergraph convolution, which was associated to highlight similarities and high pass filters to underscore various learning paths between the learners. Though its effectiveness in modeling structural relationships through hypergraph learning, EduGraph did not regulate the temporal relevance of historical learning interactions, which affected the long-term recommendation accuracy.

Ma et al. (2024) implemented a Contrastive Learning and Graph Convolution Network-based Attentive Decay Network (CLGADN) for associating learner's knowledge background to enhance CR. The CLGADN operated a contrastive learning to recognize various background knowledge for learners, to overcome the issue of bias. A monotonic attention decay model was associated with CLGADN for knowledge forgetting curve providing a structure to understand new courses based on knowledge obtained in the past. However, the CLGADN incorporated attention-based decay mechanisms, the monotonic decay strategy reduced the influence of semantically important historical courses that remained relevant to future learning pathways.

Existing course recommendation approaches effectively capture sequential behaviors, semantic

relationships, and structural course interactions. However, these traditional sequential and graph-based MOOC recommendation baselines primarily capture localized temporal behaviors, causing severe position decay and a breakdown in semantic continuity over long-term enrollment gaps. To address these limitations, the proposed SATG-TT framework integrates a TT with a temporal gating mechanism, which enables the simultaneous modeling of temporal dependencies and semantic continuity for improving course recommendation performance. Consequently, SATG-TT addresses the research gap by traditional sequence models to achieve enhanced position-aware ranking precision in the course recommendation system.

Materials and Methods

In this section, the proposed SATG-TT is used to analyze the user behavior to recommend relevant courses based on individual preferences. The proposed SATG-TT regulates temporal dependencies using time intervals, which model semantic similarities between previously learned courses. The SATG selectively preserves temporally distant connections but it is semantically relevant by suppressing historical behaviors.

The KG embedding of course helps in capturing the semantic relationship among the courses. The MOOCube and XuetangX datasets are used as a primary input for the model and Fig. 1 provides the architecture of the proposed SATG-TT for the student course recommendation system.

In the research, the MOOCube (Deng et al., 2023) and XuetangX (Ma et al., 2024) datasets are used to evaluate and estimate the model performance, which are described below:

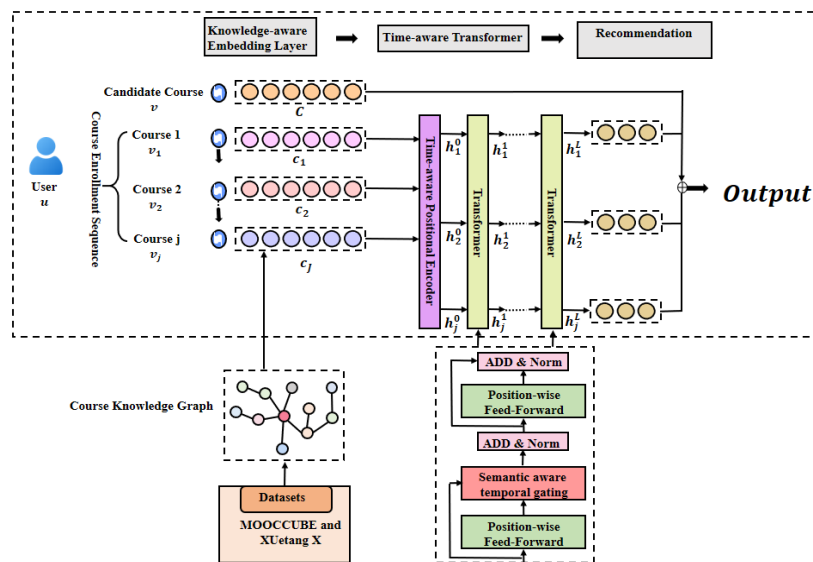


Fig. 1: Architecture of the proposed SATG-TT for student course recommendation system

Dataset Acquisition

- MOOCCube: The dataset consists of 1,99,199 students, 706 courses, 38,181 videos, 114,563 concepts, 1738 trainers, and 208 universities
- XuetangX dataset: XuetangX consists of 9986 students, 7020 courses, 43,405 videos, and 1029 concepts. The dataset contains various relationships such as user to course, user to video, and user to concepts

Recommendation Using Time Transformer

In this research, the MOOCCube and XuetangX datasets are used to create KG, which maps courses-based on metadata. The knowledge aware embedding converts the course into a low dimensional vector and captures the semantic relationship between courses. In contrast, TT helps frame users' interests over time by using a specific order of the classes with the help of the past courses enrollment sequence.

Course Knowledge Graph (KG)

The KG embedding captures semantic relationships between courses using various entities, which include trainers, schools, concepts, and connections. Additionally, KG helps organize extensive data for improving online course recommendations by analyzing semantic relationships. The KG transforms each entity and connection into a vector and simultaneously preserves the arrangement of the graph. Particularly, KG is prolonged to contain high entities and relations from high data sources.

Time Aware Transformer

The time aware transformer interprets each student's individual learning interest based on course enrollment arrangements. While the conventional positional encoding considers only relative order, the TT calculates a time-aware positional value for each course. The stacked layers from a single transformer create a deep neural network to learn difficult changing outlines that are present in courses. The conventional positional encoding assumes constant speed between courses and neglects the knowledge decay; hence, it struggles with real world student schedules. The order of course and time taken for individual improvement is important to capture the actual learning interest of the students. The time aware position index $k(1 \leq k \leq j)$ for each individual position is calculated by using Eq. 1:

$$TPos_{(v_k)} = k + t(v_k) \quad (1)$$

Here, $TPos_{(v_k)}$ represents the final time modulated positional value of k^{th} course in the student's enrollment

history, combining both sequential order and temporal progression. Next, k denotes the sequential position index, referring to discrete index indicating the relative chronological order of the enrolled course within the sequence ($1 \leq k \leq j$). Subsequently, v_k indicates the specific course enrollment event occurring at index k and $t(v_k)$ refers to the relative time interval, which measure temporal distance that is associated with the enrollment of course v_k . $TPos_{(v_k)}$ is extended to n dimensional positional embedding P_k , $i(1 \leq i \leq n)$ is extended within P_k in order to encode the time aware positional value. Each of the dimension value is mentioned in Eq. 2:

$$P_k^{(i)} = \begin{cases} \sin\left(\frac{TPos_{(v_k)}}{10000^{\frac{2i}{n}}}\right), & \text{if } i \text{ is even} \\ \cos\left(\frac{TPos_{(v_k)}}{10000^{\frac{2i}{n}}}\right), & \text{if } i \text{ is odd} \end{cases} \quad (2)$$

Here, $P_k^{(i)}$ indicates the specific value of the i^{th} coordinate within the continuous time-aware positional vector for the k^{th} course. $TPos_{(v_k)}$ denotes the time-modulated position index, which is calculated from Equation (1), reflecting both the sequential ordering and the actual elapsed time. Next, i denotes the specific dimension component that is calculated within the vector space, ranging from 1 to n . Subsequently, $10000^{\frac{2i}{n}}$ represents the geometric frequency scaling factor, which is used to assign distinct frequencies to different dimensions of the embedding, ensuring the model can learn and distinguish learning patterns across varying timescales. The $\sin$ and $\cos$ specifies the sinusoidal functions used to map the continuous scalar index into a multi-dimensional periodic space, allowing the network to easily learn relative time-aware positions through linear transformations. In the provided course v_k , the input is h_k^0 , which is created by concatenating the KG embedding c_k , and time aware positional embedding P_k , the mathematical formula is mentioned in Eq. 3:

$$h_k^0 = c_k + P_k \quad (3)$$

The embedding matrices of all courses are denoted as $H = [h_1^0, h_2^0, \dots, h_j^0]$, which is stacked transformer layers, which is used to capture the dependencies between courses within a particular sequence. The transformer layer contains two layers, namely the Position Wise Feed Forward Layer (PWFFL) and the Multi Head Self Attention Layer (MHSAL). The MHSAL enables the transformer model to integrate data from the distinct representation subspaces at various positions in the enrollment sequence. PWFFL provides a nonlinearity transformer with connections among various dimensions, the MHSAL mathematical formula is mentioned in Eqs. 4 and 5:

$$MultiHead(H) = concat(head_1; head_2; \dots; head_h)W^o \quad (4)$$

$$head_i = Attention(HW^Q, HW^K, HW^V) \quad (5)$$

Here, *MultiHead* denotes the MHSAL output; *concat(.)* refers to concatenation operator *h*, which specifies the number of attention heads; *H* represents the input representation matrix and HW^Q, HW^K, HW^V , which represents head of projection weight matrices. W^o denotes the output projection matrix and $head_i$ indicates the output of i^{th} attention head. The *Attention(.)* exhibits core attention function. The attention function for the scaled dot product attention is provided in Eq. 6:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d}}\right)V \quad (6)$$

Where *Attention(Q, K, V)* denotes the Attention output matrix where *Q*, *K* and *V* denote query, key and value. K^T indicates the transpose of key matrix, $\sqrt{d}$ specifies a scaling factor, which generates a softer attention distribution to avoid extremely small gradients and *softmax(.)* represents the Softmax activation function.

Semantic Aware Temporal Gating

The TT integrates temporal information by using time aware positional encoding through an indirect and discrete manner rather than depending on explicit continuous state transitions. TT efficiently captures temporal data using time aware positional encoding but temporal dependencies are not learnt through explicit state transitions; hence, this indirect approach is considered as a limitation. The SATG overcomes the challenge by directly controlling the temporal relevance by preserving semantically related long-term interactions. Furthermore, SATG efficiently controls the temporal influence based on time and semantic relevance, which makes it better for sparse and long term learning behaviors that is observed in the dataset. The SATG controls the temporal contribution based on the learning relevance of student history $\mathcal{S}$ and operates temporal distance, which is mathematically represented in Eqs. 7 and 8:

$$\mathcal{S} = \{(c_i, t_i)\}_{i=1}^n \quad (7)$$

$$\Delta t_{ij} = |t_i - t_j| \quad (8)$$

Here, $\mathcal{S}$ denotes the student enrollment sequence; c_i and t_i designates the course interaction item and the enrollment timestamp where (c_i, t_i) demonstrates the interaction pair tuple. *n* represents the sequence length, Δt_{ij} specifies the Temporal distance and $|\cdot|$ refers to the absolute value operator. The semantic similarity ($SemSim_{ij}$) between courses are obtained using semantic

representations is mathematically expressed in Eq. 9:

$$g_{ij} = \sigma(\alpha \cdot SemSim_{ij} - \beta \cdot \Delta t_{ij}) \quad (9)$$

Where g_{ij} represents the SATG coefficient; σ denotes the Sigmoid activation function; α and β indicates the Semantic weight and temporal decay hyperparameters. $SemSim_{ij}$ refers to the semantic similarity score and $\cdot$ denotes the scalar multiplication operator. The conventional transformer-based recommendation models incorporate temporal information only through positional encoding, whereas the proposed SATG mechanism directly regulates the self-attention process using a semantic-temporal gating coefficient. The semantic similarity term $SemSim_{ij}$ preserves historically relevant learning interactions, whereas the temporal distance term Δt_{ij} suppresses the outdated interactions. Consequently, the gating coefficient explicitly controls the contribution of each historical course before attention aggregation. This differs from conventional temporal transformers, where temporal information influences only the input representation and does not directly participate in attention regulation. The gating coefficient is applied to the self-attention mechanism, which is expressed in the following Eq. 10:

$$A_{ij}^{SATG} = g_{ij} \times A_{ij} \quad (10)$$

Here, A_{ij} and A_{ij}^{SATG} represents the actual attention score and gated attention score, the proposed SATG does not merely encode temporal information at the input level. Instead, the gating coefficient directly modifies the attention score according to Eq. (10), enabling semantic relevance and temporal distance to jointly influence the attention distribution. Therefore, temporally distant but semantically important courses retain their influence, while unrelated historical interactions receive lower attention weights. The gating coefficient is applied to the self-attention mechanism, which confirms semantic relevance and obtains only the temporally distant course while suppressing irrelevant interactions. The semantic temporal regulation supports the model to prioritize knowledge stability and enables the TT to effectively capture long-term learning trajectories, suggesting courses to align with student interest. The MHSAL generates PWFFL output, which recognizes each position by including two linear transformations with ReLU activation, which is mentioned in Eq. 11:

$$PWFFL(x) = max(0, (xW_1 + b_1)W_2 + b_2) \quad (11)$$

Where $PWFFL(x)$ denotes the PWFFL output, x indicates the input feature matrix and $max(0, \cdot)$ specifies the ReLU activation function. W_1 , b_1 and W_2 , b_2

represent the first and second linear layer weight as well as bias matrix. After L layers, the output is $H^L = [h_1^L, h_2^L, \dots, h_j^L]$ for all courses in an input sequence. The proposed SATG-TT controls the temporal relevance while preserving semantically related long-term interactions with the attention layer by integrating temporal distance Δt_{ij} and SemSim_{ij} . Existing transformer-based recommendation models generally utilize temporal embedding, relative time encoding, or graph propagation mechanisms to model temporal behavior. In these approaches, temporal information primarily affects the input representation stage, whereas the proposed SATG-TT introduces a semantic-temporal gating mechanism that clearly regulates self-attention using both semantic similarity and temporal distance. Consequently, the proposed framework simultaneously preserves semantic continuity and temporal relevance during attention computation, providing a temporal-semantic integration that is not considered in conventional transformer-based course recommendation architectures.

Additionally, the SATG-TT model produces adaptive gating coefficients, which dynamically regulate the involvement of each past history and interaction of students. This enables the proposed SATG-TT to preserve semantically meaningful learning experiences after prolonged period without affecting the student's learning course interest and recommendation. The integration of the proposed SATG-TT mechanism semantically guides the temporal model, which increases recommendation performance in the student course recommendation system.

Algorithm for Proposed SATG-TT for Student Course Recommendation

```

Input: MOOCCube and XuetangX
Initialize course KG and learn course semantic
embedding
Initialize Time Transformer parameters and SATG
parameters
Load user enrollment sequences from MOOCCube and
XuetangX dataset
    For each user u in dataset Do
        Extract historical course sequence
        Extract corresponding enrollment timestamps
    End For
    For each training epoch Do
        For each user u Do
            Knowledge-aware embedding
            Obtain semantic embedding of each enrolled course
            from KG
            Time-aware positional encoding
            Compute time interval  $\Delta t$  between course enrollments
            Generate time-aware positional embedding
            Combine semantic embedding and time-aware positional
    
```

```

embedding
    Compute Query, Key, Value matrices
    Calculate scaled dot-product attention
    Semantic-Aware Temporal Gating (SATG)
    Compute temporal distance between historical
interactions
    Compute semantic similarity between courses
    Calculate gating coefficient using temporal and
semantic relevance
    Apply gating coefficient to attention scores
    Transformer feed-forward layer
    Apply position-wise feed-forward network
    Stack transformer layers
    Repeat steps 18–26 for L layers
    Generate final user interest representation
    Course ranking
    Predict scores for candidate courses
    Rank courses based on predicted scores
    Compute loss using ground-truth enrolled courses
Update model parameters using back-propagation
    End For
End For
    For each test user u Do
        Generate Top-K recommended courses
    End For
    Return Top-K course recommendations
Output:
Top-K recommended courses for each student
    
```

Results and Discussion

The proposed SATG-TT model's performance has been estimated using MOOCCube and XuetangX datasets, the system configuration is described below:

- Processor: 12th Gen Intel(R) Core(TM) i7-12700K (3.60 GHz)
- Installed RAM: 64.0 GB
- System type: 64-bit operating system
- OPERATING-OS: Windows 11 Pro
- SSD: 1.86 TB

The proposed SATG-TT model is used to recommend courses and classify the courses into two classes such as completed and not completed. The performance metrics such as Mean Reciprocal Rank (MRR), Hit Rate of Top-K items ($HR@K$), and Normalized Discounted Cumulative Gain of Top-K items ($NDCG@K$) are applied to evaluate the proposed SATG-TT model, as described below:

$HR@K$: The $HR@K$ refers to hit rate where $K = 5$ and 10 , which measures the percentage of ground truth instances of the user, which is captured within the top k recommended course list and it is mathematically provided in Eq. 12:

$$HR@K = \frac{1}{N} \sum_{i=1}^N hit(i) \quad (12)$$

Where N represents the Total number of evaluation samples and i directs the sample/student index. $\sum_{i=1}^N hit(i)$ refers to the cumulative sum of individual recommendation results, whereas $hit(i)$ depicts the binary indicator, which is used to identify the success of a recommendation for the learner i . The value of one is assigned if the course selected by the learner is presented in the recommended list, while zero is assigned otherwise.

NDCG@K: The NDCG is used to estimate the ranking output, NDCG@K is a precision-based measure, which explains the recommended course items position in the ranking list. The mathematical formula for NDCG@K is mentioned in Eq. 13:

$$NDCG@K = \frac{1}{|U|} \sum_{u \in U} \sum_{i=1}^K \frac{I(rank_{u,i} \in S_u)}{\log_2(i+1)} \quad (13)$$

Where $NDCG@K$ represents the position aware ranking metric that evaluates quality of the top-k recommended courses. $|U|$ indicates the total number of the evaluation error $K = 5$ and 10 , where u denotes the user index. Next, $\sum_{u \in U}$ represents the user summation operator and $\sum_{i=1}^K$ denotes the Rank position summation operator, where i specifies the rank position index. Subsequently, $I(.)$ represents the indicator function, which provides the value as 1 if the condition is true and 0 if the condition is false. The $rank_{u,i}$ indicates the actual course item, which is recommended to the user u at the position index i in the output list. Additionally, S_u denotes the ground-truth interaction set, which contains actual courses that the user u enrolled during the test phase. The $\log_2(i + 1)$ specifies the logarithmic position discount factor.

MRR: The MRR is utilized to calculate an efficient ranking output of a course recommendation system, this metric applies the average reciprocal rank across all learners, which mathematical formula is mentioned in Eq. 14:

$$MRR = \frac{1}{N} \sum_{i=1}^N \frac{1}{P_i} \quad (14)$$

Where MRR denotes the ranking evaluation metric that measures the quality of the recommendation system. $\sum_{i=1}^N$ denotes the cumulative sum of the individual reciprocal ranks that are computed across all evaluation instances from $i = 1$ to N . P_i refers the position of the target course and $\frac{1}{P_i}$ specifies the reciprocal rank value.

Experimental Setup

To ensure the experimental integrity and reproducibility, uniform preprocessing pipelines and evaluation protocols are applied in both datasets. For both MOOCCube and XuetangX, the noisy interactions are filtered by removing users with fewer than 5 course enrollments and courses with fewer than 5 total student interactions, mitigating data sparsity issues. The chronological interaction history for each student is partitioned using a standard leave-one-out evaluation split strategy. The last chronological course interaction in the sequence is reserved for the test set, the second-to-last interaction is assigned to the validation set for hyperparameter tuning, and all preceding historical interactions are used for model training. To evaluate ranking performance efficiently, a paired negative sampling strategy is adapted during evaluation. For each ground-truth positive item in the validation and test sets, 99 negative course items that the user has never interacted are randomly sampled, creating a candidate ranking pool of 100 items per test instance. The core parameter configurations are used during optimization that are detailed in Table 1. All the baseline models are reimplemented as per the evaluation protocol that considers for the proposed model. The proposed recommendation system is simulated using PyTorch and optimized using the Adam optimizer with a batch size of 256 over 100 epochs, utilizing early stopping with a patience of 10 epochs based on validation loss.

Knowledge Graph (KG) Construction and Semantic Representation

The course KG is constructed using the entity relationships, which are available in the MOOCCube and XuetangX datasets.

Table 1: Core Architecture and optimization parameters of the proposed student course recommendation system

Parameter Group	Hyperparameter Name	Operational Value (MOOCCube)	Operational Value (XuetangX)
Model Embedding	Hidden Dimension (d)	128	128
	Max Sequence Length (n)	50	50
Transformer Block	Attention head (h)	4	4
	Dropout rate	0.2	0.2
Gating Network	Semantic Scale (α)	0.45	0.50
	Temporal decay (β)	0.15	0.12
Optimization	Learning rate	0.001	0.001
	L2 regularization (weight decay)	1×10^5	1×10^5

Each course represents as a node and is connected with associated entities including concepts, instructors, universities, and learning resources through typed relations. The resulting heterogeneous graph preserves semantic associations between courses and educational entities. To generate semantic course representations, KG embeddings are learned by mapping each entity and relation into a shared low-dimensional vector space. The final course embedding is obtained by aggregating the embeddings of the course node and is directly connected to the neighboring entities. Semantic similarity between two courses are computed using cosine similarity between their corresponding KG embedding vectors. This similarity score is subsequently incorporated into the Semantic-Aware Temporal Gating mechanism to preserve semantically relevant historical learning interactions.

Negative Sampling Strategy

During validation and testing, one ground-truth positive course is paired with 99 negative course candidates, randomly sampled from the set of courses with which the corresponding learner has no historical interaction. The selection of 99 negative samples follows leave-one-out evaluation protocol that is used in sequential recommendation research, where one positive item is ranked among a candidate set of 100 items. This setting provides a balanced trade-off between computational efficiency and ranking difficulty while enabling fair comparison with existing recommendation methods. To avoid information leakage, negative samples are selected without overlapping learner's training, validation, or test interactions. The resulting candidate set consist of one positive and 99 negative courses that are used to compute HR@K, NDCG@K, and MRR evaluation metrics.

Performance Analysis

In the performance analysis section, the proposed SATG-TT is compared with various DL approaches such as Gated Recurrent Unit (GRU), Bidirectional-GRU (Bi-GRU), and Long Short-Term Memory (LSTM). The performance estimation of the proposed SATG-TT is estimated and evaluated across two datasets, namely MOOCCube and XuetangX, and its performance is analyzed by various metrics such as HR@5, HR@10, NDCG@5, NDCG@10, and MRR. Table 2 demonstrates that the proposed SATG-TT model attains better recommendation output, while compared with other existing approaches. The performance analysis of the proposed SATG-TT model is performed across various models such as GRU, Bi-GRU, LSTM, Time Frequency Sequential Recommendation (TFSR), Self-Attentive Sequential Recommendation (SASR), and Temporal Graph Transformer (TGT) across both MOOCCube and XuetangX datasets. The recurrent models capture local sequential dependencies, while the proposed SATG-TT utilizes self-attention for global temporal modeling and associates SATG to preserve relevant historical courses. These integrations enhance recommendation quality, providing precise course suggestions that improve students' decision-making capabilities.

Table 3 denotes the computational complexity analysis of the proposed SATG-TT for student learning course recommendation. This computational analysis demonstrates that SATG-TT minimizes storage requirements and reduces per-batch (with a batch size of 64) training, testing, and inference times on both the MOOCCube and XuetangX datasets. The proposed SATG-TT framework achieves a favorable trade-off between computational efficiency and recommendation performance.

Table 2: Performance analysis of the proposed SATG-TT is compared with other DL-based approaches for student course recommendation

Datasets	Methods	HR@5 (↑)	HR@10 (↑)	NDCG@5 (↑)	NDCG@10 (↑)	MRR (↑)
MOOCCube	LSTM	60.18	77.64	42.05	50.12	35.83
	Bi-LSTM	62.45	79.50	44.20	52.10	37.40
	GRU	61.24	78.35	43.12	51.06	36.94
	Bi-GRU	63.87	80.92	45.76	53.48	38.57
	TFSR	65.10	81.90	47.15	55.02	39.80
	SASR	66.95	83.15	48.90	56.95	41.15
	TGT	68.14	84.60	50.33	58.40	44.65
	Proposed SATG-TT	72.57	87.97	54.82	62.78	46.87
XuetangX dataset	LSTM	52.18	71.03	40.11	45.02	36.74
	Bi-LSTM	54.90	73.90	42.15	48.10	38.60
	GRU	53.46	72.81	41.07	46.35	37.92
	Bi-GRU	56.89	75.46	43.52	44.27	39.85
	TFSR	58.02	76.85	44.60	48.70	40.50
	SASR	59.35	77.90	45.45	49.10	41.15
	TGT	60.25	78.90	46.12	49.35	41.90
	Proposed SATG-TT	63.87	81.57	49.16	51.02	44.58

Table 3: Performance analysis of computational complexity analysis of the proposed SATG-TT for student learning course recommendation

Datasets	Methods	Training Time (s)	Testing Time (s)	Inference Time(s)	Storage (KB)
MOOCCube	LSTM	195	0.102	0.089	2760
	Bi-LSTM	228	0.124	0.098	3450
	GRU	182	0.095	0.082	2480
	Bi-GRU	215	0.118	0.091	3120
	TFSR	245	0.115	0.102	4120
	SASR	260	0.122	0.110	4380
	TGT	190	0.088	0.074	2890
	Proposed SATG-TT	165	0.071	0.058	2240
	LSTM	181	0.097	0.083	2620
	Bi-LSTM	216	0.119	0.093	3210
XuetangX dataset	GRU	168	0.089	0.076	2360
	Bi-GRU	204	0.112	0.091	2980
	TFSR	225	0.108	0.095	3940
	SASR	238	0.114	0.101	4150
	TGT	178	0.081	0.069	2710
	Proposed SATG-TT	152	0.066	0.052	2180

Compared to recurrent architectures that process learner interactions sequentially, the TT leverages parallel self-attention operations to model long-range learning dependencies efficiently. However, the SATG mechanism introduces additional semantic similarity and temporal relevance computations, these operations are executed through parallel matrix multiplications, which does not alter the overall complexity order. As a result, SATG-TT attains a slightly higher computational cost than the TGT model but remains more efficient than bidirectional recurrent architectures in practical implementations.

Fig. 2. demonstrates the cross-generalization analysis, the training dataset is XuetangX and testing dataset is MOOCCube. The results are obtained from the cross-dataset evaluation, which confirm that the proposed SATG-TT generalizes effectively when trained on XuetangX dataset and tested on the MOOCCube dataset. This enhancement is achieved by a TT self-attention model, which learns long range temporal patterns and SATG mechanism is used to suppress dataset noise while emphasizing meaningful learning behaviors. Additionally, SATG-TT achieves exceptional cross-dataset generalization because its gating matrix optimizes domain-invariant features specifically relative to time distances (Δt_{ij}) and relative theme similarities (SemSim_{ij}) rather than memorizing platform-specific course IDs. Consequently, by transferring zero-shot learning to a new educational platform, vocabulary mismatches are successfully bypassed to accurately map student behaviors.

Fig. 3 presents the ablation study of the proposed SATG-TT mechanism across two datasets for the student course recommendation system. The ablation study validates the individual and combine the contribution of SATG-TT in the student course recommendation system. The TT model converts discrete and irregular enrollment

intervals into continuous vectors, enabling the architecture to mathematically model student knowledge decay instead of treating varying timeline gaps uniformly. The proposed SATG acts as a non-linear information filter, preventing historical noise and information overload by adjusting attention weights based on contextual relevance rather than simple interaction frequency.

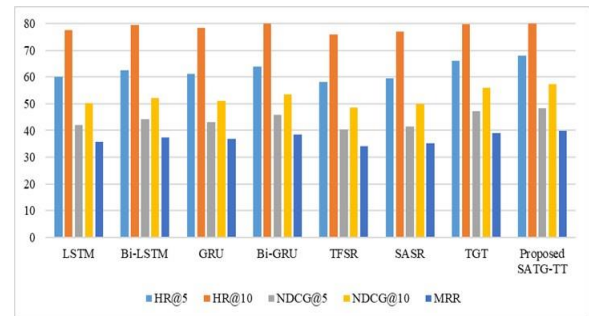


Fig. 2: Cross generalization analysis, training dataset is XuetangX and testing dataset is MOOCCube

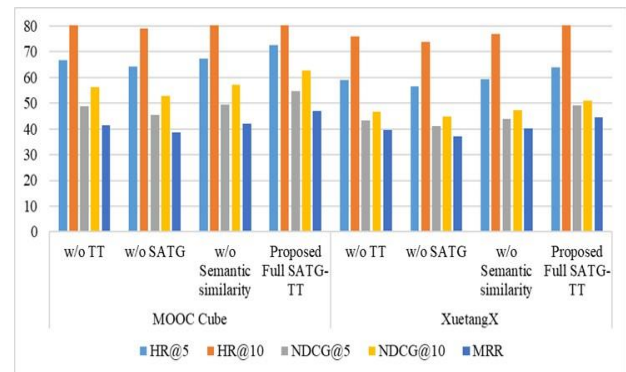


Fig. 3: Ablation study of the proposed SATG-TT mechanism across two datasets for student course recommendation system

With in this gate, the Semantic similarity component creates a competitive dynamic; it provides a structural override that preserves old, foundational prerequisite courses if they match the current learning trajectory, mitigating strict recency bias. To quantitatively evaluate the contribution of each component, an ablation study is conducted by removing the Time Transformer (TT), Semantic-Aware Temporal Gating (SATG), and semantic similarity module individually. On the MOOCCube dataset, removing TT reduces the NDCG@10 from 63 to 56, corresponding to a 11.11% performance degradation, while removing SATG reduces NDCG@10 from 63 to 53, corresponding to a 15.87% degradation. Similarly, by removing semantic similarity reduces NDCG@10 from 63 to 58, resulting in a 7.94% degradation. For HR@10, the removal of TT, SATG, and semantic similarity reduces the performance by 9.09%, 10.23%, and 9.09%, respectively. In the XuetangX dataset, removing TT reduces NDCG@10 from 51 to 47, corresponding to a 7.84% degradation, whereas removing SATG reduces NDCG@10 from 51 to 45, corresponding to an 11.76% degradation. By eliminating the semantic similarity module, it reduces NDCG@10 from 51 to 47, corresponding to a 7.84% degradation. Similarly, HR@10 is decreased by 8.54%, 10.98%, and 7.32% when TT, SATG, and semantic similarity are removed, respectively.

These results demonstrate that all three components of the proposed model contribute positively to the student course recommendation performance. Among them, SATG provides the largest individual contribution on both datasets, confirming the importance of semantic-temporal gating in preserving relevant historical learning interactions. The combined integration of TT, SATG, and semantic similarity achieves the highest recommendation accuracy, validating the complementary effect of all components within the proposed SATG-TT framework.

Comparative Analysis

In the comparative analysis section, the proposed SATG-TT is compared with existing approaches such as

KSCR (Deng et al., 2023), MRCR (Hao et al., 2023), Multi-HIN (Wang et al., 2023), EduGraph (Li et al., 2024), and CLGADN (Ma et al., 2024). Compared to other existing approaches, the proposed model attains the high level of HR@5, HR@10, NDCG@5, NDCG@10, and MRR. Table 4 provides the comparative analysis of the proposed SATG-TT, which is compared with other existing approaches based on the MOOCCube dataset.

Table 5 illustrates the comparative analysis output of the proposed SATG-TT, the existing approaches fail to control long-term temporal relevance, which limits the capability to manage irregular learning behaviors. By associating SATG-TT, the proposed model efficiently captures student interests, which leads to enhance student course recommendations in both MOOCCube and XuetangX datasets.

Comparative analysis of the proposed method on the MOOCCube dataset, reporting competitive performance when compared to other approaches across all performance metrics under considered evaluation settings. The KSCR and MRCR are based on limited relational depth, which provides low recommendation performance but the graph-based approaches such as multi HIN and EduGraph increase semantic representation. However, the graph-based approaches remain insufficient as they fail to extract long term dependencies. The proposed model integrates global temporal modeling with semantic aware temporal gating, which allows the selective preservation of relevant historical learning interactions.

Discussion

This section describes the proposed SATG-TT advantages and limitations of existing approaches in course recommendation systems. Conventional recommendation models such as KSCR and MRCR models are performed based on shallow sequential and static relational representations, affecting the capability to capture long term dependencies and estimate learning recommendations.

Table 4: Comparative analysis of the proposed SATG-TT is compared with other existing approaches for course recommendation system

Dataset	Methods	HR@5	HR@10	NDCG@5	NDCG@10
MOOCCube dataset	KSCR ((Deng et al., 2023)	30.42	45.57	19.39	24.21
	MRCR (Hao et al., 2023)	51.12	65.87	37.65	42.41
	Multi-HIN (Wang et al., 2023)	63.93	84.12	41.82	51.30
	EduGraph (Li et al., 2024)	68.23	83.93	51.25	58.40
	Proposed SATG-TT	72.57	87.97	54.82	62.78

Table 5: Comparative analysis of the proposed SATG-TT is compared with existing model based on XuetangX dataset

Dataset	Methods	HR@5	HR@10	NDCG@5	NDCG@10	MRR
XuetangX dataset	MRCR (Hao et al., 2023)	60.22	77.90	42.51	48.28	40.19
	CLGADN (Ma et al., 2024)	58.98	74.73	43.24	48.35	41.45
	Proposed SATG-TT	63.87	81.57	49.16	51.02	44.58

The inability of the model to efficiently retain a long term contextual information, leading to suboptimal recommendation quality. The graph-based models such as Multi-HIN, EduGraph, and CLGADN illustrate the heterogeneous information and structured relationship between courses due to the lack of temporal influence of historical connections that leads to minimizing the ranking precision. The proposed SATG-TT overcomes these limitations, the TT allows the model to attend all historical learning connections, which permits to capture global learning patterns rather than sequential dependencies. The proposed SATG-TT model confirms that the semantically relevant courses are preserved even across extended temporal gaps, while outdated and irrelevant connections are effectively suppressed. The performance analysis section provides HR@5 and HR@10, which depict that the proposed SATG-TT is efficient in gaining the relevant courses with top ranked recommendation list. Evaluation using NDCG@5 and NDCG@10 confirms that the SATG-TT model delivers effective recommendations by accurately predicting next-course selections and ensuring top-ranked relevance.

Research Implication

The proposed SATG-TT framework provides practical benefits for online learning platforms by generating personalized course recommendations based on both temporal learning behavior and semantic course relationships. It helps the learners to identify relevant courses from large repositories, reducing information overload and improving learning experiences. By retaining long-term interaction data despite extended gaps, the framework maintains accurate learning pathways for professionals and perpetual learners. Additionally, improved recommendation accuracy enhances learner engagement, course completion rates, and overall educational effectiveness in MOOC environments.

Conclusion

In this manuscript, the proposed SATG-TT model captures sequential dependencies within historical user behavior to generate sequential recommendations and suggest relevant courses. The proposed SATG module is integrated with the self-attention layer of the TT, which integrates the required attention weights. Subsequently, SATG produced a gating coefficient by considering temporal distance between interactions and semantic similarity from course knowledge representations. The proposed SATG-TT allowed the model to select temporally distant learning behavior by handling outdated and irrelevant historical behaviors. The high level of MRR values ensured that the relevant courses were recommended accurately, which increased user satisfaction and learning efficiency of the student. The proposed SATG-TT attained 54.82 and 51.02%

Normalized Discounted Cumulative Gain items (NDCG) @5, @10 on MOOCCube, and XuetangX datasets, respectively. In future work, recommendation system should recommend out-of-scope courses, which is semantically related to student's current learning path and hence, the students will be able to explore wide range of interest.

Acknowledgment

I am appreciative of Dr. Dahlia Sam's crucial advice, inspiration, and assistance during my research. Her knowledge and perceptions greatly impacted the course of our inquiry. My sincere appreciation goes out to Dr. N. Kanya for their invaluable input and helpful critiques, which substantially improved the caliber of this work. Special thanks should be given to the Dr. M.G.R Educational and Research Institute's Department of Computer Science and Engineering (CSE) for providing the tools, resources, and support that made this work feasible.

Funding Information

This research received no external funding.

Author's Contributions

Joshi Vinay Kumar: Conceptualized and designed the study, supervised the overall research work, and contributed to manuscript writing and critical revision.

Dahlia Sam: Conducted the research work, performed data collection and experimentation, and contributed to drafting the manuscript.

N. Kanya: Carried out the data analysis, validation of results, and assisted in the interpretation of findings. All authors reviewed, revised, and approved the final version of the manuscript.

Ethics

This article does not contain any studies with human participants or animals performed by the authors. All ethical standards applicable to this type of research have been duly followed.

Notation Table

Notations	Description
$k(1 \leq k \leq j)$	Time aware position index
$TPos_{(v_k)}$	Final time modulated position value k^{th} course in the student's enrollment history, combining both sequential order and temporal progression
k	Sequential position index refers to discrete index indicating the relative chronological order of the enrolled course within the sequence ($1 \leq k \leq j$)

v_k	Specific course enrollment event occurring at index k	c_i, t_i	Interaction pair tuple
$t(v_k)$	Relative time interval which measure temporal distance associated with the enrollment of course v_k relative to the first course enrollment v_1 in the sequence, accounting for irregular time gaps and knowledge decay	n	Sequence length
$TPos(v_k)$	Extended to n dimensional positional embedding P_k , $i(1 \leq i \leq n)$ extended within P_k in order to encode the time aware positional value	Δt_{ij}	Temporal distance
$P_k^{(i)}$	Specific value of the i^{th} coordinate within the continuous, time-aware positional vector for the k^{th} course	t_j	Comparison timestamp
$TPos(v_k)$	Time-modulated position index calculated from Equation (1), reflecting both the sequential ordering and the actual elapsed time; i denotes specific dimension component being calculated within the vector space, ranging from 1 to n	$. $	Absolute value operator
$10000^{\frac{2i}{n}}$	Geometric frequency scaling factor used to assign distinct frequencies to different dimensions of the embedding, ensuring the model can learn and distinguish learning patterns across varying timescales	g_{ij}	SATG coefficient
$\sin$ and $\cos$.	Sinusoidal functions used to map the continuous scalar index into a multi-dimensional periodic space, allowing the network to easily learn relative time-aware positions through linear transformations	σ	Sigmoid activation function
$v_k,$ h_k^0 c_k P_k H $= [h_1^0, h_2^0, \dots, h_j^0]$	Provided course Input Knowledge graph embedding Time aware positional embedding Embedding matrices of all courses	α	Semantic weight hyper parameter
$MultiHead$ H $concat(.)$	Multi-Head Attention output Representation matrix Concatenation operator h specifies number of attention heads	$SemSim_{ij}$	Semantic similarity score
W^o $head_i$ $Attention(.)$ HW^Q, HW^K, HW^V $Attention(Q, K, V)$ Q, K and V K^T $\sqrt{d}$	Output projection matrix Output of i^{th} attention head Core attention function Head of projection weight matrices Attention output matrix Query, Key and Value Transpose of key matrix Specifies a scaling factor which generates a softer attention distribution to avoid extremely small gradients	$\cdot$	Scalar multiplication operator
$softmax(.)$ $\mathcal{S}$ c_i t_i	Softmax activation function Student enrollment sequence Course interaction item Enrollment timestamp	β	Temporal decay hyper parameter
		A_{ij}	Actual attention score
		A_{ij}^{SATG}	Gated attention score
		$PWFFL(x)$	$PWFFL$ output
		x	Input feature matrix
		$max(0, .)$	ReLU activation function
		W_1, b_1 and W_2, b_2	First and second linear layer weight and bias matrix
		H^L $= [h_1^L, h_2^L, \dots, h_j^L]$	Output is for all courses in an input sequence
		HR	Hit rate
		N $\sum_{i=1}^N hit(i)$	Total number of evaluation samples Cumulative sum of individual recommendation results whereas $hit(i)$ depicts a binary indicator, which is used to identify the success of a recommendation for learner i
		i	Sample/student index
		$NDCG@K$	Position aware ranking metric that evaluate quality of the top- k recommended courses
		$ U $	Total number of evaluation error; u denotes user index
		$\sum u \in U$	User summation operator
		$\sum_k$ $\sum_{i=1}^k$ i $I(.)$	Rank position summation operator Rank position index Indicator function which gives the value as 1 if condition is true and 0 if the condition is false
		$rank_{u,i}$	Actual course item recommended to user u at position index i in the output list
		S_u	Ground-truth interaction set which contains actual courses that user u interacted with or enrolled in during the test phase
		$log_2(i + 1)$	Logarithmic position discount factor
		MRR	Ranking evaluation metric that measures the quality of the recommendation system
		N $\sum_{i=1}^N$	Total number of evaluation samples Cumulative sum of the individual reciprocal ranks computed across all evaluation instances from $i = 1$ to N
		P_i $\frac{1}{P_i}$	Position of target course Reciprocal rank value

References

- Ahmad, H. K., Qi, C., Wu, Z., & Muhammad, B. A. (2023). ABiNE-CRS: course recommender system in online education using attributed bipartite network embedding. *Applied Intelligence*, 53(4), 4665–4684. <https://doi.org/10.1007/s10489-022-03758-z>
- Bhanuse, R. S., & Mal, S. (2023). Optimal e-learning course recommendation with sentiment analysis using hybrid similarity framework. *Multimedia Tools and Applications*, 83(6), 16417–16446. <https://doi.org/10.1007/s11042-023-16138-7>
- Boufaïda, S. O., Benmachiche, A., Bennour, A., Maatallah, M., Derdour, M., & Ghabban, F. (2025). Enhancing MOOC Course Classification with Convolutional Neural Networks via Lion Algorithm-Based Hyperparameter Tuning. *SN Computer Science*, 6(6), 707. <https://doi.org/10.1007/s42979-025-04179-8>
- Chanaa, A., & El Faddouli, N. (2024). Prerequisites-based course recommendation: recommending learning objects using concept prerequisites and metadata matching. *Smart Learning Environments*, 11(1), 16. <https://doi.org/10.1186/s40561-024-00301-0>
- Chen, X., Zou, D., Cheng, G., & Xie, H. (2024). Deep neural networks for the automatic understanding of the semantic content of online course reviews. *Education and Information Technologies*, 29(4), 3953–3991. <https://doi.org/10.1007/s10639-023-11980-6>
- Deng, W., Zhu, P., Chen, H., Yuan, T., & Wu, J. (2023). Knowledge-aware sequence modelling with deep learning for online course recommendation. *Information Processing & Management*, 60(4), 103377. <https://doi.org/10.1016/j.ipm.2023.103377>
- Feixiang, X. (2024). Intelligent Personalized Recommendation Method Based on Optimized Collaborative Filtering Algorithm in Primary and Secondary Education Resource System. *IEEE Access*, 12, 28860–28872. <https://doi.org/10.1109/access.2024.3365549>
- George, G., & Lal, A. M. (2024). PERKC: Personalized kNN With CPT for Course Recommendations in Higher Education. *IEEE Transactions on Learning Technologies*, 17, 885–892. <https://doi.org/10.1109/tlt.2023.3346645>
- Hao, P., Li, Y., & Bai, C. (2023). Meta-relationship for course recommendation in MOOCs. *Multimedia Systems*, 29(1), 235–246. <https://doi.org/10.1007/s00530-022-00989-5>
- Haque, M. M. ul, Kotaiah, B., & Ahamed, J. (2025). Hybrid course recommendation system using SVD, NMF and attention-based neural network. *International Journal of Information Technology*, 17(4), 2449–2456. <https://doi.org/10.1007/s41870-025-02440-0>
- Jiang, X., Bai, L., Yan, X., & Wang, Y. (2023). LDA-based online intelligent courses recommendation system. *Evolutionary Intelligence*, 16(5), 1619–1625. <https://doi.org/10.1007/s12065-022-00810-2>
- Kukkar, A., Mohana, R., Sharma, A., & Nayyar, A. (2024). A novel methodology using RNN + LSTM + ML for predicting student's academic performance. *Education and Information Technologies*, 29(11), 14365–14401. <https://doi.org/10.1007/s10639-023-12394-0>
- Li, M., Li, Z., Huang, C., Jiang, Y., & Wu, X. (2024). EduGraph: Learning Path-Based Hypergraph Neural Networks for MOOC Course Recommendation. *IEEE Transactions on Big Data*, 10(6), 706–719. <https://doi.org/10.1109/tbdata.2024.3453757>
- Ma, F. (2025). Learning behavior analysis and personalized recommendation system of online education platform based on machine learning. *Computers and Education: Artificial Intelligence*, 8, 100408. <https://doi.org/10.1016/j.caeai.2025.100408>
- Ma, W., Chen, W., Lu, L., & Fan, X. (2024). Integrating learners' knowledge background to improve course recommendation fairness: A multi-graph recommendation method based on contrastive learning. *Information Processing & Management*, 61(4), 103750. <https://doi.org/10.1016/j.ipm.2024.103750>
- Mansouri, N., Soui, M., & Abed, M. (2024). SFS feature selection with decision tree classifier for massive open online courses (MOOCs) recommendation. *Journal of Computers in Education*, 11(4), 1089–1110. <https://doi.org/10.1007/s40692-023-00291-x>
- Mrhar, K., & Abik, M. (2025). A deep learning framework for optimizing personalized online course recommendation and selection. *Decision Analytics Journal*, 16, 100616. <https://doi.org/10.1016/j.dajour.2025.100616>
- Shuiping, W. (2026). Personalised higher education course recommendation system based on generative AI. *Australian Journal of Electrical and Electronics Engineering*, 7(1), 1–15. <https://doi.org/10.1080/1448837x.2025.2603675>
- Siswipraptini, P. C., Warnars, H. L. H. S., Ramadhan, A., & Budiharto, W. (2024). Personalized Career-Path Recommendation Model for Information Technology Students in Indonesia. *IEEE Access*, 12, 49092–49105. <https://doi.org/10.1109/access.2024.3381032>
- Wang, X., Jia, L., Guo, L., & Liu, F. (2023). Multi-aspect heterogeneous information network for MOOC knowledge concept recommendation. *Applied Intelligence*, 53(10), 11951–11965. <https://doi.org/10.1007/s10489-022-04025-x>