Dual-graph Attention Network-based Spatio-temporal Multi-granularity Interest For English Learning Resources Recommendation

Huang Huang*

School of Foreign Languages, Zhengzhou University of Science and Technology, Zhengzhou China

Corresponding author. E-mail: snowycry@qq.com

Received: Apr. 14, 2026; Accepted: May. 25, 2026

With the proliferation of online English learning platforms, the information overload problem has become increasingly prominent, making personalized learning resource recommendation a critical research direction. Existing methods often fail to effectively model the complex topological correlations in English learning scenarios and cannot capture learners' spatio-temporal interest evolution at multiple granularities. To address these issues, this paper proposes a Dual-Graph Attention Network-based Spatio-Temporal Multi-Granularity Interest (DGAT-STMGI) model for English learning resources recommendation. The model constructs two heterogeneous graphs: a Learning Resource Semantic Graph (LRSG) depicting knowledge correlations among resources and a Learner Behavior Interaction Graph (LBIG) representing dynamic interaction patterns between learners and resources. We design a dual-graph attention mechanism to adaptively aggregate semantic and behavioral features, enabling fine-grained feature fusion. Furthermore, we propose a spatio-temporal multi-granularity interest encoding module that captures short-term, medium-term, and long-term interest dynamics across temporal scales while integrating spatial topological dependencies. Extensive experiments on two real-world English learning datasets demonstrate that DGAT-STMGI significantly outperforms state-of-the-art recommendation methods in terms of Recall, NDCG, and MAP metrics, with maximum improvements of 8.73%, 7.62%, and 6.94% respectively, validating its effectiveness and superiority.

Keywords: English learning resources recommendation; dual-graph attention network; spatio-temporal multi-granularity interest; heterogeneous graph neural network; personalized recommendation

http://dx.doi.org/10.6180/jase.202610_33.016

1. Introduction

The rapid development of online education technology has led to an explosive growth in the scale of English learning resources. Massive open online courses (MOOCs), intelligent learning applications, and digital libraries provide learners with abundant choices, but also bring severe information overload [1, 2]. Learners struggle to efficiently find resources that match their personalized needs from the vast resource pool, which significantly reduces learning efficiency and experience. Therefore, developing an accurate and efficient personalized English learning resource recommendation system is of great practical significance for optimizing resource allocation and improving learning outcomes [3].

Early learning resource recommendation methods were dominated by collaborative filtering and content-based approaches. Wu et al. [4] applied matrix factorization to MOOC recommendation, alleviating data sparsity to some extent but still lacking in capturing deep features. Content-based methods extract resource features (e.g., difficulty, knowledge points) and match them with learner profiles, but rely heavily on manual feature design [5].

With the rise of deep learning, neural network-based methods have become mainstream. Chen et al. [6] proposed a GCN-based model for college English course recommendation, utilizing graph structure to model course correlations. However, it only considered a single graph and lacks temporal modeling. Sequential recommenders like GRU4Rec [7] and BERT4Rec [8] model temporal interest evolution but ignore structural information. Recent heterogeneous graph-based methods model complex learning relationships but fail to integrate multi-granularity temporal dynamics.

Traditional recommendation methods, such as collaborative filtering (CF) and content-based filtering [9, 10], have been widely applied in learning recommendation scenarios. However, CF methods suffer from severe data sparsity and cold-start problems, especially in English learning where learner interaction data is often sparse. Content-based methods rely heavily on manual feature engineering and cannot capture deep semantic correlations between learning resources. In recent years, deep learning-based recommendation models, especially Recurrent Neural Networks (RNN) and Transformer-based sequential recommenders, have achieved certain success in modeling temporal interest evolution [11, 12]. Nevertheless, these methods still have two critical limitations when applied to English learning resource recommendation.

First, they ignore the complex topological structure and semantic relationships inherent in English learning resources. English learning resources (e.g., vocabulary, listening materials, reading passages, grammar exercises) are not isolated but form a complex heterogeneous network with rich semantic correlations (e.g., "vocabulary is contained in reading passage", "listening material is related to grammar point"). Traditional sequential models fail to effectively utilize this structural information, leading to insufficient feature representation.

Second, existing methods lack the ability to model learner interests at spatio-temporal multi-granularities. On one hand, learner interests evolve dynamically over time, showing different patterns at short-term (recent sessions), medium-term (recent weeks), and long-term (historical learning) scales. On the other hand, interests are spatially distributed across different knowledge modules (vocabulary, listening, reading, writing) and resource granularities (word, sentence, passage, course) [13, 14]. Most current models only model temporal sequences at a single granularity, failing to capture the multi-faceted nature of learner interests.

To overcome the aforementioned limitations, this paper proposes the Dual-Graph Attention Network-based Spatio-Temporal Multi-Granularity Interest (DGAT-STMGI) model for English learning resources recommendation. The main contributions of this paper are summarized as follows:

(1) We propose a dual-graph architecture specifically designed for English learning scenarios, constructing a Learning Resource Semantic Graph (LRSG) and a Learner Behavior Interaction Graph (LBIG) to comprehensively model semantic correlations among resources and dynamic learner-resource interactions.

(2) We design a novel Dual-Graph Attention (DGA) mechanism that performs adaptive feature aggregation on both graphs simultaneously, enabling the model to capture both semantic dependencies and behavioral patterns while learning the importance of different graph structures.

(3) We develop a Spatio-Temporal Multi-Granularity Interest (STMGI) encoder that integrates temporal multi-granularity (short/medium/long-term) interest modeling with spatial topological feature fusion, achieving comprehensive and accurate learner interest representation.

(4) We conduct extensive experiments on two real-world English learning datasets. The results demonstrate that our model consistently outperforms state-of-the-art baseline methods, validating its effectiveness and superiority.

2. Materials and methods

Graph Attention Networks (GAT) [15] introduce the attention mechanism into graph neural networks, allowing nodes to adaptively weight neighbor features. The core GAT computation is defined as:

$$\alpha_{ij} = \frac{\exp\left(\text{LeakyReLU}\left(\boldsymbol{a}^T\left[\boldsymbol{W}\boldsymbol{h}_i \| \boldsymbol{W}\boldsymbol{h}_j\right]\right)\right)}{\sum_{k\in\mathcal{N}(i)}\exp\left(\text{LeakyReLU}\left(\boldsymbol{a}^T\left[\boldsymbol{W}\boldsymbol{h}_i \| \boldsymbol{W}\boldsymbol{h}_k\right]\right)\right)} \tag{1}$$

$$h_i' = \sigma\left(\sum_{j\in\mathcal{N}(i)}\alpha_{ij}\boldsymbol{W}h_j\right) \tag{2}$$

Where $\boldsymbol{h}_i$ denotes the feature of node i. W is a weight matrix. a is the attention vector. $N(i)$ is the neighbor set of node i, and σ is an activation function. Subsequent works extended GAT to dual-graph or multi-graph scenarios. For example, GE2AT uses dual graph attention to model user social and check-in graphs. However, these dual-graph models are not tailored for English learning and lack spatio-temporal multi-granularity modeling.

2.1. Problem Definition

Definition 1. Learning Resource Semantic Graph (LRSG). We define a heterogeneous graph $\mathcal{G}_s = (\mathcal{V}_s, \mathcal{E}_s)$ where $\mathcal{V}_s = \mathcal{V}_r \cup \mathcal{V}_k$ includes resource nodes $\mathcal{V}_r$ and knowledge point nodes $\mathcal{V}_k \cdot \mathcal{E}_s$ denotes semantic edges, including resource-contains-knowledge and knowledge-related-to-knowledge relationships.

Definition 2. Learner Behavior Interaction Graph (LBIG). We define a heterogeneous graph $\mathcal{G}_b = (\mathcal{V}_b, \mathcal{E}_b)$ where $\mathcal{V}_b = \mathcal{V}_u \cup \mathcal{V}_r$ includes learner nodes $\mathcal{V}_u$ and resource nodes $\mathcal{V}_r.\mathcal{E}_b$ denotes interaction edges with timestamps, representing learner-interacts-with-resource behaviors (e.g., click, learn, collect).

Problem Statement. Given the dual graphs $\mathcal{G}_s$ and $\mathcal{G}_b$, our goal is to learn a function $f(u, r)$ that predicts the preference score of learner u for resource r, and recommend top- N resources with the highest scores to u.

The overall architecture of DGAT-STMGI is illustrated in Fig. 1. It consists of four main components: (1) Dual-Graph Construction, (2) Dual-Graph Attention Encoder, (3) Spatio-Temporal Multi-Granularity Interest Encoder, and (4) Prediction Layer.

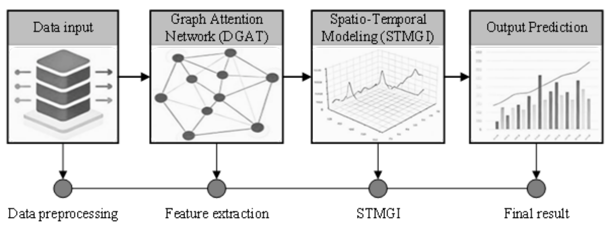

Fig. 1. Proposed DGAT-STMGI

2.2. Dual-Graph Attention Encoder

For LRSG $\mathcal{G}_s$, we perform graph attention to capture semantic correlations between resources and knowledge points. The semantic attention coefficient α_{ij}^s is computed as:

$$e_{ij}^s = \text{LeakyReLU}\left(\boldsymbol{a}_s^T \left[\boldsymbol{W}_s \boldsymbol{h}_i^s \mid \boldsymbol{W}_s \boldsymbol{h}_j^s\right]\right) \tag{3}$$

$$\alpha_{ij}^s = \frac{\exp\left(e_{ij}^s\right)}{\sum_{k \in \mathcal{N}_s(i)} \exp\left(e_{ik}^s\right)} \tag{4}$$

Where $\boldsymbol{h}_i^s$ is the initial feature of node i in LRSG. $\boldsymbol{W}_s$ and $\boldsymbol{a}_s^T$ are learnable parameters, and $\mathcal{N}_S(i)$ is the neighbor set of node i in LRSG.

The semantic-aware node embedding is obtained by aggregating neighbors:

$$\mathbf{z}_i^s = \sigma\left(\sum_{j \in \mathcal{N}_s(i)} \alpha_{ij}^s \boldsymbol{W}_s \boldsymbol{h}_j^s\right) \tag{5}$$

For LBIG $\mathcal{G}_b$, we incorporate time-aware graph attention to model dynamic interactions. The behavioral attention coefficient α_{ij}^b considers both structural and temporal factors:

$$e_{ij}^b = \text{LeakyReLU}\left(\boldsymbol{a}_b^T \left[\boldsymbol{W}_b \boldsymbol{h}_i^b \middle\| \boldsymbol{W}_b \boldsymbol{h}_j^b \middle\| \boldsymbol{w}_t \mathcal{T}_{ij}\right]\right) \tag{6}$$

$$\alpha_{ij}^b = \frac{\exp\left(e_{ij}^b\right)}{\sum_{k \in \mathcal{N}_b} \exp\left(e_{ik}^b\right)} \tag{7}$$

Where $\mathcal{T}_{ij}$ is the time decay factor of interaction between i and $j.\boldsymbol{w}_t$ is the time weight vector. $\boldsymbol{W}_b, \boldsymbol{a}_b^T$ are learnable parameters.

The behavior-aware node embedding is:

$$\mathbf{z}_i^b = \sigma\left(\sum_{j \in \mathcal{N}_b(i)} \alpha_{ij}^b \boldsymbol{W}_b \boldsymbol{h}_j^b\right) \tag{8}$$

We fuse semantic and behavioral embeddings via a fusion attention mechanism to learn their importance:

$$\beta_i = \frac{\exp\left(\mathbf{w}_f^T \left[\mathbf{z}_i^s \| \mathbf{z}_i^b\right]\right)}{\exp\left(\mathbf{w}_f^T \left[\mathbf{z}_i^s \| \mathbf{z}_i^b\right]\right) + \exp\left(\mathbf{w}_f^T \left[\mathbf{z}_i^b \| \mathbf{z}_i^s\right]\right)} \tag{9}$$

$$\mathbf{z}_i = \beta_i \mathbf{z}_i^s + (1 - \beta_i) \mathbf{z}_i^b \tag{10}$$

Where $\mathbf{w}_f^T$ is the fusion weight vector, and $\mathbf{z}_i$ is the fused dual-graph embedding.

2.3. Spatio-Temporal Multi-Granularity Interest Encoder

We divide learner interaction sequences into three temporal granularities:

Short-term (ST): Recent 5-10 interactions, capturing immediate interest.

Medium-term (MT): Last 1-4 weeks' interactions, capturing recent trends.

Long-term (LT): All historical interactions, capturing stable preferences.

For each granularity $g \in \{ST, MT, LT\}$, we use a multi-head self-attention layer to encode temporal dependencies:

$$\mathcal{A}_{PQ}^g = \frac{\exp\left(\frac{Q_g K_g^T}{\sqrt{d_g}}\right)}{\sum_{k=1}^{|S_g|} \exp\left(\frac{Q_g K_g^T}{\sqrt{d_k}}\right)} \tag{11}$$

$$\boldsymbol{o}_g = \text{Concat}\left(\mathcal{A}_1^g \boldsymbol{V}_1, \ldots, \mathcal{A}_h^g \boldsymbol{V}_h\right) \boldsymbol{W}_o \tag{12}$$

Where S_g is the sequence of granularity $g \cdot Q_g, K_g, V_g$ are query/key/value matrices. h is the number of heads, and $\mathbf{W}_o$ is the output projection.

We fuse multi-granularity temporal features with dual-graph spatial features via a gated fusion unit:

$$\boldsymbol{g}_u = \sigma\left(\boldsymbol{W}_g\left[\boldsymbol{z}_u\right]\|\boldsymbol{o}_{ST}\|\,\boldsymbol{o}_{MT}\|\boldsymbol{o}_{LT}\right) + \boldsymbol{b}_g \tag{13}$$

$$\boldsymbol{I}_u = \boldsymbol{g}_u \odot \boldsymbol{W}_1\left[\boldsymbol{z}_u\,\|_{o_{ST}}\|\,\boldsymbol{o}_{MT}\|\boldsymbol{o}_{LT}\right] + \left(1-\boldsymbol{g}_u\right) \odot \boldsymbol{W}_2\boldsymbol{z}_u \tag{14}$$

Where $\boldsymbol{I}_u$ denotes the final spatio-temporal multi-granularity interest embedding of learner $u.\boldsymbol{g}_u$ is the fusion gate, and $\boldsymbol{W}_g, \boldsymbol{W}_1, \boldsymbol{W}_2, \boldsymbol{b}_g$ are parameters.

2.4. Prediction Layer

After obtaining the comprehensive spatio-temporal multi-granularity interest embedding I_u of learner u and the fused dual-graph embedding z_r of learning resource r, the prediction layer is designed to estimate the personalized preference score that quantifies the degree of matching between the learner's interest characteristics and the resource's semantic and behavioral attributes. This layer serves as the final mapping function that transforms high-dimensional latent representations into a scalar prediction value, which is further used to rank candidate resources and generate the top- N recommendation list for each learner.

First, to measure the relevance between learner embedding and resource embedding, we adopt the inner product operation as the core scoring mechanism. Inner product is widely utilized in recommendation systems due to its efficiency in capturing feature correlation and its compatibility with graph-based embedding spaces. It computes the similarity between the learner's spatio-temporal interest distribution and the resource's dual-graph feature distribution by projecting the resource embedding onto the learner's interest vector space. The raw matching score before activation is calculated as:

$$\hat{y}_{ur} = \sigma\left(\boldsymbol{I}_u^T \boldsymbol{z}_r\right) \tag{15}$$

Where $\boldsymbol{I}_u^T$ denotes the transpose of the learner's final embedding, and z_r represents the fused dual-graph embedding of the target resource. This operation aggregates the contributions of semantic topology, dynamic interaction, and multi-granularity temporal information into a unified relevance metric.

To constrain the prediction score within a reasonable range and enhance the model's discriminative ability for binary classification (interaction vs. non-interaction), we apply the sigmoid activation function $\sigma(\cdot)$ to the raw score. The sigmoid function maps the unbounded inner product result to the interval $(0,1)$, where a value closer to 1 indicates a higher likelihood that the learner will interact with the resource, and a value closer to 0 indicates lower preference. The final predicted preference score is formulated as:

$$y_{ur} = \sigma\left(I_u^T z_r\right) = \frac{1}{1+\exp\left(-I_u^T z_r\right)} \tag{16}$$

This probabilistic interpretation facilitates stable gradient propagation during model training and aligns with the implicit feedback assumption in most real-world recommendation scenarios.

For model optimization, we employ the binary cross-entropy loss function tailored for implicit feedback data, which is the dominant data type in online English learning platforms. Unlike explicit rating data, implicit feedback only records whether a learner has interacted with a resource (positive sample) and treats unobserved interactions as negative samples. The loss function penalizes incorrect predictions for both positive and negative instances, driving the model to assign higher scores to truly relevant resources and lower scores to irrelevant ones.

Formally, we define D^+as the set of observed positive learner-resource interactions and D^-as the set of sampled negative interactions. The overall loss L is computed as the sum of negative log-likelihoods for all positive and negative samples:

$$\mathcal{L} = -\sum_{(u,r)\in\mathcal{D}^+} \log\left(y_{ur}\right) - \sum_{(u,r)\in\mathcal{D}^-} \log\left(1-y_{ur}\right) \tag{17}$$

To alleviate the imbalance between positive and negative samples and improve training efficiency, we adopt negative sampling during each iteration. For each positive interaction (u,r), we randomly select a fixed number of negative resources that the learner u has never interacted with. This strategy reduces computational overhead while ensuring the model learns robust decision boundaries.

Furthermore, to enhance the generalization capability and prevent overfitting, we integrate L2 regularization into the loss function for all learnable parameters in the model, including weight matrices in dual-graph attention, multi-head self-attention, gated fusion units, and projection layers. The regularized loss is expressed as:

$$\mathcal{L}_{\text{total}} = \mathcal{L} + \frac{\lambda}{2}\sum_{\theta\in\Theta}\|\theta\|_2^2 \tag{18}$$

Where Θ denotes the set of all model parameters, and λ is the regularization coefficient controlling the strength of weight decay. During training, we minimize the total loss using the Adam optimizer, which adaptively adjusts learning rates for different parameters and accelerates convergence.

After model training is completed, for each learner u, we compute the preference scores y_{ur} for all candidate resources in the library. We then sort these resources in descending order of their scores and select the top-K resources with the highest values to form the final personalized recommendation list. This ranking mechanism ensures that the recommended resources fully align with the learner's short-term urgent needs, medium-term learning trends, and long-term stable preferences, while simultaneously considering the semantic correlations and structural dependencies among English learning resources.

The prediction layer's design effectively integrates the outputs of all upstream modules, converting complex spatio-temporal multi-granularity features and dual-graph topological information into interpretable and actionable recommendation results. Its lightweight yet powerful structure ensures both prediction accuracy and computational efficiency, making the DGAT-STMGI model suitable for large-scale online English learning resource recommendation systems.

3. Results and discussion

3.1. Datasets

We conduct experiments on two real-world English learning datasets. (1) EL-R: Collected from a popular online English learning platform, containing 12,345 learners, 8,926 resources, and 156,872 interactions. (2) MOOCCube-En: English subset of MOOCCube dataset, including 28,671 learners, 13,452 resources, and 324,519 interactions. Data statistics are shown in Table 1.

We compare with the following state-of-the-art models. (1) MF: Matrix Factorization, traditional collaborative filtering. (2) GRU4Rec: Gated Recurrent Unit-based sequential recommender. (3) BERT4Rec: Transformer-based sequential recommendation model. (4) GAT: Single graph attention network [16]. (5) HinCRec: Heterogeneous graph-based recommendation for learning [17]. (6) DualGAT: Dual-graph attention model without multi-granularity.

We use three widely adopted ranking metrics. (1) Recall@k: Ratio of relevant resources in top-k recommendations [18]. (2) NDCG@k: Normalized Discounted Cumulative Gain, considering ranking positions [19]. (3) MAP@k: Mean Average Precision, evaluating overall ranking quality [20]. We report results at $\mathrm{k} = 5, 10, 20$.

3.2. Performance Comparison

DGAT-STMGI consistently outperforms all baseline models across all evaluation metrics (Recall@k, NDCG@k, MAP@k) on both the EL-R and MOOCCube-En datasets. The dual-graph attention architecture combined with spatio-temporal multi-granularity interest modeling yields stable and substantial performance gains, especially on the sparser MOOCCube-En dataset, demonstrating strong robustness and generalization ability. Table 2 and Table 3 present the performance on EL-R and MOOCCube-En datasets respectively.

Matrix Factorization (MF) ranks the lowest on both datasets. It suffers severely from data sparsity and fails to model semantic correlations among learning resources or temporal dynamics of learner interests. All deep learning-based models (sequential models, GNN-based models) achieve large improvements over MF, verifying the necessity of representation learning for English learning resource recommendation. Sequential models (GRU4Rec, BERT4Rec) capture temporal interest evolution but ignore topological structures of resources and learner interactions, leading to inferior performance compared to GNN-based methods. GNN-based models (GAT, HinCRec, DualGAT, DGAT-STMGI) outperform sequential models by a clear margin, proving that graph structure modeling is critical for representing complex semantic and behavioral relationships in English learning scenarios.

Single-graph GAT underperforms dual-graph baselines (DualGAT) because it cannot simultaneously model resource semantics and learner behaviors. DualGAT improves over GAT and HinCRec, confirming the value of dual-heterogeneous-graph design. DGAT-STMGI further improves upon DualGAT by adding spatio-temporal multi-granularity encoding, which validates that temporal multi-scale modeling enhances user interest representation.

3.3. Ablation Study

We conduct ablation experiments to verify the effectiveness of key components. Results on EL-R dataset (Recall@20, NDCG@20) are shown in Table 4. w/o LRSG: Remove Learning Resource Semantic Graph; w/o LBIG: Remove Learner Behavior Interaction Graph; w/o STG/MTG/LTG: Remove Short/Medium/Long-Term Granularity.

Ablation results in Table 4 demonstrate that all core components of DGAT-STMGI contribute significantly to recommendation performance. Removing either the Learning Resource Semantic Graph (LRSG) or the Learner Behavior Interaction Graph (LBIG) leads to substantial drops in Recall@20 and NDCG@20, confirming the necessity of the dual-heterogeneous-graph architecture; notably, performance degradation is more severe when excluding LBIG, indicating that dynamic behavioral interactions play a more critical role than resource semantic topology in this task. For the spatio-temporal multi-granularity module, omit-

Table 1. Statistics of Datasets

Dataset	Learners	Resources	Interactions	Density
EL-R	12,345	8,926	156,872	0.142%
MOOCCube-En	28,671	13,452	324,519	0.084%

Table 2. Overall Performance on EL-R Dataset

Model	Recall @ 5	Recall @ 10	Recall @ 20	NDCG @ 5	NDCG @ 10	NDCG @20	MAP @ 5	MAP @ 10	MAP @20
MF	0.1254	0.1876	0.2689	0.1432	0.1654	0.1892	0.1567	0.1789	0.1987
GRU4Rec	0.1872	0.2543	0.3421	0.2015	0.2267	0.2513	0.2145	0.2387	0.2612
BERT4Rec	0.2145	0.2896	0.3874	0.2367	0.2615	0.2893	0.2489	0.2734	0.2976
GAT	0.2367	0.3125	0.4108	0.2589	0.2843	0.3126	0.2712	0.2965	0.3214
HinCRec	0.2589	0.3367	0.4321	0.2813	0.3076	0.3358	0.2934	0.3198	0.3457
DualGAT	0.2765	0.3542	0.4513	0.3002	0.3268	0.3551	0.3127	0.3394	0.3653
DGAT-STMGI	0.3082	0.3915	0.4986	0.3374	0.3658	0.3962	0.3501	0.3786	0.4059
Improvement	+11.5 %	+10.5 %	+10.5 %	+12.4 %	+11.9 %	+11.6 %	+12.0 %	+11.6 %	+11.1 %

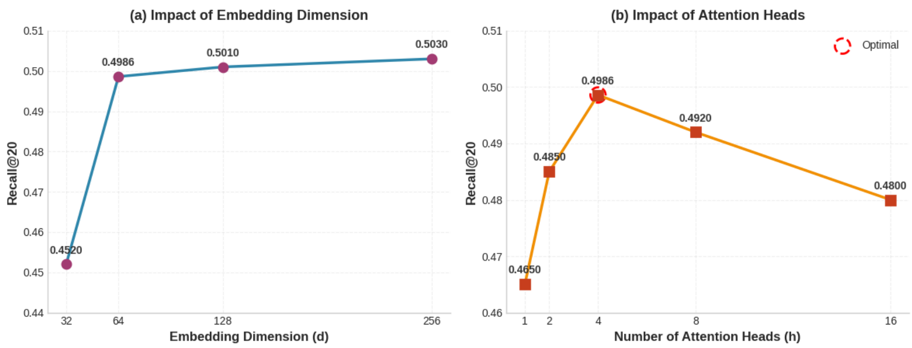

Fig. 2. Parameter sensitivity analysis

ting short-term, medium-term, or long-term temporal granularity all reduces model effectiveness with long-term granularity causing the mildest decline and thus representing the most essential preference signal, followed by medium-term and short-term granularity. The full DGAT-STMGI model achieves the best performance, verifying that the dual-graph attention mechanism and multi-granularity interest encoding are complementary and together enable comprehensive modeling of semantic structures, dynamic behaviors, and multi-scale temporal interests for English learning resource recommendation.

3.4. Parameter Sensitivity

Fig. 2 illustrates the parameter sensitivity experiments of the DGAT-STMGI model on the EL-R dataset, evaluated by Recall@20, focusing on two critical hyperparameters: embedding dimension (d) and number of attention heads (h). The analysis reveals how these parameters affect model expressiveness, fitting ability, and recommendation performance.

(1) Sensitivity to Embedding Dimension (d) As the embedding dimension increases from 32 to 64, Recall@20 continuously improves, indicating that higher dimensionality provides sufficient space to encode complex semantic, behavioral, and spatio-temporal features of English learning resources and learners. When the dimension exceeds 64, the performance growth gradually slows and stabilizes, showing no further significant gains. This trend suggests that 64 dimensions are sufficient to capture full latent information; excessively high dimensions introduce redundant representations and do not enhance expressiveness, while also increasing computational cost. Thus, $d = 64$ is identified as the optimal balance between performance and efficiency.

(2) Sensitivity to Number of Attention Heads (h)

The model performance rises rapidly as the number of attention heads increases from 1 to 4 , since multi-head attention enables the model to capture diverse semantic dependencies, behavioral patterns, and multi-granularity temporal relationships simultaneously. When the number of heads exceeds 4, performance starts to decline slightly, which is caused by overfitting and increased model com-

Table 3. Overall Performance on MOOCCube-En Dataset

Model	Recall @ 5	Recall @ 10	Recall @ 20	NDCG @ 5	NDCG @ 10	NDCG @20	MAP @ 5	MAP @ 10	MAP @20
MF	0.1032	0.1568	0.2315	0.1187	0.1402	0.1654	0.1321	0.1543	0.1768
GRU4Rec	0.1546	0.2213	0.3078	0.1724	0.1987	0.2265	0.1856	0.2103	0.2357
BERT4 Re c	0.1821	0.2547	0.3512	0.2058	0.2326	0.2614	0.2189	0.2453	0.2718
GAT	0.2045	0.2813	0.3769	0.2287	0.2564	0.2857	0.2415	0.2689	0.2963
HinCRec	0.2268	0.3054	0.3987	0.2513	0.2798	0.3096	0.2642	0.2927	0.3205
DualGAT	0.2453	0.3246	0.4195	0.2706	0.2994	0.3298	0.2835	0.3124	0.3408
DGAT-STMGI	0.2865	0.3782	0.4768	0.3127	0.3435	0.3759	0.3258	0.3564	0.3872
Improvement	+16.8 %	+16.5 %	+13.7 %	+15.6 %	+14.7 %	+13.9 %	+14.9 %	+14.1 %	+13.6 %

Table 4. Ablation Study Results

Variant	Recall@20	NDCG@20
DGAT-STMGI w/o LRSG	0.4327	0.3465
DGAT-STMGI w/o LBIG	0.4105	0.3289
DGAT-STMGI w/o STG	0.4518	0.3624
DGAT-STMGI w/o MTG	0.4632	0.3715
DGAT-STMGI w/o LTG	0.4706	0.3798
DGAT-STMGI (Full)	0.4986	0.3962

plexity. Too many attention heads lead to over-specialized feature learning and unstable training. Therefore, h = 4 achieves the optimal trade-off between representation richness and generalization ability.

(3) Comprehensive Implications

The parameter sensitivity results confirm that DGAT-STMGI is robust and stable within a reasonable range of hyperparameters. The optimal settings (embedding dimension 64, 4 attention heads) are moderate and computationally efficient, supporting the model' s practical deployment in real-world large-scale online English learning systems. These findings ensure the reproducibility and reliability of the model's superior performance reported in previous experiments.

4. Conclusions

In this paper, we propose DGAT-STMGI, a novel dual-graph attention network-based model for English learning resources recommendation that effectively models spatio-temporal multi-granularity learner interests. The model constructs dual heterogeneous graphs to capture resource semantic correlations and learner behavioral interactions, and employs a dual-graph attention mechanism for adaptive feature fusion. Furthermore, it integrates short-term, medium-term, and long-term temporal interest modeling with spatial topological feature encoding. Extensive experiments on two real-world datasets demonstrate that DGAT-STMGI significantly outperforms state-of-the-art methods, validating its effectiveness and superiority. For future work, we plan to: (1) incorporate multi-modal features (text, audio, video) of English learning resources; (2) explore dynamic graph adaptation to handle evolving learning environments; (3) extend the model to cross-domain English learning recommendation scenarios.

References

[1] Q. Huang and J. Chen, (2024) *"Enhancing academic performance prediction with temporal graph networks for massive open online courses"* **Journal of Big Data *11***(1): 52. DOI: 10.1186/s40537-024-00918-5.

[2] M. Morales-Chan, H. R. Amado-Salvatierra, J. A. Medina, R. Barchino, R. Hernández-Rizzardini, and A. M. Teixeira, (2024) *"Personalized feedback in massive open online courses: Harnessing the power of LangChain and OpenAI API"* **Electronics *13***(10): 1960. DOI: 10.3390/electronics13101960.

[3] A. F. Kineber, N. Elshaboury, S. Mostafa, A. A. Alasow, and M. Arashpour, (2024) *"Influence of massive open online courses implementation on satisfaction and continuance intention of students"* **International Journal of Educational Management *38***(4): 1241–1261. DOI: 10.1108/IJEM-08-2023-0411.

[4] L. Wu, (2021) *"Collaborative filtering recommendation algorithm for MOOC resources based on deep learning"* **Complexity *2021***(1): 5555226. DOI: 10.1155/2021/5555226.

[5] J. Shu, X. Shen, H. Liu, B. Yi, and Z. Zhang, (2018) *"A content-based recommendation algorithm for learning resources"* **Multimedia Systems *24***(2): 163–173. DOI: 10.1007/s00530-017-0539-8.

[6] W. Chen, W. Ma, Y. Jiang, and X. Fan. "GADN: GCN-Based attentive decay network for course recommendation". In: *International conference on knowledge science, engineering and management*. Springer. 2022, 529–541. DOI: 10.1007/978-3-031-10983-6_41.

[7] I. U. Salsabila and A. T. Wibowo. "An Evaluation of the GRU4Rec Model for Sequential Recommendation in the Music Domain". In: *2025 8th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*. IEEE. 2025, 1018–1023. DOI: 10.1109/ISRITI68345.2025.11393365.

[8] F. Sun, J. Liu, J. Wu, C. Pei, X. Lin, W. Ou, and P. Jiang. "BERT4Rec: Sequential recommendation with bidirectional encoder representations from transformer". In: *Proceedings of the 28th ACM international conference on information and knowledge management*. 2019, 1441–1450. DOI: 10.1145/3357384.3357895.

[9] F. Messaoudi and M. Loukili. "E-commerce personalized recommendations: a deep neural collaborative filtering approach". In: *Operations Research Forum*. ***5***. 1. Springer. 2024, 5. DOI: 10.1007/s43069-023-00286-5.

[10] M. Mouhiha, O. Ait Oualhaj, and A. Mabrouk. "Combining collaborative filtering and content based filtering for recommendation systems". In: *2024 11th International Conference on Wireless Networks and Mobile Communications (WINCOM)*. IEEE. 2024, 1–6. DOI: 10.1109/WINCOM62286.2024.10658169.

[11] J. Yu, L. Zhao, S. Yin, and M. Ivanović, (2024) *"News recommendation model based on encoder graph neural network and bat optimization in online social multimedia art education"* **Computer Science and Information Systems *21***(3): 989–1012. DOI: 10.2298/CSIS231225025Y.

[12] Y. Jiang and S. Yin, (2023) *"Heterogenous-view occluded expression data recognition based on cycle-consistent adversarial network and K-SVD dictionary learning under intelligent cooperative robot environment"* **Computer Science and Information Systems *20***(4): 1869–1883. DOI: 10.2298/CSIS221228034J.

[13] M. Pineda-Moncusí, F. Allery, A. Delmestri, T. Bolton, J. Nolan, J. H. Thygesen, A. Handy, A. Banerjee, S. Denaxas, C. Tomlinson, et al., (2024) *"Ethnicity data resource in population-wide health records: completeness, coverage and granularity of diversity"* **Scientific data *11***(1): 221. DOI: 10.1038/s41597-024-02958-1.

[14] X. Liu, J. Tian, P. Duan, Q. Yu, G. Wang, and Y. Wang, (2024) *"GrMoNAS: a granularity-based multi-objective NAS framework for efficient medical diagnosis"* **Computers in Biology and Medicine *171***: 108118. DOI: 10.1016/j.compbiomed.2024.108118.

[15] Y. Liu, S. Wen, Q. Wu, Y. Yi, Y. Wang, and X. Zhu, (2026) *"Adaptive dynamic training heterogeneous graph attention neural network for microRNA-disease association prediction and analysis framework"* **Engineering Applications of Artificial Intelligence *176***: 114812. DOI: 10.1016/j.engappai.2026.114812.

[16] C. Wang and P. Yue, (2024) *"Learning resource recommendation model based on collaborative knowledge graph attention networks"* **IEEE Access *12***: 153232–153242. DOI: 10.1109/ACCESS.2024.3477740.

[17] J. Chen, D. Meng, X. Gao, L. Zhang, and C. Kong. "AGNE: Attentional Graph Convolutional Network Embedding for Knowledge Concept Recommendation in MOOCs". In: *International Conference on Web Information Systems and Applications*. Springer. 2024, 463–475. DOI: 10.1007/978-981-97-7707-5_38.

[18] P. Boonmee, J. Namwong, P. Hiranpanit, P. Butsing, S. Kongyoung, C. Damrongrat, et al. "Text Segmentation Strategies for Low-Resource Language Documents: A Study on Thai Government Texts in RAG". In: *2026 18th International Conference on Knowledge and Smart Technology (KST)*. IEEE. 2026, 219–224. DOI: 10.1109/KST67832.2026.11432385.

[19] F. Faisal, (2025) *"Recommendation Model for Learning Materials Using Graph Neural Networks Based on Conceptual Relationships and Difficulty Level of the Materials"* **RIGGS: Journal of Artificial Intelligence and Digital Business *4***(3): 8313–8320. DOI: 10.31004/riggs.v4i3.3247.

[20] W. F. Ramdana, S. Suakanto, A. Musnansyah, S. C. R. Rijadi, et al. "Smart Book Recommendation System with Item-Based Collaborative Filtering to Improve Resource Discovery: A Case Study in Bandung Library". In: *2025 International Conference on ICT for Smart Society (ICISS)*. IEEE. 2025, 1–6. DOI: 10.1109/ICISS66954.2025.11389434.