ARTICLE

HMGS: Hierarchical Matching Graph Neural Network for Session-Based Recommendation

Pengfei Zhang[1], Rui Xin[1], Xing Xu[1], Yuzhen Wang[1], Xiaodong Li[2], Xiao Zhang[2], Meina Song[2] and Zhonghong Ou[3,*]

[1]State Grid Hebei Information and Telecommunication Branch, Shijiazhuang, 050000, China
[2]School of Computer Science (National Pilot Software Engineering School), Beijing University of Posts and Telecommunications, Beijing, 100876, China
[3]State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing, 100876, China
*Corresponding Author: Zhonghong Ou. Email: zhonghong.ou@bupt.edu.cn

ABSTRACT: Session-based recommendation systems (SBR) are pivotal in suggesting items by analyzing anonymized sequences of user interactions. Traditional methods, while competent, often fall short in two critical areas: they fail to address potential inter-session item transitions, which are behavioral dependencies that extend beyond individual session boundaries, and they rely on monolithic item aggregation to construct session representations. This approach does not capture the multi-scale and heterogeneous nature of user intent, leading to a decrease in modeling accuracy. To overcome these limitations, a novel approach called HMGS has been introduced. This system incorporates dual graph architectures to enhance the recommendation process. A global transition graph captures latent cross-session item dependencies, while a heterogeneous intra-session graph encodes multi-scale item embeddings through localized feature propagation. Additionally, a multi-tier graph matching mechanism aligns user preference signals across different granularities, significantly improving interest localization accuracy. Empirical validation on benchmark datasets (Tmall and Diginetica) confirms HMGS's efficacy against state-of-the-art baselines. Quantitative analysis reveals performance gains of 20.54% and 12.63% in Precision@10 on Tmall and Diginetica, respectively. Consistent improvements are observed across auxiliary metrics, with MRR@10, Precision@20, and MRR@20 exhibiting enhancements between 4.00% and 21.36%, underscoring the framework's robustness in multi-faceted recommendation scenarios.

KEYWORDS: Session-based recommendation; graph network; multi-level matching

1 Introduction

Recommendation systems are pivotal in facilitating efficient and personalized user decision-making processes. However, access to longitudinal user interaction histories and comprehensive profile data is frequently constrained in practical applications [1,2], leading to compromised efficacy in conventional recommendation frameworks. This limitation has driven significant research interest in session-based recommendation (SBR) [3–5], a paradigm that predicts subsequent user engagements by analyzing anonymized, temporally contiguous interaction sequences.

Current SBR methodologies are classified into three categories: heuristic similarity-driven approaches, recurrent neural network (RNN)-based frameworks, and graph neural network (GNN)-based architectures. Heuristic techniques, exemplified by co-occurrence-centric models [6], prioritize intra-session item

adjacency metrics while disregarding temporal sequential dependencies. In contrast, RNN-based implementations such as GRU4Rec [7] and NARM [8] operationalize sessions as chronologically ordered item sequences to infer user intent. Recent advancements have introduced GNN-based strategies [6,9–11], which model sessions as graph structures where nodes denote items and edges encode transitional relationships. State-of-the-art GNN implementations focus on capturing intra-session transition dynamics through iterative message-passing mechanisms; these transitions are subsequently aggregated to derive session-level embeddings. Empirical evaluations across standardized benchmarks confirm the superiority of GNN-based architectures over heuristic and sequential models, attributable to their capacity to represent complex relational patterns.

Beyond predominant GNN-based methodologies, notable contributions to session-based recommendation (SBR) systems have been made through non-graph approaches. Memory network architectures and attention-driven models, for instance, have been enhanced to address data sparsity challenges inherent in session-based paradigms [6,12]. Though less effective than GNNs in modeling intricate inter-item relational dynamics, such techniques remain critical when session data exhibits structural incompatibility with graph representations. The HMGS, however, has been proposed to surpass conventional non-GNN SBR methods. Diverging from traditional systems dependent on simplistic item correlation metrics or linear latent embeddings derived from matrix factorization [12], HMGS incorporates a multi-tiered graph alignment mechanism to dynamically model interaction hierarchies. This framework enables granular behavioral analysis by synthesizing both atomic interaction events and emergent multi-level session patterns, thereby capturing user intent across varying contextual scales.

Persistent limitations in existing methodologies are identified in two principal aspects. First, inter-session item transition dynamics—behavioral dependencies extending beyond individual session boundaries—are frequently disregarded. SBR's reliance on transient interaction sequences exacerbates data paucity due to absent longitudinal behavioral context, rendering isolated intra-session analysis insufficient for robust preference inference. Although solutions such as CSRM [12] and GCE-GNN [6] employ global graphs to mitigate this, CSRM's undifferentiated aggregation of historical sessions introduces noise through the inclusion of behaviorally incongruent data. GCE-GNN, while analogous, inadequately models heterogeneous behavioral influences by prioritizing homogeneous interaction patterns, thus failing to accommodate diverse preference signals. Second, while individual items reflect atomic user intent, holistic session semantics remain underutilized. Through hierarchical session decomposition, multi-scale behavioral patterns can be extracted with heightened precision, enabling fine-grained intent representation.

To surmount these problems, a novel HMGS is introduced for the session-based recommendation, designed to holistically model multi-granular user behavioral intent. The framework is structured as follows: first, a global graph layer is implemented to capture cross-session transitional dependencies, thereby mitigating the aforementioned challenge. Second, to disentangle heterogeneous and multi-scale user preferences, variable-length session subsequences are analyzed to derive hierarchical intent representations. To harmonize these dual components—the global graph for cross-session behavioral correlations and the local graph for intra-session contextual enrichment—a hierarchical graph matching mechanism is integrated, aligning candidate items with session embeddings across granularity levels. Finally, a predictive layer is incorporated to estimate probabilities of subsequent user interactions. The principal contributions of this work are delineated as follows:

(i)   A hierarchical session graph matching framework (HMGS) is introduced, enabling precise user interest localization through multi-level alignment of candidate items with session representations.

(ii) Dual architectural components are developed: a global graph layer to model cross-session behavioral synergies and a local graph layer to augment contextual interaction modeling. These layers collectively address two foundational SBR challenges.

(iii) Empirical validation of HMGS on two benchmark datasets demonstrates state-of-the-art performance. Ablation studies and hyperparameter sensitivity analyses further substantiate the efficacy of the proposed components.

2 Related Work

In the context of session-based recommendations, which are confined to anonymous sessions, that is, users' short-term interactions, several main research directions within the field of sessional recommendation are reviewed in this section.

2.1 Rnn-based Session Recommendation

Since RNN can effectively model sequential information, RNN-based methods have been used in SBR for many years. GRU4REC proposed by Hidasi et al. [7] applies the RNN network to SBR for the first time. It uses a multi-layer gated recurrent unit (GRU) to model item interaction sequences to learn item representations. Later, Tan et al. [2] introduced data augmentation based on GRU4REC to further improve the performance of the model. The NARM proposed by Li et al. [8] integrates the attention mechanism into the stack GRU encoder and adjusts the weights to capture item transitions information that is more representative of the user's short-term interests. Liu et al. [13] proposed an attention-based short-term memory network STAMP to replace RNN to capture users' short-term interests. Among them, NARM and STAMP both emphasize the importance of the last click by using the attention mechanism. SASRec [14] captures correlations between different projects by stacking multiple layers. ISLF [3] considers the situation of user interest transfer, and uses variational autoencoder (VAE) and RNN to achieve the purpose of accurately capturing the user's sequence behavior characteristics. However, RNN-based methods focus more on modeling the sequential transition of adjacent items and infer user preferences by analyzing the temporal order of a given sequence, and therefore have a poor ability to capture information about more complex item transitions (e.g., non-adjacent item transitions).

2.2 Gnn-based Session Recommendation

In recent years, Graph Neural Networks (GNNs) have emerged as a leading approach in session-based recommendation systems due to their ability to model complex user-item interactions. The early work of SRGNN [15] was the first to apply GNNs in session modeling. By representing each session as a directed, unweighted graph, SRGNN used a gating mechanism to capture session representations, effectively modeling short-range item transitions within sessions. While SRGNN laid the groundwork for session-based recommendation, its reliance on a simple session graph limited its ability to capture long-range dependencies between items.

To address this limitation, GC-SAN [10] enhanced SRGNN by incorporating an attention mechanism, which allowed the model to better capture long-distance item dependencies across a session. The attention mechanism enabled the model to dynamically focus on more relevant items, improving performance. However, while GC-SAN effectively captures item dependencies within a session, it still faces challenges in modeling the broader contextual shifts in user behavior across sessions.

Building upon this idea, Xia et al. introduced DHCN [11], which further advanced session-based recommendation by leveraging a dual-channel hypergraph. This hypergraph captures super-pairwise relationships and integrates a self-supervised learning process. By enhancing the mutual information between two session

representations, DHCN improved the model's ability to capture more complex patterns of user behavior. However, DHCN still relies on traditional graph structures, which may struggle to model the dynamic nature of user interests in real-world applications.

In a similar vein, Wang et al. proposed CSRM [12], an end-to-end neural network that first encodes each session using NARM at the item level. CSRM then enriches the current session's representation by considering neighboring sessions, which helps capture session-level contextual information. The Fusion gating mechanism used in CSRM allows for the combination of multiple feature sources to learn a robust session representation. Although CSRM demonstrates improved session modeling by incorporating neighboring sessions, it still faces the challenge of effectively capturing highly diverse user interests within a session.

The introduction of global information into session recommendation was further explored in MGIR [16], which incorporates global incompatibilities and combines positive and negative relationships to refine the session representation. However, the reliance on global information may introduce noise and dilute the session-specific focus. Similarly, SPARE [17] introduced a multi-hop information aggregation process and employed shortcut connections to improve efficiency. While SPARE's multi-layer aggregation is effective, its performance may degrade in highly sparse graphs or when modeling user behaviors that change rapidly across sessions. GCE-GNN [6] and COTREC [18] introduced more sophisticated approaches by incorporating both global and session-level graphs to learn item embeddings at different levels of granularity. GCE-GNN integrates a soft attention mechanism to modify learned item embeddings, enhancing the global representation. COTREC, on the other hand, employs contrastive learning to refine session representations by strengthening internal and external connections. While both models improve the overall representation learning process, they face limitations in handling the dynamic nature of user interests. In particular, relying solely on individual items to form the final session representation may fail to fully capture the diverse and evolving nature of user behavior.

Overall, the existing GNN-based methods have significantly advanced the field of session-based recommendation, with each approach offering unique strengths in capturing item dependencies, long-range interactions, and user behaviors. However, challenges remain in effectively modeling the dynamic and diverse interests of users across sessions. Our proposed method, HMGS, seeks to address these limitations by introducing a hierarchical graph structure that dynamically adapts to shifting user preferences, providing a more flexible and accurate recommendation system.

3 Technique

The session-based recommendation problem is formally defined in this section, followed by a delineation of the construction of global and multi-granular session graphs employed for learning latent item embeddings.

3.1 Problem Definition

Let $V = \{v_1, v_2, v_3, \ldots, v_M\}$ denote the universal item corpus, where M represents the cardinality of the item set. A session, which we'll refer to as S, is written as $S = \{v_1^s, v_2^s, \ldots, v_l^s\}$, and this reflects the sequential order of user interactions that occur during the current session. Here, v_i^s represents the identifier of the item that was interacted with at the i-th stage within session S, and l represents the length of the session. For a given session S, the goal of the session-based recommendation task is to predict the top N (where $1 \leq N \leq M$) items that are most likely to be interacted with next, by leveraging the information from the items in the current session S. The task necessitates the analysis of relational dependencies and sequential dynamics within segmented user behavior to generate precise probabilistic forecasts and actionable user recommendations.

3.2 Graph Models: Global Graph and Multi-Granularity Session Graph
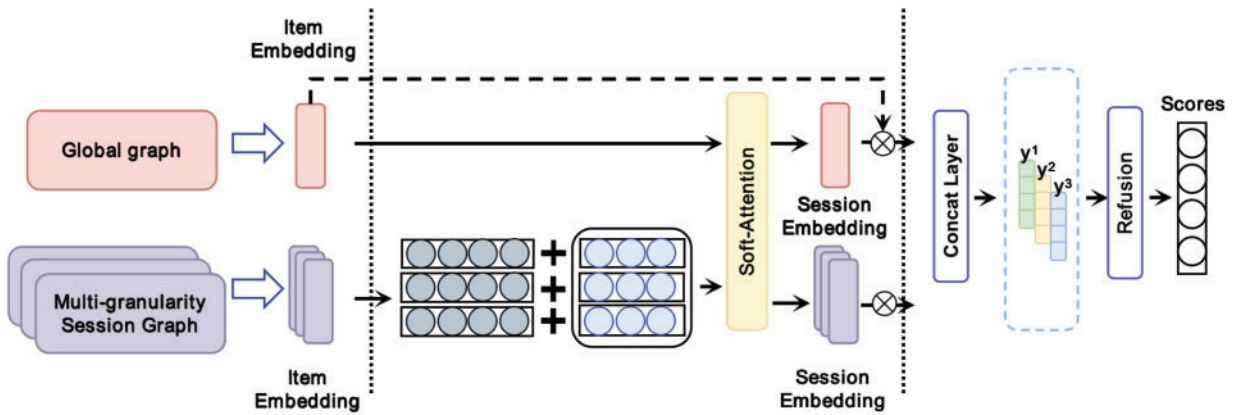
3.2.1 Global Graph Construction

For the derivation of global-level item embeddings, dyadic item transitions observed across all sessions are incorporated to integrate cross-session behavioral patterns. The global interaction graph is constructed by representing each item as a node within the aggregated interaction corpus, enabling the extraction of cross-user co-occurrence patterns to yield generalizable item embeddings. Building upon GCE-GNN's theoretical foundation, a global transition modeling approach is proposed. Specifically, the graph topology is derived from first-order adjacency relations within the complete interaction history. Formally, the global graph is defined as: $G_g = (V_g, E_g)$ represents the global graph, where V_g denotes all the items encompassed in the global graph, and $E_g = \{e_{ij}^g | (v_i, v_j | v_i \in V, v_j \in N_E(v_i)\}$ stands for the collection of edges in the global graph. Hence, the global graph is an undirected graph.
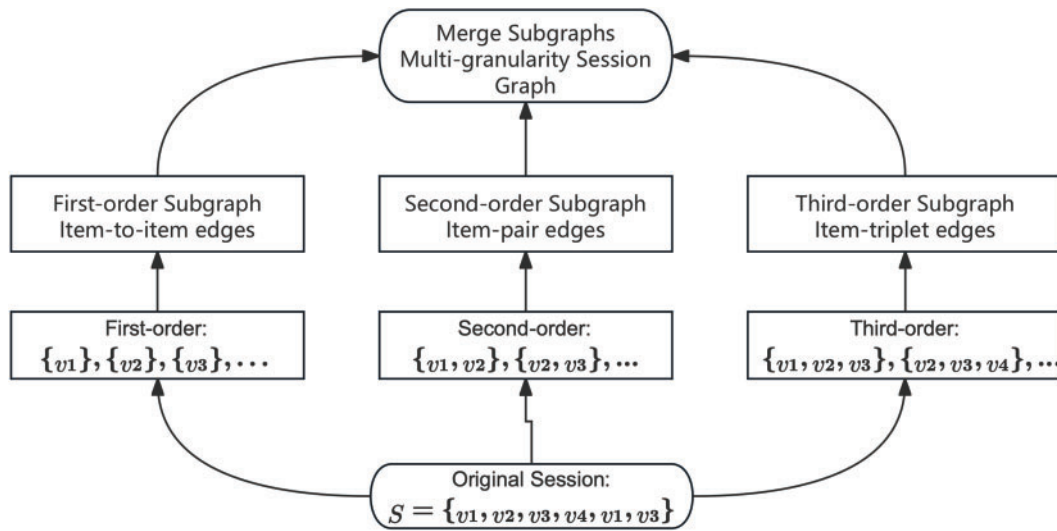
3.2.2 Multi-Granularity Session Graph

In current session-based recommendation methods, each item is individually considered, modeling a single item as a node. This approach overlooks the information embedded in consecutive session segments and is insufficient for accurately representing users' multi-granularity interests. In this work, we address this limitation by partitioning sessions into segments of length k and modeling user interests at different granularity levels. We define v_j^k as a continuous interest unit where items come from a continuous segment, starting from the j-th item in the session and taking a segment of length k, i.e., $v_j^k = (v_j, \ldots, v_{j+k+1})$. The length k also represents the granularity level of the interest unit. As illustrated in Figs. 1 and 2, consider a session $S = \{v_1, v_2, v_3, v_4, v_1, v_3\}$. The first-level interest unit corresponds to the original sequence of S. Session segments $(v_1, v_2), (v_2, v_3), \ldots, (v_1, v_3)$ serve as 2-level interest unit, denoted as $v_1^2, \ldots, v_5^2$. This segmentation method refers to pairing consecutive items within the session to form a lower-order interest unit. Additionally, session segments $(v_1, v_2, v_3), \ldots, (v_4, v_1, v_3)$ represent another 2-level interest unit, denoted as $v_1^3, \ldots, v_4^3$, where three consecutive items are grouped to form a higher-order interest unit. For the first-order interest unit, we assign a learnable embedding vector v_j^1 to represent it. For higher-order k-level interest unit, we use GRU (Gated Recurrent Unit) to aggregate each item's representation within the interest unit, generating its initial representation, which is v_j^k.

Initially, k subgraphs with heterogeneous granularity tiers are constructed. Consider the k-th tier session graph: this hierarchically structured session graph is formulated to encapsulate spatial adjacency in user-item interactions, instantiated as a directed weighted graph $G_s^k = (V_s^k, E_s^k)$. Where vertices correspond to k-th tier interest units and edges connect adjacent units within the session sequence. The first-tier graph captures atomic intent transitions between items, modeling fine-grained intra-session interaction dynamics. Progressively higher-tier graphs aggregate broader behavioral motifs, identifying transitions between abstracted, macro-scale interest units. Such hierarchical architecture facilitates the simultaneous modeling of micro-level interactional immediacy and macro-level longitudinal preference evolution, ensuring multi-scale behavioral pattern comprehension.

Subsequently, a multi-granular session graph is synthesized by interconnecting subgraphs across granularity tiers. Inter-tier edges are introduced to link nodes from distinct granularity levels, differing fundamentally from intra-tier edges. These cross-tier edges establish bidirectional connectivity between an interest unit and its antecedent/consequent nodes at adjacent granularity tiers, enabling hierarchical preference propagation.

Figure 1: The overall framework of HMGS. Given a specific session and the entire item pool, the HMGS first constructs both a global graph and a multi-granularity session graph. These two graphs are then integrated through multi-head graph attention networks, each with diverse structures. Next, a soft attention mechanism is used to combine the unit representations within the session at various levels corresponding to different user interests. Finally, the click-through rate (CTR) is predicted through a fine-grained fusion process

Figure 2: Construction of a multi-granularity session graph for enhanced session-based recommendation

3.3 Global-Level Item Representation Learning Layer

The transitions between items in different sessions is crucial for learning global representations of projects. Next, we introduce how to aggregate node-related neighborhood information in the global graph. As the importance of different neighbors of a node varies, we use a Graph Attention Network (GAT) to learn the representation of each project.

$$h_{v_i}^{(l+1)} = \sum_{v_j \in N_{v_i}^g} \alpha_g^l h_{v_j}^l \tag{1}$$

In this context, α_g^l is utilized to assess the significance weight of diverse neighbor nodes. Specifically, the closer the representation of a certain project is to v_i, the greater its corresponding weight will be. For v_i

that belongs to S and its neighboring nodes v_j within $N^g_{v_i}$, the importance of project v_j is contingent upon the adjacency relationship between v_i and v_j.

$$\tilde{\alpha}^l_g = q_1^{{(l)}^T} \text{LeakyRelu}\left(W_1^l \left[h^l_{v_i}; h^l_{v_j}\right]\right) \tag{2}$$

Here, we use LeakyRelu as the activation function, [;] denotes the concatenation operation, $q_1^{(l)} \in \mathbb{R}^{2d \times 1}$, and $W_1^l \in \mathbb{R}^{d \times d}$ are trainable parameters. h^0_v is set to h_v during the initial propagation. Then, the softmax function is applied to normalize all coefficients related to the neighborhood of v_i:

3.4 Session-Level Item Representation Learning Layer

Multi-level session graphs are established through dyadic item transitions observed in localized session sequences. Subsequent to this structural formalization, the derivation of session-level item embeddings is methodologically elaborated. A Multi-Head Heterogeneous Graph Attention Network (MHGAT) is employed to hierarchically derive interest unit representations, operating across distinct granularity tiers within the session graph architecture. Suppose there is a directed edge (s, e, t), where s and t represent the source and target interest units, respectively, and e represents the edge. It should be emphasized that s and t can be at any granularity level. More specifically, we denote s^k and t^k as the specific granularity levels of s and t, respectively, and e represents the edge type. For any k-MHGAT layer, the values of s^k and t^k are within the range of $\{1, \ldots, K\}$, and ϕ_e belongs to the set $\{inter_1, \ldots, inter_k, intra_1, \ldots, intra_k\}$.

For each layer, a bidirectional attention mechanism is employed to aggregate the representations of neighboring units. Given a neighbor set N_{ϕ_e}, the neighbor aggregation mechanism is as follows:

$$x_t^{(l+1)} = \sum_{\phi_e} \sum_{s \in N_{\phi_e}} \alpha_s^l W_{\phi_e}^l x_s^l \tag{3}$$

Here, $W_{\phi_e}^l \in \mathbb{R}^{d \times d}$ is a non-shared learnable weight for different layers and edge types, and α_s^l is the importance weight of different neighbor nodes.

$$\tilde{\alpha}_s^l = q_2^{{(l)}^T} \text{LeakyRelu}\left(W_{\phi_e}^l \left[x_s^l; x_t^l\right]\right) \tag{4}$$

Here, we use LeakyRelu as the activation function, [;] denotes the concatenation operation, and $q_2^{(l)} \in \mathbb{R}^{2d \times 1}$. Then, the softmax function is applied for normalization:

$$\alpha_s^l = \text{Softmax}(\tilde{\alpha}_s^l) \tag{5}$$

Similar to the global graph, MHGAT also adopts a multi-head attention structure, and the node representation is output through a readout function after obtaining the representation for each head.

$$x_t^{(l+1)} = \underset{i=1,\ldots,H}{\text{R}}(x_t^{(l+1),i}) \tag{6}$$

3.5 Session Representation Learning Layer

Within the proposed framework, session embeddings are constructed through the integration of heterogeneous granularity tiers of user interest units, derived from a hierarchical session graph architecture. To optimally leverage multi-scale behavioral patterns, discrete session embeddings are generated for each granularity tier. An attention-based aggregation mechanism is implemented to synthesize intra-session

interest units, producing the composite representation S_g^k. Concurrently, the terminal interest unit at each granularity level is interpreted as the session's localized behavioral signature, formalized as S_l^k.

To capture the complete user intent from each level, we represent the composition of interest units from different granularities in the current session using a context set, i.e., $C = \{x_i^k \mid i = 1, \ldots, n_k, k = 1, \ldots, K\}$. We obtain S_g^k through the following formula:

$$S_g^k = \sum_{c=1}^{|C|} \text{Softmax}(\beta_c^k) x_c \tag{7}$$

Here, the priority β_c^k is determined by the corresponding user interest unit and context. Referring to GCE-GCN, we combine reverse positional information and session information. However, GCE-GCN only uses single granularity positional information. To more accurately model multi-granularity user interests, we present a collection of learnable position embedding matrices denoted as $P = [P_1, \ldots, P_k]$. Here, $P_k = [p_1^k, \ldots, p_l^k]$, where p_i (belonging to $\mathbb{R}^d$) represents the position vector corresponding to a particular position i. Meanwhile, l indicates the number of interest units at level-k within the current session. The position details are incorporated by means of concatenation and a subsequent non-linear transformation, as follows:

$$x_i^k = \tanh\left(W_2^k[x_i^k \| p_{(l-i+1)}^k] + b_1^k\right) \tag{8}$$

Here, $W_2^k \in \mathbb{R}^{d \times 2d}$ is a trainable parameter, and $b_1^k \in \mathbb{R}^d$ is a bias term.

Subsequently, a soft attention mechanism is employed by us to acquire the weights relevant to the corresponding items.

$$\hat{x}^k = \frac{1}{l} \sum_{i=1}^{l} x_i^k \tag{9}$$

$$\beta_i^k = q_3^T \sigma\left(W_3^k x_i^k + W_4^k \hat{x}^k + b_2^k\right) \tag{10}$$

In this regard, W_3^k, W_4^k which are elements of $\mathbb{R}^{d \times d}$, q_3^T belonging to $\mathbb{R}^d$ are parameters that can be trained. Additionally, b_2^k in $\mathbb{R}^d$ serves as a bias term. The variable $\hat{x}^k$ represents the average of the interest units within the current session at the k-th level. The symbol σ denotes the sigmoid function.

Ultimately, we integrate the local and global representations of the session to produce the session representation for each level of the interest units:

$$S^k = W^k[S_g^k; S_l^k] \tag{11}$$

Here, [;] represents the concatenation operation, and W^k is the projection matrix.

3.6 Hierarchical Matching Prediction Layer

After generating embeddings for different levels of sessions and global item embeddings, a multi-level matching mechanism is proposed to capture comprehensive user preferences. Specifically, recommendations are first made based on the intent of each level, and the results are then fused to provide the final recommendation.

For each level of session embedding, candidate embeddings obtained from the global graph are multiplied with different level session embeddings to get interest scores for the candidate set at different

levels. In particular, for the candidate item set $V = \{v_1, v_2, v_3, \ldots, v_m\}$, we obtain user interest scores for the candidate set items at different levels:

$$y_i^k = \langle S^k, h_i \rangle \tag{12}$$

Here, $\langle, \rangle$ represents the dot product operation, and y_i^k is the interest score for the k-th level item v_i in the session S. Subsequently, a gating mechanism is utilized to fuse interest scores at different levels, yielding the final score for each item:

$$G_i = \text{sigmoid}(W_1^y y_i^1, \ldots, W_k^y y_i^k + b_3) \tag{13}$$

$$\hat{y}_i = G_i(y_i^1, \ldots, y_i^k) \tag{14}$$

Here, $W_1^y, W_k^y \in \mathbb{R}^{k \times d}$ are trainable parameters, and $b_3 \in \mathbb{R}^k$ is the bias matrix. In order to dynamically adjust the layer weights according to varying session types and user behaviors, we introduce a learning-based approach to adaptively tune the weights during training. Specifically, the weight matrices W_1^y, W_k^y are not fixed; instead, they are adjusted by a learnable function that depends on the session's characteristics and the user's interaction history. This enables the model to prioritize different levels of user interests based on the context, ensuring that the gating mechanism is responsive to the dynamic shifts in user behavior over time. This dynamic adjustment process ensures that the model remains flexible and adaptive to varying user interaction patterns.

We use the cross-entropy loss function as the optimization objective for parameter learning, and the loss equation is:

$$L(\hat{y}) = -\sum_{i=1}^{|I|} y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i) \tag{15}$$

Here, y represents the one-hot encoding vector of the ground truth item. Specifically, for the i-th item, if it is the target item of the given session, then $y_i = 1$; conversely, if it is not the target item, $y_i = 0$.

4 Experiments

In this section, we first delineate the datasets, baselines, and evaluation metrics. Subsequently, we conduct comprehensive analyses of the experimental results.

4.1 Datasets

For gauging the effectiveness of HMGS, we perform experiments on two datasets prevalently utilized in session-based recommendation studies: $Tmall$[1] and $Diginetica$[2]. The $Tmall$ dataset, sourced from the IJCAI-15 competition, holds anonymous user shopping data from the Tmall online marketplace. It offers a practical view of users' purchasing behaviors. The $Diginetica$ dataset, provided for the CIKM Cup 2016's personalized e-commerce search challenge, contains transition records apt for session-based recommendation analysis. It presents valuable details about users' search and selection patterns.

In line with the approaches taken in prior research works [6,11,15,19], we implement a filtering process to remove short sessions that have a length of less than 2, as well as items that occur with a frequency lower than 5. Subsequently, we employ the data augmentation strategies detailed in [2,6,15] to preprocess the dataset. To

---

[1] https://tianchi.aliyun.com/dataset/42 (accessed on 7 March 2025)
[2] https://competitions.codalab.org/competitions/11161 (accessed on 7 March 2025)

further enhance the data, we utilize a sequence-splitting method to augment and label both the training and testing datasets. Moreover, for a session $S = [s_1, s_2, \ldots, s_n]$, we generate sequences and corresponding labels by a sequence splitting preprocessing, i.e., $([s_1], s_2), ([s_1, s_2], s_3), \ldots, ([s_1, s_2, \ldots, s_{n-1}], s_n)$ for both training and testing across all the two datasets. The comprehensive statistics of these datasets are illustrated in Table 1.

Table 1: Statistical information regarding the datasets such as Tmall and Diginetica

Statistics	Tmall	Diginetica
#clicks	818,479	981,620
#train	351,268	716,835
#test	25,898	60,194
#items	40,728	42,596
Average length	6.69	5.12

4.2 Evaluation Metrics

Adhering to the approaches described in [6,11,15], the assessment of the top-N recommendation performance is carried out by means of two prevalently utilized ranking-oriented metrics, namely Precision (P@N) and Mean Reciprocal Rank (MRR@N). In this context, the variable N designates the count of recommended items. Precision (P@N) serves as a measure of the ratio of accurately recommended items within the top-N selection. Meanwhile, MRR@N is calculated as the mean of the reciprocal ranks of the correctly recommended items, taking into account the sequence of the recommendation ranking. A greater MRR value is indicative of the correct recommendations being positioned nearer to the beginning of the ranking list, thereby denoting a more proficient performance with respect to the ranking order.

4.3 Baselines

We compare HMGS with the following representative methods:

GRU4REC[3] [7] employs Gated Recurrent Units (GRU) to model the sequences of user interactions for recommendation tasks.

NARM[4] [8], a state-of-the-art RNN model, incorporates an attention mechanism to prioritize the main intention of users and combines sequential behaviors to generate effective recommendations.

SRGNN[5] [15] method, a gated graph convolutional layer is applied to extract item embeddings, and a soft-attention mechanism is utilized to compute session embeddings. This combination enables a more comprehensive understanding of the relationships between items and sessions, thus enhancing the quality of the generated recommendations.

GCE-GNN[6] [6] develops two types of session-based graphs to capture both local and global relationships at different levels, improving the recommendation process.

TAGNN[7] [20] proposes a Target-Aware Graph Neural Network (GNN) that learns the dynamic interest representations of users to adapt to different target items.

---

[3] https://github.com/hidasib/GRU4Rec (accessed on 7 March 2025)
[4] https://github.com/lijingsdu/sessionRec_NARM (accessed on 7 March 2025)
[5] https://github.com/CRIPAC-DIG/SR-GNN (accessed on 7 March 2025)
[6] https://github.com/CCIIPLab/GCE-GNN (accessed on 7 March 2025)
[7] https://github.com/CRIPAC-DIG/TAGNN (accessed on 7 March 2025)

S^2-**DHCN**[8] [11] uses two distinct types of hypergraphs to model the relationships within and between sessions and integrates self-supervised learning to strengthen session-based recommendations.

COTREC[9] [18] decomposes session data into two complementary views. These views are used to model both the internal and external connectivity of sessions and then are utilized to enhance each other's learning, leading to more comprehensive and accurate recommendations.

CORE[10] [21] introduces the CORE framework, which is designed to unify the representation space for both the encoding and decoding processes in session-based recommendation models.

MGIR[11] [16] proposes a multi-dimensional model. This model encodes diverse item relationships through different aggregation layers. Combining both positive and negative relations, generates enhanced session representations, resulting in more accurate and relevant recommendations.

SPARE[12] [17] suggests an approach that explicitly models multi-hop information aggregation over several layers using shortest-path edges, leveraging knowledge from the sequential recommendation domain.

CARE [22] enhances session-based recommendation by introducing a context-aware attention mechanism that captures dynamic and evolving user interest distributions within sessions.

The proposed HMGS model offers several advantages over the baseline methods. Unlike traditional RNN-based models such as GRU4REC and NARM, which rely solely on sequential data and attention mechanisms to capture user intent, HMGS introduces a hierarchical session graph matching framework that effectively models multi-granularity user interests. This allows it to better capture the intricate patterns of user behavior at different levels, enhancing recommendation accuracy. Moreover, while methods like SRGNN and GCE-GNN utilize graph-based techniques to model item relationships, HMGS goes a step further by incorporating both a global and local layer to capture both global item transformations and fine-grained session-specific patterns, leading to a more comprehensive understanding of user preferences. Additionally, HMGS addresses the limitations of models like TAGNN and S^2-DHCN, which focus on specific user interests or self-supervised learning, by incorporating a global graph layer to account for broader behavioral correlations across sessions. Overall, HMGS demonstrates a superior ability to balance both local and global contexts, making it more effective than existing methods in capturing complex user behaviors and providing higher-quality recommendations.

4.4 Implementation Details

To determine the optimal hyperparameters, a grid search strategy is employed for all methods. We randomly allocate 10% of the training data as the validation set, and the best hyperparameter combination is selected based on its performance, with the last item of each user's interaction history serving as the prediction target. For baseline models built on Graph Neural Networks, we investigate the performance of graph layers within the range of $\{1, 2, 3, 4, 5\}$ to pick the optimal value. In the case of our proposed HMGS model, we further search for the best granularity level K within the same range of $\{1, 2, 3, 4, 5\}$. The model optimization utilizes the Adam algorithm, where the initial learning rate is set to $1e^{-3}$ and the weight decay is $5e^{-4}$. The maximum session length is set at 30. In line with related research, a learning rate decay strategy is adopted, reducing the rate by a factor of 0.1 every 3 epochs. The embedding dimension and batch size are fixed at 256 and 512, respectively. Finally, for evaluating model performance, the Precision and MRR metrics are used for comparison.

---

[8] https://github.com/xiaxin1998/DHCN (accessed on 7 March 2025)
[9] https://github.com/xiaxin1998/COTREC (accessed on 7 March 2025)
[10] https://github.com/RUCAIBox/CORE (accessed on 7 March 2025)
[11] https://github.com/zc-97/MGIR (accessed on 7 March 2025)
[12] https://github.com/dbis-uibk/SPARE (accessed on 7 March 2025)

5 Experimental Results

5.1 Overall Performance

In this segment, we conduct a comparison between HMGS and the contemporary state-of-the-art (SOTA) baselines to confirm its effectiveness. We emphasize the top results from both the baselines and our own HMGS. The symbol * is used to denote that HMGS outperforms the current SOTA metric. In Table 2, the "Enhancement" represents the percentage improvement that HMGS attains relative to the best result among the baseline models. All of these improvements are statistically significant, with a p-value less than 0.01. The statistical significance tests were carried out in pairs.

Table 2: Performance comparison of HMGS against the baselines. * denotes a significant improvement of HMGS over the best baseline using a paired t-test ($p < 0.01$), and bold font represents the second-best result

Method	Tmall				Diginetica			
	P@10	M@10	P@20	M@20	P@10	M@10	P@20	M@20
GRU4REC	9.47	5.78	10.93	5.89	17.93	7.33	29.45	8.33
NARM	19.17	10.42	23.30	10.70	35.44	15.13	49.70	16.17
SRGNN	23.41	13.45	27.57	13.72	36.86	15.52	50.73	17.59
GCE-GNN	28.01	15.08	33.42	15.42	41.54	18.29	54.64	19.20
TAGNN	**34.80**	17.58	**42.01**	18.08	38.58	17.15	51.31	18.03
S^2-DHCN	26.22	14.60	31.42	15.05	41.16	18.15	53.18	18.44
COTREC	30.62	17.65	36.35	18.04	**41.88**	18.16	54.18	19.07
CORE	33.16	19.10	39.16	19.52	39.90	17.58	52.84	18.47
MGIR	30.71	17.03	36.31	17.42	40.63	17.86	53.73	18.77
SPARE	33.61	**19.78**	39.28	**20.07**	41.36	18.12	**55.60**	19.01
CARE	34.33	19.48	39.72	19.86	41.53	**18.39**	55.05	**19.32**
HMGS	**41.95***	**20.57***	**49.17***	**21.36***	**47.17***	**19.16***	**57.56***	**20.07***
Improvement	20.54%	4.00%	16.89%	6.43%	12.63%	4.19%	3.52%	3.88%
p-value	$1e^{-4}$	$3e^{-4}$	$1e^{-4}$	$3e^{-4}$	$1e^{-4}$	$3e^{-4}$	$1e^{-4}$	$1e^{-3}$

The experimental results of all methods are presented in Table 2. From this table, three key observations can be made:

First, graph-based models such as GCE-GNN and SR-GNN generally perform better than RNN-based methods like NARM and STAMP. This highlights the efficacy of Graph Neural Networks (GNNs) in capturing patterns within session data. Among these, GCE-GNN shows superior performance compared to SR-GNN, suggesting that the integration of local and global information is crucial for accurately inferring user intentions in session-based recommendations. Additionally, models that incorporate unsupervised tasks, such as S^2-DHCN and COTREC, further emphasize the importance of cross-session information. They demonstrate how additional unlabeled data can enhance the learning process and improve prediction accuracy.

Second, HMGS exhibits excellent performance across two datasets, particularly on the Tmall dataset. This showcases its ability in multi-level user interest modeling. Moreover, the MGIR model outperforms all baseline models on various metrics, highlighting the importance of matching users and items in a consistent representation space.

Third, our HMGS method significantly outperforms all baselines. This indicates that session-based recommendations can benefit from our proposed framework that combines global graphs with multi-granularity user interest graphs. We attribute these significant improvements to the following factors: (1) The proposed hierarchical user interest graph can explore user intentions at multiple granularity levels and model the intricate transitions between different user intentions; (2) The proposed global graph can capture common preferences among different users, enabling a more accurate representation of candidate items; (3) The proposed hierarchical matching framework can effectively utilize the advantages of both the global graph and the multi-level user interest graph, thus promoting efficient and precise matching between users and the candidate set.

5.2 Ablation Study

To enhance the comprehensiveness and demonstrate the robustness and effectiveness of the HMGS model, we conducted a large number of ablation experiments to explore the impacts of the values of k and the different numbers of layers in the global graph on the model's performance, as shown in Tables 3 and 4. First of all, we carried out ablation studies on the selection of k values. The experimental results show that the model achieves the best performance when k = 3, indicating that under the condition of generally short session lengths, items with closer relationships are more similar. In addition, we conducted ablation experiments on different layers in the global graph and found that performing aggregation twice on the premise of first-order neighbors yields the best results.

Table 3: Experiments on the selection of k values, and the bold font represents the optimal result

	Tmall				Diginetica			
	P@10	P@20	M@10	M@20	P@10	P@20	M@10	M@20
k = 1	40.12	47.24	19.77	19.25	45.16	55.49	18.71	19.22
k = 2	40.82	48.22	20.01	20.27	46.75	57.20	19.12	19.87
k = 3	**41.95**	**49.17**	**20.57**	**21.36**	**47.17**	**57.56**	**19.16**	**20.07**
k = 4	41.28	48.71	20.08	21.12	47.01	57.66	18.77	19.10
k = 5	40.92	46.51	19.78	19.28	46.88	56.10	18.91	19.44

Table 4: Experiments on different layers in the global graph, and the bold font represents the optimal result

	Tmall				Diginetica			
	P@10	P@20	M@10	M@20	P@10	P@20	M@10	M@20
N = 1	40.66	47.18	19.86	19.38	46.11	55.17	18.29	19.33
N = 2	**41.95**	**49.17**	**20.57**	**21.36**	**47.17**	**57.56**	**19.16**	**20.07**
N = 3	40.92	49.01	19.78	20.77	47.01	57.22	19.03	20.01
N = 4	41.20	48.43	19.07	20.95	47.56	57.12	18.80	19.03
N = 5	40.11	47.56	19.45	19.18	46.83	56.91	18.99	19.41

6 Conclusion

In session-based recommendation systems, aligning user interests with candidate items poses a central challenge. Recent research efforts have been centered around leveraging Graph Neural Networks (GNNs)

to learn distinct representations for sessions and items. However, it is our contention that depending on fixed neighborhood structures restricts the model's capacity for expression, impeding its ability to comprehensively capture the diverse interest preferences of users. This limitation not only undermines the model's performance but also neglects the interaction cues between items across different sessions. To tackle these problems, we present a multi-level matching GNN designed for session-based recommendations. Our methodology adaptively accumulates item neighborhood information at multiple tiers, learning both global item representations and multi-level user interest unit representations. We introduce a multi-level matching framework that bolsters the correspondence between users and candidate items by seizing user interests at different levels of granularity. Comprehensive experiments conducted on two benchmark datasets reveal that our approach, HMGS, considerably surpasses the existing state-of-the-art models.

Looking to the future, we intend to explore the generalization of the HMGS framework from both theoretical and practical standpoints. Specifically, we plan to investigate how the adaptive and hierarchical capabilities of HMGS can be tailored for domains beyond session-based retail recommendations, such as music or video streaming services. These domains, which also feature highly dynamic user preferences and complex interaction patterns, could benefit substantially from HMGS's ability to capture nuanced user interests at various granularity levels. Moreover, we aim to examine the integration of item attribute information to more effectively model user-item interactions and ensure consistent matching across sessions. This will include studying the impact of incorporating genre, artist, or user engagement metrics in music recommen, or director, actor, and viewer ratings in video recommendations, to see how these features can enhance the predictive accuracy of HMGS in these contexts.

Author Contributions: Xiaodong Li led data collection, paper drafting, and figure preparation. Xiao Zhang assisted in data work, co-wrote key sections, and contributed to figures. Zhonghong Ou comprehensively revised the manuscript. Pengfei Zhang oversaw the project. Rui Xin verified theoretical consistency, Xing Xu validated practical implications, Yuzhen Wang checked figures and captions, and Meina Song proofread for language accuracy. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The data that support the findings of this study are openly available in Tianchi at https://tianchi.aliyun.com/dataset/42 (accessed on 7 March 2025) and in the Competition at https://competitions.codalab.org/competitions/11161 (accessed on 7 March 2025).

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. Huang P, He X, Gao J, Deng L, Acero A, Heck LP. Learning deep structured semantic models for web search using clickthrough data. In: He Q, Iyengar A, Nejdl W, Pei J, Rastogi R, editors. 22nd ACM International Conference on Information and Knowledge Management, CIKM'13; 2013 Oct 27–Nov 1; San Francisco, CA, USA: ACM; 2013. p. 2333–8. doi:10.1145/2505515.2505665.
2. Tan YK, Xu X, Liu Y. Improved recurrent neural networks for session-based recommendations. In: Karatzoglou A, Hidasi B, Tikk D, Shalom OS, Roitman H, Shapira B et al., editors. Proceedings of the 1st Workshop on Deep

Learning for Recommender Systems, DLRS@RecSys 2016; 2016 Sep 15; Boston, MA, USA: ACM; 2016. p. 17–22. doi:10.1145/2988450.2988452.

3. Song J, Shen H, Ou Z, Zhang J, Xiao T, Liang S. ISLF: interest shift and latent factors combination model for session-based recommendation. In: Kraus S, editor. Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019; 2019 Aug 10–16; Macao, China; 2019. p. 5765–71. doi:10.24963/ijcai.2019/799.

4. Chen T, Wong RC. Handling information loss of graph neural networks for session-based recommendation. In: Gupta R, Liu Y, Tang J, Prakash BA, editors. KDD '20: The 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining; 2020 Aug 23–27; Virtual Event, CA, USA: ACM; 2020. p. 1172–80. doi:10.1145/3394486.3403170.

5. Yuan F, Karatzoglou A, Arapakis I, Jose JM, He X. A simple convolutional generative network for next item recommendation. In: Culpepper JS, Moffat A, Bennett PN, Lerman K, editors. Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining, WSDM 2019; 2019 Feb 11–15; Melbourne, VIC, Australia: ACM; 2019. p. 582–90. doi:10.1145/3289600.3290975.

6. Wang Z, Wei W, Cong G, Li X, Mao X, Qiu M. Global context enhanced graph neural networks for session-based recommendation. In: Huang JX, Chang Y, Cheng X, Kamps J, Murdock V, Wen J et al., editors. Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2020; 2020 Jul 25–30; China: ACM; 2020. p. 169–78. doi:10.1145/3397271.3401142.

7. Hidasi B, Karatzoglou A, Baltrunas L, Tikk D. Session-based recommendations with recurrent neural networks. In: Bengio Y, LeCun Y, editors. 4th International Conference on Learning Representations, ICLR 2016; 2016 May 2–4; San Juan, Puerto Rico: Conference Track Proceedings; 2016.

8. Li J, Ren P, Chen Z, Ren Z, Lian T, Ma J. Neural attentive session-based recommendation. In: Lim E, Winslett M, Sanderson M, Fu AW, Sun J, Culpepper JS et al., editors. Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, CIKM 2017; 2017 Nov 06–10; Singapore: ACM. p. 1419–28. doi:10.1145/3132847.3132926.

9. Huang C, Chen J, Xia L, Xu Y, Dai P, Chen Y, et al. Graph-enhanced multi-task learning of multi-level transition dynamics for session-based recommendation. In: Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021; 2021 Feb 2–9; Virtual Event: AAAI Press; 2021. p. 4123–30. doi:10.1609/aaai.v35i5.16534.

10. Xu C, Zhao P, Liu Y, Sheng VS, Xu J, Zhuang F, et al. Graph contextualized self-attention network for session-based recommendation. In: Kraus S, editors. Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019; 2019 Aug 10–16; Macao, China; 2019. p. 3940–6. doi:10.24963/ijcai.2019/547.

11. Xia X, Yin H, Yu J, Wang Q, Cui L, Zhang X. Self-supervised hypergraph convolutional networks for session-based recommendation. In: Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021; 2021 Feb 2–9; Virtual Event: AAAI Press; 2021. p. 4503–11. doi:10.1609/aaai.v35i5.16578.

12. Wang M, Ren P, Mei L, Chen Z, Ma J, de Rijke M. A collaborative session-based recommendation approach with parallel memory modules. In: Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR'19; 2019; New York, NY, USA: Association for Computing Machinery; p. 345–54. doi:10.1145/3331184.3331210.

13. Liu Q, Zeng Y, Mokhosi R, Zhang H. STAMP: Short-term attention/memory priority model for session-based recommendation. In: Guo Y, Farooq F, editors. Proceedings of the 24th ACM SIGKDD International Conference on Kowledge Discovery & Data Mining, KDD 2018; 2018 Aug 19–23; London, UK: ACM; 2018. p. 1831–9. doi:10.1145/3219819.3219950.

14. Kang W, McAuley JJ. Self-attentive sequential recommendation. In: IEEE International Conference on Data Mining, ICDM 2018; 2018 Nov 17–20; Singapore: IEEE Computer Society; 2018. p. 197–206. doi:10.1109/ICDM.2018.00035.

15. Wu S, Tang Y, Zhu Y, Wang L, Xie X, Tan T. Session-based recommendation with graph neural networks. In: The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019; 2019 Jan 27–Feb 1; Honolulu, HI, USA: AAAI Press; 2019. p. 346–53. doi:10.1609/aaai.v33i01.3301346.

16. Han Q, Zhang C, Chen R, Lai R, Song H, Li L. Multi-faceted global item relation learning for session-based recommendation. In: Amigó E, Castells P, Gonzalo J, Carterette B, Culpepper JS, Kazai G, editors. SIGIR '22: The 45th International ACM SIGIR Conference on Research and Development in Information Retrieval; 2022 Jul 11–15; Madrid, Spain: ACM; 2022. p. 1705–15. doi:10.1145/3477495.3532024.

17. Peintner A, Mohammadi AR, Zangerle E. SPARE: shortest path global item relations for efficient session-based recommendation. In: Zhang J, Chen L, Berkovsky S, Zhang M, Noia TD, Basilico J et al., editors. Proceedings of the 17th ACM Conference on Recommender Systems, RecSys 2023; 2023 Sep 18–22; Singapore: ACM; 2023. p. 58–69. doi:10.1145/3604915.3608768.

18. Xia X, Yin H, Yu J, Shao Y, Cui L. Self-supervised graph co-training for session-based recommendation. In: Demartini G, Zuccon G, Culpepper JS, Huang Z, Tong H, editors. CIKM '21: The 30th ACM International Conference on Information and Knowledge Management, Virtual Event; 2021 Nov 1–5; Queensland, Australia: ACM; 2021. p. 2180–90. doi:10.1145/3459637.3482388.

19. Qiu R, Li J, Huang Z, Yin H. Rethinking the item order in session-based recommendation with graph neural networks. In: Zhu W, Tao D, Cheng X, Cui P, Rundensteiner EA, Carmel D, editors. Proceedings of the 28th ACM International Conference on Information and Knowledge Management, CIKM 2019; 2019 Nov 3–7; Beijing, China: ACM; 2019. p. 579–88. doi:10.1145/3357384.3358010.

20. Yu F, Zhu Y, Liu Q, Wu S, Wang L, Tan T. TAGNN: target attentive graph neural networks for session-based recommendation. In: Huang JX, Chang Y, Cheng X, Kamps J, Murdock V, Wen J, editors. Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2020; 2020 Jul 25–30; Virtual Event, China: ACM; 2020. p. 1921–4. doi:10.1145/3397271.3401319.559.

21. Hou Y, Hu B, Zhang Z, Zhao WX. CORE: simple and effective session-based recommendation within consistent representation space. In: Amigó E, Castells P, Gonzalo J, Carterette B, Culpepper JS, Kazai G, editors. SIGIR '22: The 45th International ACM SIGIR Conference on Research and Development in Information Retrieval; 2022 Jul 11–15; Madrid, Spain: ACM; 2022. p. 1796–801. doi:10.1145/3477495.3531955.

22. Tong P, Zhang Z, Liu Q, Wang Y, Wang R. CARE: context-aware attention interest redistribution for session-based recommendation. Expert Syst Appl. 2024;255(2):124714. doi:10.1016/j.eswa.2024.124714.