Multi-Granular Graph Learning with Fine-Grained Behavioral Pattern Awareness for Session-Based Recommendation

Ming Li[1], Zihao Yan[2], Yuting Chen[3*], Lixin Cui[4*], Lu Bai[5], Feilong Cao[6], Ke Lv[7,8], Zhao Li[9]

[1]Zhejiang Key Laboratory of Intelligent Education Technology and Application, Zhejiang Normal University, Jinhua, China
[2]School of Computer Science and Technology, Zhejiang Normal University, Jinhua, China
[3]Centre for Learning Sciences and Technologies, The Chinese University of Hong Kong, Hong Kong, China
[4]Central University of Finance and Economics, Beijing, China.
[5]School of Artificial Intelligence, Beijing Normal University, Beijing, China
[6]School of Mathematical Sciences, Zhejiang Normal University, Jinhua, China
[7]School of Engineering Science, University of Chinese Academy of Sciences, Beijing, China
[8]Peng Cheng Laboratory, Shenzhen, China
[9]Zhejiang Lab, Hangzhou, China
mingli@zjnu.edu.cn, yzhaoian2@zjnu.edu.cn, yuting.chen@cuhk.edu.hk, cuilixin@cufe.edu.cn, bailu@bnu.edu.cn, caofeilong88@zjnu.edu.cn, luk@ucas.ac.cn, lzjoey@gmail.com

Abstract

Session-based recommendation aims to predict users' next actions by modeling their ongoing interaction sequences, particularly in scenarios where long-term user profiles are unavailable. While existing methods have achieved promising results by leveraging sequential and graph-based structures, they often rely on global aggregation strategies that emphasize dominant user interests while overlooking the transient and fine-grained behavior patterns embedded in sessions. In practice, user intent evolves across sessions and is reflected through diverse behavioral patterns, ranging from immediate preferences to segmented co-occurrence interests and long-range goals. To address these limitations, we propose **GraphFine**, a novel multi-granular graph learning framework that achieves fine-grained behavioral pattern awareness for session-based recommendation. Our approach models user behavior at different temporal and semantic granularities through a combination of graph and hypergraph neural networks. Specifically, we employ a position-aware graph to capture short-term item transitions, and construct segmented co-occurrence hypergraphs to uncover high-order semantic relations among co-occurred items. To preserve diverse user intents, we further introduce a multi-view intent readout mechanism that extracts and adaptively integrates intent signals from short-term actions, segmented co-occurrence patterns, and entire sessions. Extensive experiments on benchmark datasets demonstrate that **GraphFine** consistently outperforms existing state-of-the-art methods, confirming its effectiveness in capturing fine-grained and dynamic user preferences for more accurate recommendation.

1 Introduction

Recommendation systems play a pivotal role in alleviating information overload and enhancing user experience across diverse domains, including e-commerce, streaming platforms, and digital content delivery (Wu et al. 2022; Wang et al. 2021; Li et al. 2024a; Zhang et al. 2025b). While traditional recommendation approaches primarily depend on long-term user profiles and historical behavior records, these assumptions often break down in real-world scenarios where users interact in anonymous or transient sessions. In such contexts, Session-Based Recommendation (SBR) has emerged as a critical paradigm, which aims to predict the next item of interest based solely on a user's ongoing interaction sequence (Li et al. 2024b).

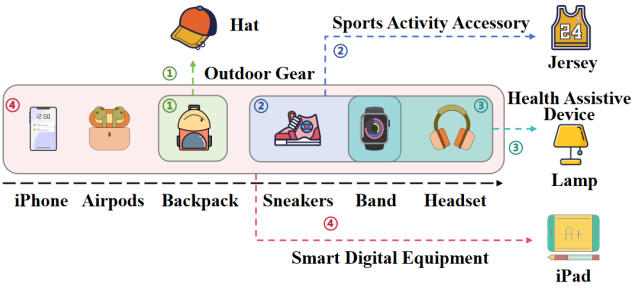

Figure 1: An example in session-based recommendation, illustrating the fine-grained intent within a session and the coarse-grained intent captured by global aggregation.

To effectively model session dynamics, earlier approaches have explored sequential architectures such as recurrent neural networks (Hidasi et al. 2016) and attention mechanisms (Ouyang et al. 2023; Zhang et al. 2025a). More recently, graph-based models have shown compelling results by capturing complex item transitions and session structures (Wu et al. 2019; Gupta et al. 2019; Zhang et al. 2024), and recent works further employ hypergraphs to encapsulate high-order dependencies within sessions (Xia et al. 2021b). These models typically focus on encoding global session representations by aggregating item-level embeddings, often through attention mechanisms. However, global aggregation

*Corresponding authors: Yuting Chen, Lixin Cui

strategies, while effective in highlighting dominant preferences, tend to overlook the diversity of user interests, particularly transient or segmented intent shifts. In reality, user behavior within a session is rarely monolithic. Instead, it often comprises multiple, coexisting intent patterns that vary in temporal and semantic granularity. For example, as illustrated in Figure 1, the user's session begins with an interest in digital equipment, such as an iPhone, but soon exhibits a shift toward a backpack (see ①), suggesting a short-term intent potentially related to outdoor activities. This is followed by interactions with sneakers and a band (see ②), indicating the emergence of interest in sports-related accessories. Later in the session, continued engagement with the band and the addition of a headset (see ③) may reflect an evolving concern with sleep or personal health. Despite the session as a whole (see ④) being dominated by digital product interests, these interactions reveal distinct and temporally localized transitions in user intent. Such patterns highlight the dynamic and multi-faceted nature of user behavior, which often involves rapid, context-dependent shifts. However, existing models largely emphasize the dominant intent captured through global aggregation, thereby failing to account for these fine-grained and transitional interest signals.

Motivated by these observations, we argue that an effective session-based recommendation framework should explicitly model user behavior patterns at multiple granularities to capture both global and fine-grained interests. To this end, we identify three representative behavioral patterns frequently manifested within sessions:

- **Short-term intent**: Temporary deviations from the dominant preference that reflect context-dependent needs or exploratory behavior (*w.r.t.* ①);
- **Segmented co-occurrence interest**: Coherent intent patterns induced by co-occurring items within contiguous session segments, often reflecting specific sub-goals (*w.r.t.* ② ③);
- **Long-term preference**: Dominant and persistent preferences maintained throughout the session, indicative of overarching user objectives (*w.r.t.* ④).

These patterns are not mutually exclusive but often coexist and evolve dynamically. This raises two major challenges: (i) how to model session behavior across multiple temporal and semantic granularities, and (ii) how to preserve and effectively fuse diverse user intents during prediction.

In this paper, we propose **GraphFine**, a multi-granular graph learning framework with fine-grained behavioral pattern awareness for session-based recommendation. The core idea is to construct behavior representations from multiple views using both graph and hypergraph structures, enabling the model to disentangle and retain diverse user intents. Specifically, our framework consists of three key components: i) **Position-Aware Graph Learning**: We model short-term intent by constructing position-sensitive item transition graphs, capturing how user interests evolve over time; ii) **Segmented Co-occurrence Hypergraph Modeling**: We partition each session into variable-length subsequences and construct a hypergraph to model high-order item co-occurrence patterns within these segments, enabling the network to effectively capture the semantic structure of segmented co-occurrence interest; iii) **Multi-View Intent Readout and Integration**: We design a multi-branch decoding mechanism that extracts intent representations from short-term positions, segmented patterns, and the entire session. These representations are then adaptively fused to produce the final recommendation score. This design departs from traditional global aggregation by explicitly modeling and integrating multi-granular intent signals, leading to more nuanced and accurate intent inference.

Our contributions can be summarized as follows:

- We propose a novel fine-grained user behavior modeling framework, **GraphFine**, which employs graph and hypergraph neural networks to capture session dynamics across multiple granularities.
- We introduce a multi-view intent readout module that decouples short-term, segmented, and long-term preferences and adaptively fuses them to produce a more comprehensive session representation.
- We conduct extensive experiments on benchmark datasets, demonstrating that our method consistently outperforms state-of-the-art SBR models.

2 Preliminaries

2.1 Problem Statement

Let $\Psi = \{v_1, v_2, \ldots, v_N\}$ represent the set of all items, where N denotes the total number of unique items. A session is defined as an anonymous sequence of user interactions, denoted by $S = [v_1, v_2, \ldots, v_k, \ldots, v_L]$, where v_k is the item interacted with at the k-th position in session S, and L is the total length of the session. The entire session dataset is denoted by $\Omega = \{S_1, S_2, \ldots, S_M\}$, where M is the number of sessions available for training and evaluation. Given a session $S = [v_1, v_2, \ldots, v_k, \ldots, v_L]$, the goal of session-based recommendation is to predict the next item $v_{L+1} \in \Psi$ that the user is most likely to interact with. Technically, the task involves learning a model that assigns a relevance score to each candidate item $v \in \Psi$, and ranking the items accordingly to recommend the top-K items with the highest predicted scores.

2.2 Multi-Granular Graph Construction

To model diverse and fine-grained user behavior patterns within sessions, we construct three types of graph structures: contextual item graph, semantic item graph, and segmented co-occurrence hypergraph, each targeting different aspects of session dynamics across multiple granularities.

Contextual Item Graph. SBR often suffers from data sparsity due to the fragmented and transient nature of user sessions. To mitigate this issue, we construct a contextual item graph $G_c = (\mathcal{V}_c, \mathcal{E}_c)$, where $\mathcal{V}_c := \Psi$ represent the global set of items. An edge is added between two items v_i and v_j if they co-occur within a sliding window of size r in any session. The edge weight reflects the frequency of their co-occurrence across all sessions, capturing general item transition patterns observed at the global level.

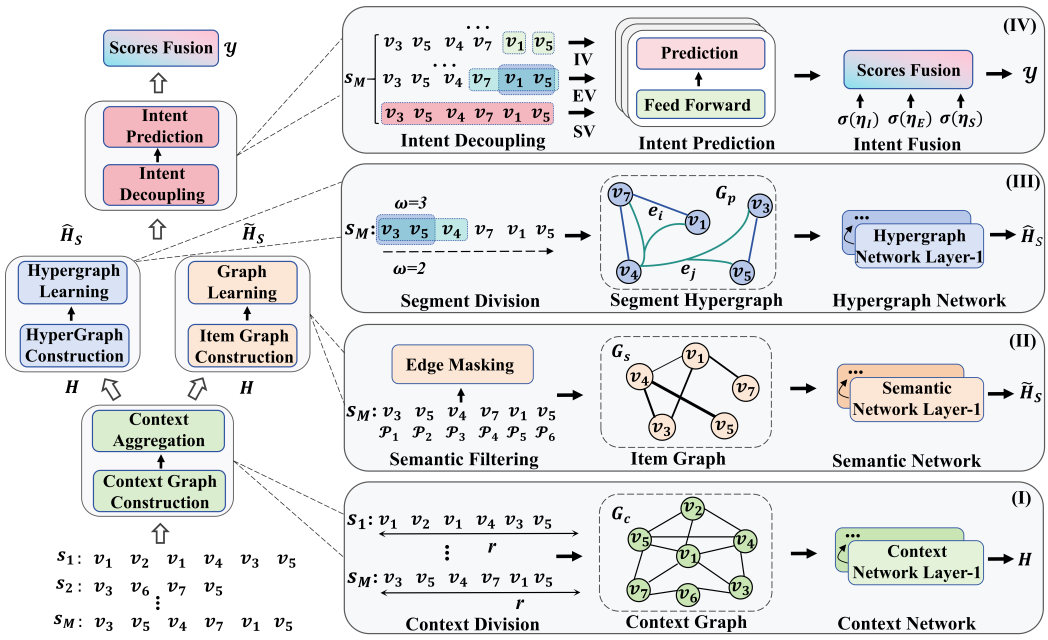

Figure 2: The architecture of **GraphFine**, consisting of four components: (I) contextual aggregation across sessions, (II) short-term semantic modeling with position-aware encoding, (III) high-order behavior extraction via a segmented co-occurrence hypergraph, and (IV) multi-view intent prediction.

Semantic Item Graph. To enhance short-term intent modeling and capture latent sequential transitions among items, we construct a position-aware semantic item graph $G_s = (\mathcal{V}_s, \mathcal{E}_s)$. $\mathcal{V}_s \subseteq \Psi$ denotes the set of items within the current session, and $\mathcal{E}_s$ encodes pairwise semantic similarities. An edge exists between items (v_i, v_j) if their semantic similarity is positive: $\epsilon_{ij} = \frac{\exp(\text{CosSim}(\tilde{\mathbf{h}}_i, \tilde{\mathbf{h}}_j))}{\sum_{k \in \mathcal{N}_i} \exp(\text{CosSim}(\tilde{\mathbf{h}}_i, \tilde{\mathbf{h}}_k))}$, where $\mathcal{N}_i$ denotes the set of semantically relevant neighbors of item v_i. This graph emphasizes local semantic proximity among items based on their learned representations.

Segmented Co-occurrence Hypergraph. In practical scenarios, users often engage with clusters of semantically or functionally related items within localized fragments of a session. To model such behavior, we construct a segmented co-occurrence hypergraph $G_p = (\mathcal{V}_p, \mathcal{E}_p)$ for each session, where $\mathcal{V}_p \subseteq \Psi$ corresponds to the set of items in the session and $\mathcal{E}_p$ contains hyperedges that connect groups of co-occurring items. Sliding windows of varying sizes are applied to generate hyperedges, which are then merged as $\mathcal{E}_p = \cup_{\omega=2}^{\mu} \mathcal{E}_p^{\omega}$, allowing the hypergraph to capture high-order, multi-scale behavioral patterns. This design enables the model to account for localized intent segments that may not be apparent in pairwise interactions.

3 Proposed Method: GraphFine

Figure 2 shows the overall framework of **GraphFine**, which comprises four components: (**I**) cross-session item network for contextual aggregation, (**II**) semantic network for modeling short-term behaviors at different temporal positions, (**III**) hypergraph network for capturing behavior segments, and (**IV**) multi-view prediction network that integrates user intent across multiple granularities for recommendation. In the following, we describe each component in detail.

3.1 Contextual Aggregation Network

Let $\mathbf{X} \in \mathbb{R}^{N \times d}$ denote the item embedding matrix, where d is the dimensionality of each item vector and $\mathbf{x}_i \in \mathbb{R}^d$ represents the embedding of item v_i. To ensure numerical stability and consistent scaling across embeddings, we apply l_2 normalization to each vector: $\hat{\mathbf{x}}_i = \mathbf{x}_i / \|\mathbf{x}_i\|$, resulting in the normalized embedding matrix $\hat{\mathbf{X}}$. To incorporate contextual information from global co-occurrence patterns, we enhance item representations using a neighbor aggregation network constructed over the contextual item graph $G_c = (\mathcal{V}_c, \mathcal{E}_c)$. In this graph, each item v_i aggregates information from its neighboring items $\mathcal{N}(i)$. The aggregation process begins with initializing each item's hidden state as $\mathbf{h}_i^{(0)}$. At each iteration, similarity scores between the target item and its contextual neighbors are calculated. An attention-based mechanism is then applied to adaptively fuse the neighbor embeddings. The aggregation coefficient $a_{ij}^{(\tau+1)}$ is computed as follows:

$$a_{ij}^{(\tau+1)} = \frac{\exp\left(\hat{\mathbf{x}}_j^\top \mathbf{h}_i^{(\tau)}\right)}{\sum_{j \in \mathcal{N}(i)} \exp\left(\hat{\mathbf{x}}_j^\top \mathbf{h}_i^{(\tau)}\right)} \tag{1}$$

The embedding of the target item is then updated as:

$$\mathbf{h}_i^{(\tau+1)} = \hat{\mathbf{x}}_i + \sum_{j \in \mathcal{N}(i)} a_{ij}^{(\tau+1)} \cdot \hat{\mathbf{x}}_j \tag{2}$$

This aggregation procedure is repeated for a fixed number of iterations. The final contextualized item representations are

denoted by $\mathbf{H} \in \mathbb{R}^{N \times d}$, which serve as enhanced input for downstream modules.

3.2 Item Semantic Network

While the contextual aggregation network captures cross-session item relationships, user interests within a session often evolve due to temporal drift and intent transitions. To better capture these dynamic short-term preferences, we construct a session-specific, position-aware semantic item graph G_s for each session S. Instead of statically defining edges, we dynamically compute the item-item interaction graph based on both semantic similarity and positional context. Specifically, we first incorporate sequential order via a learnable positional encoding matrix $\mathbf{P} \in \mathbb{R}^{L \times d}$, which is added to the item embeddings. Let $\mathbf{H}_S \in \mathbb{R}^{L \times d}$ denote the sequence of item embeddings in session S, the initial input to the semantic network is defined as:

$$\tilde{\mathbf{H}}_S^{(0)} = \mathbf{H}_S + \mathbf{P}. \tag{3}$$

At propagation layer τ, the session-level item relationship matrix is computed as:

$$\mathcal{W}_S^{(\tau)} = \frac{\tilde{\mathbf{H}}_S^{(\tau)} (\tilde{\mathbf{H}}_S^{(\tau)})^\top}{\|\tilde{\mathbf{H}}_S^{(\tau)}\|^2}. \tag{4}$$

In particular, to ensure only valid edges are considered, we apply an edge mask $\mathbf{M}_S^{(\tau)} \in \{-\infty, 0\}^{L \times L}$, indicating whether two items are positively correlated:

$$\mathbf{M}_{S,ij}^{(\tau)} = \begin{cases} 0, & \mathcal{W}_{S,ij}^{(\tau)} > 0 \\ -\infty, & \text{otherwise} \end{cases}. \tag{5}$$

We then apply a softmax function to $(\mathcal{W}_S^{(\tau)} + \mathbf{M}_S^{(\tau)})$ to compute the updated item relationship matrix $\mathbf{A}_S^{(\tau)}$. The item embeddings are updated by aggregating representations from semantically relevant neighbors:

$$\mathbf{H}_S^{(\tau+1)} = \mathbf{A}_S^{(\tau)} \tilde{\mathbf{H}}_S^{(\tau)}. \tag{6}$$

In addition, a residual gating mechanism is applied to preserve the initial embedding while incorporating the aggregated information:

$$\mathbf{G}_S^{(\tau)} = \sigma\left(\left[\mathbf{H}_S^{(\tau+1)} \| \tilde{\mathbf{H}}_S^{(0)}\right] \mathbf{W}_g\right), \tag{7}$$

where $\mathbf{W}_g \in \mathbb{R}^{2d \times 1}$ is a learnable parameter matrix.

This results in the updated embedding representation as follows:

$$\tilde{\mathbf{H}}_S^{(\tau+1)} = \mathbf{G}_S^{(\tau)} \odot \mathbf{H}_S^{(\tau+1)} + \left(1 - \mathbf{G}_S^{(\tau)}\right) \odot \tilde{\mathbf{H}}_S^{(0)}. \tag{8}$$

After a fixed number of propagation steps, we obtain the final session-aware item representations $\tilde{\mathbf{H}}_S \in \mathbb{R}^{L \times d}$.

3.3 Segmented Hypergraph Network

User behaviors often form coherent segments, where interactions exhibit strong temporal and semantic continuity. These segments provide structured and informative cues for inferring user intent, offering richer signals than isolated item interactions. To capture such segmented patterns, we adopt a hypergraph-based modeling strategy in which each hyperedge connects a group of consecutive items within a session. For a session S, we construct a segmented co-occurrence hypergraph $G_p = (\mathcal{V}_p, \mathcal{E}_p)$, and apply a hypergraph attention network to perform message passing between nodes and hyperedges, enabling the model to capture high-order dependencies.

Let $\hat{\mathbf{h}}_j^{(0)} \in \mathbf{H}$ denote the initial embedding of node v_j. The initial representation of a hyperedge e_k is computed by averaging the embeddings of its incident nodes: $\mathbf{m}_k^{(0)} = \frac{1}{|e_k|} \sum_{v_j \in e_k} \hat{\mathbf{h}}_j^{(0)}$, where $|e_k|$ denotes the number of nodes incident to hyperedge e_k.

In the *node-to-hyperedge* aggregation step, each hyperedge e_k updates its representation by attending to the embeddings of its connected nodes:

$$\mathbf{m}_k^{(\tau+1)} = \sum_{v_j \in e_k} \alpha_{kj} \cdot \hat{\mathbf{h}}_j^{(\tau)}, \tag{9}$$

where α_{kj} is the attention weight of node v_j in hyperedge e_k, computed as:

$$\alpha_{kj} = \frac{\exp(g(\mathbf{m}_k^{(\tau)}, \hat{\mathbf{h}}_j^{(\tau)}))}{\sum_{v_j \in e_k} \exp(g(\mathbf{m}_k^{(\tau)}, \hat{\mathbf{h}}_j^{(\tau)}))}. \tag{10}$$

Here, $g(\cdot)$ measures the similarity between a hyperedge and a node,i.e.,

$$g(\mathbf{m}_k^{(\tau)}, \hat{\mathbf{h}}_j^{(\tau)}) = \frac{\left(\mathbf{m}_k^{(\tau)} \cdot \mathbf{W}_k\right) \left(\hat{\mathbf{h}}_j^{(\tau)} \cdot \mathbf{W}_j\right)^\top}{\sqrt{d}}, \tag{11}$$

where $\mathbf{W}_k, \mathbf{W}_j \in \mathbb{R}^{d \times d}$ are learnable parameter matrices.

In the *hyperedge-to-node* aggregation stage, each node v_j aggregates information from incident hyperedges $\mathcal{E}_{v_j}$ as follows:

$$\hat{h}_j^{(\tau+1)} = \sum_{e_k \in \mathcal{E}_{v_j}} \beta_{jk} \cdot \mathbf{m}_k^{(\tau+1)}, \tag{12}$$

where the attention coefficient β_{jk} is defined as:

$$\beta_{jk} = \frac{\exp(g(\mathbf{m}_k^{(\tau+1)}, \hat{\mathbf{h}}_j^{(\tau)}))}{\sum_{e_k \in \mathcal{E}_{v_j}} \exp(g(\mathbf{m}_k^{(\tau+1)}, \hat{\mathbf{h}}_j^{(\tau)}))} \tag{13}$$

This iterative *node–hyperedge–node* message passing process, which aligns with the general paradigm of hypergraph convolution, enables the network to capture high-order structural dependencies among items and to effectively represent segmented user intent within behavioral segments.

3.4 Prediction Module

User intent within a session can manifest through diverse behavioral patterns, each reflecting different aspects of user preference. To capture this diversity, we propose a multi-view prediction module that jointly models short-term, segmented, and long-term user intentions to generate recommendation scores.

Item View. To capture short-term intent, we utilize the most recent π item embeddings from the session representation $\tilde{\mathbf{H}}_S$, as they are highly indicative of immediate user preferences. Instead of aggregating these embeddings directly, we treat each as an independent predictive signal and project them into an intent-aware space:

$$\mathbf{Z}_I = \tilde{\mathbf{H}}_{S,[L-\pi+1:L]}\mathbf{W}_I + \mathbf{b}_I, \tag{14}$$

where $\mathbf{W}_I \in \mathbb{R}^{d\times d}$ and $\mathbf{b}_I \in \mathbb{R}^d$ are learnable parameters. $\mathbf{Z}_I \in \mathbb{R}^{\pi\times d}$ contains the transformed representations for recent items.

The prediction scores for candidate items are computed based on cosine similarity between the transformed vectors $\mathbf{Z}_I$ and the candidate item embeddings $\mathbf{H}$, that is,

$$\mathbf{Y}_I = \text{CosSim}\left(\mathbf{Z}_I, \mathbf{H}\right) \tag{15}$$

We then apply a combination of max pooling and mean pooling to aggregate the score matrix $\mathbf{Y}_I \in \mathbb{R}^{\pi\times N}$, yielding the short-term prediction result:

$$\mathcal{Y}_I = \lambda \cdot \text{MaxPool}(\mathbf{Y}_I) + (1-\lambda)\cdot \text{MeanPool}(\mathbf{Y}_I), \tag{16}$$

where $\lambda \in [0,1]$ balances the two pooling strategies.

Segment View. To model segmented co-occurrence interest, we leverage the hypergraph G_p, which segments a session into multiple behavior fragments. For each sliding window of size ω, we extract the embedding of its last hyperedge $e_{\omega'}$ as the key feature for intention prediction:

$$\hat{\mathbf{z}}_{\omega'} = \left(\sum \alpha_{\omega' j}\cdot \hat{\mathbf{h}}_j\right)\mathbf{W}_E + \mathbf{b}_E, \tag{17}$$

where $\alpha_{\omega' j}$ denotes similarity score between hyperedge $e_{\omega'}$ and node v_j. $\mathbf{W}_E \in \mathbb{R}^{d\times d}$, $\mathbf{b}_E \in \mathbb{R}^d$ are learnable parameters. Let $\mathbf{Z}_E \in \mathbb{R}^{\mu\times d}$ represent the set of transformed hyperedge embeddings. The prediction scores are computed as:

$$\mathbf{Y}_E = \text{CosSim}\left(\mathbf{Z}_E, \mathbf{H}\right). \tag{18}$$

The score matrix $\mathbf{Y}_E$ is then aggregated via a similar pooling strategy:

$$\mathcal{Y}_E = \gamma\cdot \text{MaxPool}(\mathbf{Y}_E) + (1-\gamma)\cdot \text{MeanPool}(\mathbf{Y}_E), \tag{19}$$

where $\gamma \in [0,1]$ controls the pooling balance.

Session View. To capture long-term intent, we construct a session-level representation by aggregating item embeddings within the session. This is achieved using an adaptive attention mechanism that assigns different importance weights to individual items. Given the session representation $\tilde{\mathbf{H}}_S = \left[\tilde{\mathbf{h}}_1, \tilde{\mathbf{h}}_2, \ldots, \tilde{\mathbf{h}}_L\right]$, the attention weights are computed via entmax normalization (Martins and Astudillo 2016):

$$\mathbf{\Gamma} = \text{Entmax}\left(\tilde{\mathbf{H}}_S\mathbf{W}_\phi + \mathbf{b}_\phi\right), \tag{20}$$

where $\mathbf{W}_\phi \in \mathbb{R}^{d\times 1}$, $\mathbf{b}_\phi \in \mathbb{R}$ are learnable parameters. The session embedding is then derived as:

$$\mathbf{Z}_S = \left(\sum \mathbf{\Gamma}_i \cdot \tilde{\mathbf{h}}_i\right)\mathbf{W}_S + \mathbf{b}_S, \tag{21}$$

where $\mathbf{W}_S \in \mathbb{R}^{d\times d}$ and $\mathbf{b}_S \in \mathbb{R}^d$ are learnable parameters.

The final prediction scores from this view are given by:

$$\mathcal{Y}_S = \text{CosSim}\left(\mathbf{Z}_S, \mathbf{H}\right). \tag{22}$$

Training Objective. To integrate the outputs from the three views, we adopt an adaptive fusion mechanism. The final prediction scores are computed as a weighted sum of the three branches, that is,

$$\mathcal{Y} = \sigma(\eta_I)\cdot\mathcal{Y}_I + \sigma(\eta_E)\cdot\mathcal{Y}_E + \sigma(\eta_S)\cdot\mathcal{Y}_S, \tag{23}$$

where $\mathcal{Y}$ denotes the final recommendation scores over all candidate items. η_I, η_E, η_S are learnable parameters. $\sigma(\cdot)$ is a sigmoid function to ensure the weights remain in [0,1].

For model training, we use the binary cross-entropy loss:

$$\mathcal{L}(\Theta) = -\sum_{i=1}^{N}\hat{\mathbf{y}}_i\log(\mathcal{Y}_i) + (1-\hat{\mathbf{y}}_i)\log(1-\mathcal{Y}_i), \tag{24}$$

where $\hat{\mathbf{y}} \in \{0,1\}^N$ is a one-hot vector indicating the ground-truth next item, and $\mathcal{Y} \in \mathbb{R}^N$ represents the predicted scores for all candidate items.

4 Experiments

In this section, we conduct extensive experiments on three benchmark datasets to evaluate the effectiveness of the proposed **GraphFine** model. Specifically, we aim to answer the following research questions:

RQ1: How does **GraphFine** perform compared to existing baseline models?

RQ2: Does the proposed multi-view intent modeling strategy improve recommendation performance when integrated into other session-based models?

RQ3: What is the contribution of each key component in **GraphFine** to the overall model performance?

RQ4: How sensitive is the model to different hyperparameter settings?

RQ5: How well does **GraphFine** perform across sessions of varying lengths?

Statistics	Tmall	Yoochoose	RetailRocket
# of train sessions	351,268	369,869	433,648
# of test sessions	25,898	55,696	15,132
# of items	40,728	17376	36,968
# average length	6.69	6.16	5.43

Table 1: Statistics of the datasets.

4.1 Datasets and Evaluation Metrics

We evaluate the proposed model on three widely used benchmark datasets for session-based recommendation: Tmall, Yoochoose, and RetailRocket. To ensure fair comparison with baselines, we follow standard preprocessing protocols as adopted in prior work (Li et al. 2017; Wu et al. 2019; Wang et al. 2020). The detailed statistics of these datasets are summarized in Table 1.

For evaluation, we adopt two widely used metrics: Hit Rate (HR) and Mean Reciprocal Rank (MRR). HR@K measures whether the ground-truth item appears among the top-K predicted items, while MRR@K reflects the average inverse rank of the ground-truth item, emphasizing ranking

quality. All results are reported at $K = 20$ unless otherwise specified.

4.2 Baselines and Experimental Setups

We compare **GraphFine** with 17 existing session-based recommendation methods. These include: (i) **Traditional models**, such as FPMC (Rendle, Freudenthaler, and Schmidt-Thieme 2010), SKNN (Jannach and Ludewig 2017), and STAN (Garg et al. 2019); (ii) **Sequence-based models**, such as NARM (Li et al. 2017), CSRM (Wang et al. 2019), MTAW (Ouyang et al. 2023), GTPAN (Lu et al. 2024), and DPDM (Luo, Sheng, and Zhang 2024); and (iii) **Graph-based models**, such as SR-GNN (Wu et al. 2019), GCE-GNN (Wang et al. 2020), DHCN (Xia et al. 2021b), COTREC (Xia et al. 2021a), GSN-IAS (Zhang and Wang 2023), SPARE (Peintner, Mohammadi, and Zangerle 2023), RESTC (Wan et al. 2023), SDHID (Gao et al. 2023), and RAIN (Zeng et al. 2025).

For fair comparison, we follow well-established experimental protocols (Wu et al. 2019; Wang et al. 2020). The embedding size and batch size are set to 100. Parameters are initialized uniformly and optimized using Adam with an initial learning rate of 0.001, decayed by a factor of 0.1 every 3 epochs. We apply L_2 regularization with a coefficient of 1×10^{-5} and train for up to 20 epochs with early stopping based on validation performance.

Method	Tmall		Yoochoose		RetailRocket	
	MRR@20	HR@20	MRR@20	HR@20	MRR@20	HR@20
FPMC	7.32	16.06	15.01	45.62	13.82	32.37
SKNN	-	-	25.22	63.77	24.46	54.28
STAN	-	-	28.74	69.45	26.81	53.48
NARM	10.70	23.30	28.63	68.32	24.59	50.22
CSRM	13.96	29.46	29.71	69.85	26.19	51.02
MTAW	19.14	37.17	-	-	30.52	56.39
GTPAN	-	-	31.31	71.17	30.19	55.74
DPDM	-	-	<u>31.52</u>	71.68	30.79	56.29
SR-GNN	13.72	27.57	30.94	70.57	26.57	50.32
GCE-GNN	15.42	33.42	30.84	72.18	28.01	53.63
S^2-DHCN	15.05	31.42	27.89	68.34	27.30	53.66
COTREC	18.04	36.35	29.36*	70.72*	29.97	56.17
GSN-IAS	17.71*	34.95*	31.45	<u>72.34</u>	29.97	57.13
SPARE	<u>20.07</u>	39.28	25.92*	65.62*	30.22	56.91
RESTC	18.52	<u>42.47</u>	-	-	<u>30.82</u>	<u>57.81</u>
SDHID	18.38	37.69	-	-	30.24	57.51
RAIN	19.12	38.73	30.91*	72.32*	29.21	56.88
GraphFine	**21.45**	**48.48**	**31.93**	**72.64**	**31.17**	**58.85**
% Improve	6.87%	14.15%	1.30%	0.30%	1.13%	1.79%

Table 2: Performance comparison between **GraphFine** and 17 baseline models. '*' indicates the re-implemented results; '-' denotes baselines with unavailable code. The best result is shown in **bold**, and the second-best is <u>underlined</u>.

4.3 Overall Performance Comparison (RQ1)

Table 2 summarizes the performance of **GraphFine** compared to 17 baseline models. As shown, **GraphFine** consistently achieves the best results on all datasets in both HR@20 and MRR@20. Compared with traditional methods (FPMC, SKNN, STAN) that rely on simple sequential or neighbor assumptions, **GraphFine** captures rich high-order dependencies and delivers substantial gains. Sequential models (NARM, CSRM) add temporal and cross-session signals but still miss fine-grained intent shifts. Attention-based approaches (MTAW, GTPAN, DPDM) improve adaptability—DPDM is strong on Yoochoose—yet lack the structural flexibility to model behaviors at multiple levels. Graph/hypergraph models (SR-GNN, GCE-GNN, RESTC, SDHID, RAIN) strengthen representations and address noise or higher-order relations, but only read out the global user intent. In contrast, **GraphFine** unifies multi-granular contextual, semantic, and structural representations, enabling more precise intent modeling and yielding state-of-the-art performance across datasets.

4.4 Effectiveness of Fine-Grained Intent Modeling (RQ2)

To further assess the effectiveness of our multi-view intent modeling strategy, we incorporate it into two representative GNN-based session models: SR-GNN and GCE-GNN. For each model, we construct three enhanced variants: **Item-view variants** (SR-IV, GCE-IV) incorporate only the short-term intent modeling branch; **Segment-view variants** (SR-EV, GCE-EV) utilize only the segmented intent modeling branch; **Multi-view variants** (SR-MV, GCE-MV) combine both short-term and segmented intent branches for joint intent modeling. The original models (SR-BS, GCE-BS) are used as baselines for comparison.

Method	Tmall		Yoochoose		RetailRocket	
	MRR@20	HR@20	MRR@20	HR@20	MRR@20	HR@20
SR-BS	13.85	28.80	30.90	71.17	27.55	52.27
SR-IV	<u>16.27</u>	<u>36.04</u>	<u>31.58</u>	<u>71.62</u>	<u>27.96</u>	<u>54.02</u>
SR-EV	14.57	29.39	30.94	71.51	27.94	53.49
SR-MV	**16.62**	**36.19**	**31.74**	**71.88**	**28.09**	**54.27**
GCE-BS	15.17	32.75	30.71	72.01	28.93	55.89
GCE-IV	16.03	**36.51**	<u>31.33</u>	<u>72.03</u>	<u>29.41</u>	<u>56.51</u>
GCE-EV	<u>16.55</u>	34.51	30.86	**72.25**	29.16	55.98
GCE-MV	**16.68**	<u>36.07</u>	**31.66**	72.01	**29.72**	**56.54**

Table 3: Impact of fine-grained intent modeling on existing models: SR-GNN and GCE-GNN.

As shown in Table 3, the item-view variants (IV) yield clear gains, confirming that modeling recent actions at a fine-grained level enhances short-term intent inference. Similarly, the segment-view variants (EV) show substantial improvements, demonstrating the value of capturing segmented, high-order interaction patterns within sessions. Notably, the multi-view variants (MV) achieve the best performance in nearly all cases, highlighting the complementarity of short-term and segmented intent modeling. These results underscore that integrating fine-grained behavioral patterns into existing GNN frameworks offers a practical and effective enhancement for session-based recommendation.

4.5 Ablation Study (RQ3)

We conduct ablations to quantify each module's contribution. Variants: **–CA** (remove contextual aggregation integrating cross-session cues), **–IS** (remove item-semantic network), **–SH** (remove segmented hypergraph network). To probe the prediction views: **–IV** (remove item-view for short-term intent), **–EV** (remove segment-view for segmented intent), **–SV** (remove session-view for long-term intent).

As shown in Table 4, removing the contextual aggregation network (–CA) leads to performance drops, confirming its effectiveness in introducing global context and mitigating sparsity. The item semantic network (–IS) also contributes moderate gains, validating its role in semantic modeling. The segmented hypergraph network (–SH) yields stable improvements, particularly in MRR, indicating its importance for capturing segmented behavioral patterns. The item-view (–IV) proves most critical, with its removal resulting in the largest performance decline, highlighting the significance of short-term intent modeling. The Segment-view (–EV) and session-view (–SV) also contribute steady gains; while individually smaller, they collectively strengthen the model's predictive capability.

Method	Tmall		Yoochoose		RetailRocket	
	MRR@20	HR@20	MRR@20	HR@20	MRR@20	HR@20
GraphFine	**21.45**	**48.48**	**31.95**	**72.64**	**31.17**	**58.85**
-CA	21.01	47.59	31.94	72.20	30.52	57.71
-IS	21.27	47.87	31.83	72.24	30.93	58.52
-SH	21.36	48.34	31.65	72.57	30.85	58.54
-IV	17.12	34.42	31.45	72.23	30.01	56.14
-EV	21.29	48.21	31.63	72.55	30.61	58.43
-SV	21.05	48.36	31.73	72.63	30.08	58.44

Table 4: Ablation studies on different components.

4.6 Hyperparameter Analysis (RQ4)

We analyze the impact of two key hyperparameters: π, determining the number of recent items used for item-view prediction, and μ, which controls the number of sliding windows for generating the hypergraph G_p.

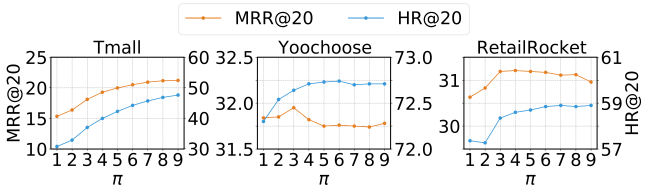

Figure 3: Impact of π on model performance.

We vary π in the range $\{1, 2, \ldots, 9\}$. As illustrated in Figure 3, performance on Tmall improves steadily with larger π, stabilizing beyond a certain threshold. However, the gains are less pronounced on Yoochoose and RetailRocket, where performance plateaus earlier. This is likely due to larger π values incorporating items from earlier in the session, which may introduce redundant or less relevant intent signals.

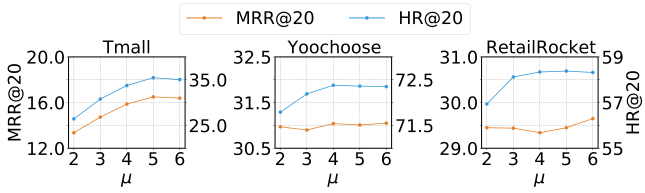

Figure 4: Impact of μ on model performance.

To investigate the impact of μ, we vary its value in $\{2, 3, 4, 5, 6\}$. The corresponding performance results are shown in Figure 4. Increasing μ generally leads to improved performance by enriching the diversity of segmented patterns, with the most significant gains observed on the Tmall dataset. Yoochoose and RetailRocket exhibit moderate yet consistent improvements. However, setting μ too high may introduce noisy or redundant patterns, potentially offsetting the benefits.

4.7 Impact of Different Session Lengths (RQ5)

We divide sessions into three categories based on their length: short ($\leq$ 5), medium (6 $\sim$ 10), and long ($>$ 10). We then compare the performance of GraphFine against two baselines, SR-GNN and GCE-GNN, within each category. Unlike SR-GNN and GCE-GNN, which rely primarily on global session-level intent, **GraphFine** jointly models fine-grained intent signals at the item, segment, and session levels. As shown in Figure 5, this multi-view modeling enables consistent and substantial performance gains across all session length groups. These results highlight **GraphFine**'s effectiveness in handling session diversity.

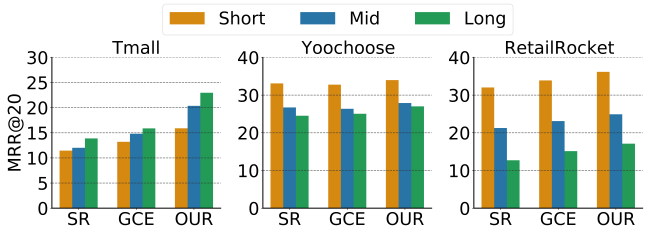

Figure 5: Impact of session length on MRR@20. (SR: SR-GNN; GCE: GCE-GNN)

5 Conclusion

In this paper, we propose **GraphFine**, a multi-granular graph learning framework for session-based recommendation that models behavioral patterns across temporal granularities via joint graph and hypergraph neural networks to effectively capture multi-granular user intent. Extensive experiments on three benchmarks show that **GraphFine** not only consistently surpasses various baselines but also effectively enhances existing GNN-based models.

Acknowledgements

This work was supported in part by the "Pioneer" and "Leading Goose" R&D Program of Zhejiang (No. 2024C03262), and the National Natural Science Foundation of China (No. U21A20473, No. 62536006, No. 62172370, No. 62576371, No. U23A20388, No. 62320106007).

References

Gao, R.; Tao, Y.; Yu, Y.; Wu, J.; Shao, X.; Li, J.; and Ye, Z. 2023. Self-supervised Dual Hypergraph learning with Intent Disentanglement for session-based recommendation. *Knowledge-Based Systems*, 270: 110528.

Garg, D.; Gupta, P.; Malhotra, P.; Vig, L.; and Shroff, G. 2019. Sequence and Time Aware Neighborhood for Session-based Recommendations: STAN. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1069–1072.

Gupta, P.; Garg, D.; Malhotra, P.; Vig, L.; and Shroff, G. M. 2019. NISER: Normalized Item and Session Representations with Graph Neural Networks. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*.

Hidasi, B.; Karatzoglou, A.; Baltrunas, L.; and Tikk, D. 2016. Session-based Recommendations with Recurrent Neural Networks. In *Proceedings of the 4th International Conference on Learning Representations*.

Jannach, D.; and Ludewig, M. 2017. When Recurrent Neural Networks meet the Neighborhood for Session-Based Recommendation. In *Proceedings of the 11st ACM Conference on Recommender Systems*, 306–310.

Li, J.; Ren, P.; Chen, Z.; Ren, Z.; Lian, T.; and Ma, J. 2017. Neural Attentive Session-based Recommendation. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, 1419–1428.

Li, M.; Li, Z.; Huang, C.; Jiang, Y.; and Wu, X. 2024a. EduGraph: Learning path-based hypergraph neural networks for mooc course recommendation. *IEEE Transactions on Big Data*, 10(6): 706–719.

Li, Z.; Yang, C.; Chen, Y.; Wang, X.; Chen, H.; Xu, G.; Yao, L.; and Sheng, M. 2024b. Graph and sequential neural networks in session-based recommendation: A survey. *ACM Computing Surveys*, 57(2): 1–37.

Lu, T.; Xiao, X.; Xiao, Y.; and Wen, J. 2024. GTPAN: Global Target Preference Attention Network for session-based recommendation. *Expert Systems with Applications*, 243: 122900.

Luo, Z.; Sheng, Z.; and Zhang, T. 2024. Dual perspective denoising model for session-based recommendation. *Expert Systems with Applications*, 249: 123845.

Martins, A.; and Astudillo, R. 2016. From softmax to sparsemax: A sparse model of attention and multi-label classification. In *ICML*, 1614–1623. PMLR.

Ouyang, K.; Xu, X.; Chen, M.; Xie, Z.; Zheng, H.-T.; Song, S.; and Zhao, Y. 2023. Mining Interest Trends and Adaptively Assigning Sample Weight for Session-based Recommendation. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2174–2178.

Peintner, A.; Mohammadi, A. R.; and Zangerle, E. 2023. SPARE: Shortest Path Global Item Relations for Efficient Session-based Recommendation. In *Proceedings of the 17th ACM Conference on Recommender Systems*, 58–69.

Rendle, S.; Freudenthaler, C.; and Schmidt-Thieme, L. 2010. Factorizing personalized Markov chains for next-basket recommendation. In *Proceedings of the 19th International Conference on World Wide Web*, 811–820.

Wan, Z.; Liu, X.; Wang, B.; Qiu, J.; Li, B.; Guo, T.; Chen, G.; and Wang, Y. 2023. Spatio-temporal Contrastive Learning-enhanced GNNs for Session-based Recommendation. *ACM Transactions on Information Systems*, 42(2).

Wang, M.; Ren, P.; Mei, L.; Chen, Z.; Ma, J.; and de Rijke, M. 2019. 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*, 345–354.

Wang, S.; Cao, L.; Wang, Y.; Sheng, Q. Z.; Orgun, M. A.; and Lian, D. 2021. A Survey on Session-based Recommender Systems. *ACM Computing Surveys*, 54(7): 1–38.

Wang, Z.; Wei, W.; Cong, G.; Li, X.-L.; Mao, X.-L.; and Qiu, M. 2020. Global Context Enhanced Graph Neural Networks for Session-based Recommendation. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 169–178.

Wu, S.; Sun, F.; Zhang, W.; Xie, X.; and Cui, B. 2022. Graph Neural Networks in Recommender Systems: A Survey. *ACM Computing Surveys*, 55(5): 1–37.

Wu, S.; Tang, Y.; Zhu, Y.; Wang, L.; Xie, X.; and Tan, T. 2019. Session-based recommendation with graph neural networks. In *Proceedings of the 33rd AAAI Conference on Artificial Intelligence*, 346–353.

Xia, X.; Yin, H.; Yu, J.; Shao, Y.; and Cui, L. 2021a. Self-Supervised Graph Co-Training for Session-based Recommendation. In *Proceedings of the 30th ACM International Conference on Information and Knowledge Management*, 2180–2190.

Xia, X.; Yin, H.; Yu, J.; Wang, Q.; Cui, L.; and Zhang, X. 2021b. Self-Supervised Hypergraph Convolutional Networks for Session-based Recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 4503–4511.

Zeng, X.; Li, S.; Zhang, Z.; Jin, L.; Guo, Z.; and Wei, K. 2025. RAIN: Reconstructed-aware in-context enhancement with graph denoising for session-based recommendation. *Neural Networks*, 184: 107056.

Zhang, Q.; Wen, H.; Yuan, W.; Chen, C.; Yang, M.; Yiu, S.-M.; and Yin, H. 2025a. HMamba: Hyperbolic Mamba for Sequential Recommendation. *arXiv preprint arXiv:2505.09205*.

Zhang, Q.; Xia, L.; Cai, X.; Yiu, S.-M.; Huang, C.; and Jensen, C. S. 2024. Graph augmentation for recommendation. In *ICDE*, 557–569.

Zhang, Q.; Yang, P.; Yu, J.; Wang, H.; He, X.; Yiu, S.-M.; and Yin, H. 2025b. A survey on point-of-interest recommendation: Models, architectures, and security. *IEEE Transactions on Knowledge and Data Engineering*.

Zhang, Z.; and Wang, B. 2023. Graph Spring Network and Informative Anchor Selection for session-based recommendation. *Neural Networks*, 159: 43–56.