Behavior Habits Enhanced Intention Learning for Session Based Recommendation

Zhida Qin, Wenhao Xue, Haotian He, Haoyao Zhang, Shixiao Yang, Enjun Du, Tianyu Huang, and John C.S. Lui, *Fellow, IEEE*

***Abstract*—Multi-behavior Session Based Recommendations (MBSBRs) have achieved remarkable results due to considering behavioral heterogeneity in sessions. Yet most existing works only consider binary or continuous behavior dependencies and aim to predict the next item under the target behavior, neglecting users' inherent behavior habits, resulting in learning inaccurate intentions. To tackle the above issues, we propose a novel Behavior Habits Enhanced Intention Learning framework for Session Based Recommendation (BHSBR). Specifically, we focus on the next item recommendation and design a global item transition graph to learn the behavior-aware semantic relationships between items, in order to mine the underlying similarity between items beyond the session. In addition, we construct a hypergraph to extract the diverse behavior habits of users and break through the limitations of temporal relationships in the session. Compared to the existing works, our behavior habit learning method learns behavior dependencies at the user level, which could capture the user's more accurate long-term intentions and reduce the impact of noise behaviors. Extensive experiments on three datasets demonstrate that the performance of our proposed BHSBR is superior to SOTA. Further ablation experiments fully illustrate the effectiveness of our various modules.**

***Index Terms*—Session based recommendation, multi-behavior recommendation, graph neural network.**

I. Introduction

SESSION-BASED recommendation algorithms aim to make the next-item prediction for anonymous, temporal sessions [1], [2], [3]. Due to the increasingly improved protection of user privacy in recent years, session-based recommendations have gradually become a hot research area. Traditional methods [4], [5] utilize Markov chains to encode sessions. However, these methods can only predict the latter term based on the previous one, which results in insufficient utilization of all sequence information of the session. To address this issue, some methods [6], [7], [8] leverage RNN to model sessions and capture temporal relationships in the sequence. Recently, due to the superiority of GNN in capturing structural features, many methods [9], [10], [11] have introduced GNN into the process of feature learning and achieved sound results.

Whereas, the above methods only consider the macro connectivity between items, without utilizing the complex behavioral information between users and items, and cannot capture fine-grained user preferences. For example, in shopping scenarios, compared to products that users have only browsed, users are clearly more interested in products that are more relevant to those they have added to their shopping cart or have already purchased. Therefore, when considering behavioral information, we are better able to capture users' intention. Recently, some methods have introduced multi-behavior information into session recommendation. TGT [12] proposed a temporal graph Transformer to capture dynamic user-item interaction patterns by exploring the evolutionary correlations between behaviors. MBHT [13] designed a multi-scale transformer to learn the behavioral order patterns of sequences and utilized hypergraphs to learn cross-behavioral dependencies. EMBSR [14] designed multi graph to aggregate micro-operations in sequences and utilized extended attention mechanisms to learn binary behavior patterns.

Although these methods have achieved delightful results in aggregating sequence representations of behavioral information and mining relationships between behavioral dependencies, there are still some issues that need to be addressed:

- *Failing to learn behavior habits:* During user-item interactions, users exhibit personalized behavioral patterns. For instance, a user might follow a specific behavioral sequence (e.g., [click, view reviews, contact customer service, add to cart, purchase]) when engaging with an item. We define such sequences as behavior habits, which directly reflect user intent - richer behavioral sequences toward an item typically indicate stronger user interest. However, user interactions with items often exhibit randomness and discontinuity. Current methods [15], [16] ignore this issue and separately model discrete interactions with the same item,

Received 18 March 2025; revised 26 July 2025; accepted 25 September 2025. Date of publication 6 October 2025; date of current version 14 January 2026. This work was supported in part by the National Key R&D Program of China under Grant 2023YFC3305700 and in part by the Open Research Fund from Guangdong Laboratory of Artificial Intelligence and Digital Economy (SZ), under Grant GML-KF-24-22. Recommended for acceptance by H. Liu. *(Corresponding author: Zhida Qin.)*

Zhida Qin is with the School of Computer Science & Technology, Beijing Institute of Technology, Beijing 100081, China, and also with the Guangdong Laboratory of Artificial Intelligence and Digital Economy (SZ), Shenzhen 518107, China (e-mail: zanderqin@bit.edu.cn).

Wenhao Xue, Haotian He, Haoyao Zhang, Shixiao Yang, and Tianyu Huang are with the School of Computer Science & Technology, Beijing Institute of Technology, Beijing 100081, China (e-mail: aqxuewenhao@bit.edu.cn; hthe@bit.edu.cn; zhanghaoyao@bit.edu.cn; ysx144_51@bit.edu.cn; huangtianyu@bit.edu.cn).

Enjun Du is with the School of Cyberspace Science and Technology, Beijing Institute of Technology, Beijing 100081, China (e-mail: enjundu666@gmail.com).

John C.S. Lui is with the Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong (e-mail: cslui@cse.cuhk.edu.hk).

Digital Object Identifier 10.1109/TBDATA.2025.3618463

thereby failing to capture users' comprehensive behavior habits toward items. This limitation means they fail to capture all behavior habits towards items for the user.

- *Predicting the targeted behaviors is not reasonable in session-based recommendations:* Current MBSBR methods often define a target behavior and treat other behaviors as auxiliary, focusing on predicting the next item for that target. However, in reality, users frequently engage in auxiliary behaviors before performing target behavior. This approach can lead to systems recommending items that users have already interacted with rather than new ones, which contradicts the purpose of recommendation systems and may degrade the user experience.Our analysis of real-world session data (Section III-B2) shows that target behaviors such as "buy" are extremely sparse, the proportion of behavior habits including "buy" is less than 0.2% overall, and the share of single "buy" actions is only 2% in the *Appliances* dataset and 0.3% in *Computers*. This indicates that relying on predefined target behaviors offers little predictive signal in practice, highlighting the need to model richer and more diverse user behavior habits.

To address the aforementioned issues, we propose a novel Behavior Habits Enhanced Intention Learning framework for Session Based Recommendation (**BHSBR**). First, we focus on achieving the next-item recommendation that users may be interested in, and design an intention learning module for global semantic fusion behavior. This module could efficiently integrate the global item transition relationships and internal behavior temporal information, which further learn the behavior-aware semantic relationships between items. Technically, we construct a global item interaction graph to extract semantic connections between items. Then, we aggregate behavior sequences under macro items and fuse semantic connections with behavioral information through gating mechanisms to capture fine-grained user intentions. Meanwhile, we capture the coarse-grained intent representation contained in the macro item sequence of the session, aggregate it with fine-grained intent, and learn the final intention representation of the user. Second, our approach leverages hypergraphs to capture user behavior habits and deduce dependencies at the user level. To elaborate, we designate items as vertices within the hypergraph and construct hyperedges that encompass all item-related behaviors occurring within a session. This technique enables us to extract user behavior habits from sessions, transcending constraints of time and number, thereby enhancing user intentions and reducing noise within behavioral sequences. Concurrently, to counteract the challenges of gradient explosion and over-smoothing, we add residual connections when constructing the hypergraph. Finally, we design a mutual information fusion module that adaptively fuses the information of the two modules to learn the final representation of the session.

In summary, the main contributions of this paper are as follows:

- We propose the **BHSBR** framework, which models user-level behavior habits through hypergraph construction. Unlike previous methods that mainly focus on item-level or single-behavior dependencies, our approach integrates multiple behavior types to capture richer user-specific patterns. By combining this with a global item interaction graph, our model learns more comprehensive session intent representations.
- We address a more flexible session-based recommendation setting that discovers items of potential interest, rather than restricting prediction to a single predefined behavior (e.g., purchase). Our method constructs a global item interaction graph to learn semantic relationships across items and adaptively integrates behavioral information to better model complex user intentions.
- We conduct extensive experiments on three real-world datasets, and our model outperforms SOTA, demonstrating the effectiveness of our framework. Meanwhile, further ablation experiments have also verified that each of our modules has its own effect.

The remainder of this paper is organized as follows. Section II briefly reviews current related work. Section III analyzes the data and defines the multi-behavior session-based recommendation. Section IV introduces the framework of BHSBR in detail. Section V analyzes the results of experiments. Finally, Section VI summarizes the paper.

II. Related Work

A. Session-Based Recommendation

Due to the lack of user information and temporal interaction, session recommendation algorithms differ from general recommendation paradigms [17], [18]. FPMC [4] combined the advantages of Matrix Factorization (MF) and Markov Chain (MC) to achieve personalized next recommendation. However, the MC-based method can only calculate similarity based on the last item in the session, ignoring the information on the previous items. RNN can effectively solve this problem. Hereby, GRU4Rec [6] first utilized the GRU method to model sessions and learn the temporal preferences of the entire session. Meanwhile, as the attention mechanism demonstrated its superiority in capturing sequence temporal features [19], NARM [7] added an attention layer to GRU4Rec for generating session representations. STAMP [20] considered both the general interest of long-term memory in the session context and the short-term interest of the last click, and utilized attention mechanisms to obtain the final representation. Bert4Rec [21] designed a deep bidirectional self-attention strategy to capture correlations between items. Most of these RNN and attention-based methods focused on capturing temporal features while paying insufficient attention to structural features in sessions.

Recently, more and more methods have introduced GNN into session recommendation. SRGNN [22] was the first to design a gated GNN framework for learning the sequential structure in a session. SGNN-HN [23] designed a star GNN model to capture high-order information that may exist between non-adjacent items in a session, and designed a Highway Network to adaptively learn embedding information from item representations to alleviate the overfitting problem in GNN. The above methods only encoded a single session and did not effectively utilize the global information of the data. In order to fully utilize the information in the dataset, GCE-GNN [24] learned sequence representation by constructing two views, a session graph and a global graph. COTREC [25] introduced contrastive learning into session recommendation, generated labels using different connection relationships, and supervised through contrastive learning strategies to obtain session representations. MCLRec [26] proposed a meta-optimized contrastive learning framework, which uses meta-learning to guide model enhancers in updating, while considering contrastive regularization terms to solve the problems of previous contrastive pairs being difficult to generalize and lacking available information. Furthermore, some methods use the idea of hypergraphs to model sessions. DHCN [27] first modeled session data into hypergraphs to learn session embedding. HIDE [28] mapped an item to multiple embeddings, each corresponding to an intention, and disentangled the intentions from both macro and micro perspectives to achieve prediction. Recent advancements have further explored the integration of multi-behavior signals and hypergraph structures in session-based recommendation. MB-GRL [29] combines graph neural networks with deep reinforcement learning to better capture user intent across multiple behaviors. MGCOT [30] employs multi-graph co-training, integrating current session graphs, similar session graphs, and global item relation graphs to comprehensively capture user intent. Additionally, HDNN [31] introduces a hypergraph denoising neural network that leverages sequential pattern learning and adaptive attention mechanisms to enhance the robustness of hypergraph modeling in session-based recommendation.

In addition to the above architectures, recent studies have explored contextual modeling to enhance session-based recommendation. DCARS [32] captures preference shifts by modeling latent contextual representations through the integration of short-term session behaviors and long-term user profiles. HAN [33] introduces a hierarchical attention network that jointly models temporal dependencies and semantic attribute relationships to enable intention-aware sequence modeling. These works highlight the advantages of incorporating contextual and semantic information. In contrast, our method focuses on capturing fine-grained multi-behavioral dynamics within sessions and disentangling latent user intentions, providing a complementary perspective to context-aware approaches.

B. Multi-Behavior Recommendation

The multi-behavior recommendation aims to use behavioral data in interactions to encode users and items, capture dynamic preferences in behavior, and improve recommendation performance [34], [35], [36], [37]. The traditional multi-behavior recommendation algorithm [38] extended the matrix factorization algorithm to achieve multi-behavior data encoding. However, such methods cannot capture higher-order information about users and items at the behavioral level. Most recent methods utilized GNN to propagate the messages contained in behaviors in order to capture complex transformation relationships [39], [40], [41]. CML [42] and MBSSL [43] combined GNN and self-supervised learning methods to capture dependencies and diversity between behaviors. However, user behavior often contains deep information in chronological order. These methods only considered the behavioral interaction information between users and items and did not delve into the deep-level temporal relationships of behaviors. RIB [44] introduces RNN layers into the model to learn fine-grained behavior sequence representations. MKM-SR [45] designed a multi-task learning framework that combined knowledge graphs with session sequences and behavioral information for loss calculation. PBAT [46] developed a personalized behavior pattern generator to extract unique user behavior patterns and designed a behavior-aware collaborative extractor that utilizes attention mechanisms to extract collaborative relationships in sequences. In the realm of multi-behavior modeling, recent studies have proposed innovative approaches to capture diverse user interactions. MiaSRec [47] introduces a multi-intent-aware session-based recommendation model that utilizes frequency embedding vectors to enhance information about repeated items and generate multiple user intent representations. Furthermore, the integration of hypergraph structures with contrastive learning has been explored in MBHCL [48], where a multi-behavior hypergraph contrastive learning framework is proposed to effectively model complex user behavior patterns in session-based recommendation. Although these methods have learned behavioral dependencies, they have not explored deeper behavioral habits and cannot better capture users' personalized behavioral preferences.

III. Preliminary

A. Behavior Habits Definition

Behavior habit is defined as a sequence composed of all behaviors the user taken towards a specific item within a session. For example, in session s_e, if there are browsing, adding to cart, and purchasing operations on item v_i, it is considered that the *behavior habit* for v_i of the user is $[browse \rightarrow carting \rightarrow purchase]$. A *behavior habit* reflects the entire interaction process of a user with an item. Due to the fact that users do not know their real needs clearly in the process of interaction, there is often discontinuity in users' interaction with an item, which may lead to a long distance between different behaviors for the same item. These habits may not be consciously aware by users, which are instrumental in assisting in predicting recommendations.

B. Data Analysis

In order to explore the actual behavioral dependencies that exist in the real world, understand how to learn session behavioral dependencies, and explore the rationality of current MBSBR tasks, we conduct data analysis on two datasets in the real world, *Appliances* and *Computers*. These two datasets are detailed in section V-A1. Specifically, we select a session from *Appliances* and count the number of behavior habits in that session. At the same time, we conduct a detailed statistical analysis of the item behavior sequences in each dataset, including all behavior habits in the dataset. In summary, we conduct data analysis from both individual and holistic perspectives, aiming to answer the following two questions:

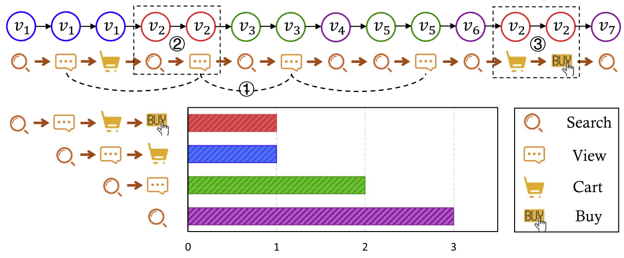

Fig. 1. Counts on all behavioral habits in session #37 in *Appliances* dataset.

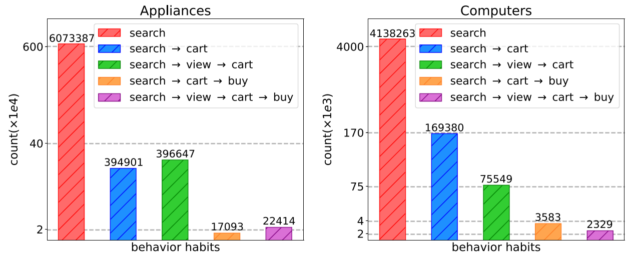

Fig. 2. Five Typical Behavior Habits on Two Datasets.

- *RQ1:* Can user behavior habits effectively reflect their intentions? Is learning user-level behavioral dependencies more reasonable than previous methods?
- *RQ2:* Currently, most MBSBR tasks set a target behavior and set other types of behavior as auxiliary behaviors, aiming to predict the next item under the target behavior. Is this task setting reasonable?

1) RQ1: In order to explore the effects of different behavioral dependency learning methods in existing works, we conduct in-depth research on the session. Specifically, we randomly select one session #37 with multiple items and behaviors in Fig. 1 from the *Appliances* dataset. First, scene ① represents behavior-level methods, which aggregates item v_1, v_2, v_3, and v_5 with "*view*" behavior. Such methods can only learn dependency relationships under different behaviors independently of each other, thereby lacking the capacity to learn complex cross-behavior dependency patterns. Second, scenarios ② and ③ show the learning objects of item-level methods. For item v_2, this method can learn the behavior dependency of [$search \rightarrow view$] in ②, and the behavior dependency of [$cart \rightarrow buy$] in ③. However, these methods either only learn binary behavior patterns, or only aggregate different behaviors of items in continuous timestamps, without mining the behavior dependency of [$view \rightarrow cart$], consequently failing to capture the complete behavior habits of users towards items. Additionally, by examining session #37's behavior habits, we deduce the user's shopping pattern: initiating with searches, followed by reviewing detailed information or comments. Upon thorough evaluation, interested items are added to the cart or purchased. Consequently, capturing user-level behavior habits can more accurately identify which items have a greater impact on user intent, achieving enhancement of user intention.

2) RQ2: To explore the rationality of the MBSBR task, we count the number of all behavior habits in two datasets. Due to the large number of behavioral habits in the dataset, we select 5 typical behavior habits for display as shown in Fig. 2. Usually, the "*buy*" behavior is considered as the target behavior. However, according to our statistics, out of nearly ten million behavioral habits, the number of individual "*buy*" behaviors of users for items in both datasets is 0. Moreover, the total number of behavior habits that include "*buy*" accounts for less than 0.2% of the total. This figure shows the two behavior habits with the highest number of "*buy*" behaviors, and it can be seen that there is a difference of hundreds of times compared to the most common [*search*] behavior habit. This phenomenon indicates that in a session if there is a "*buy*" behavior for an item, other behaviors must be taken before that behavior occurs. In this case, if the recommender system still predicts the items under the target behavior, the system will ultimately be more inclined to recommend products that the user has already interacted with, rather than recommending new items to the user. This contradicts the fundamental purpose of a recommendation system. The repetitive recommendation of items that users have already engaged with could potentially diminish the overall satisfaction and engagement of the user. In addition, our results also find that compared to other behaviors, user "*buy*" behavior accounts for very small proportion (only 2% in *Appliances* dataset and 0.3% in *Computers* dataset), such extremely small number of labels is hardly to produce accurate item predictions under the "buy" target behavior. In summary, predicting the recommendation task of items under target behavior is unreasonable on MBSBR.

C. Task Formulation

In this section, we formulate the task of our multi-behavior session-based recommendation.

Task Scenario: Suppose that we have a set of sessions denoted as $\mathcal{S} = \{s_1, s_2, \ldots, s_M\}$, where s_i represents the i-th session and M represents the number of all sessions. For session $s_i = [v_1^{b_{1,1}}, v_1^{b_{1,2}}, v_2^{b_{2,1}}, \cdots, v_n^{b_{n,m}}]$, the element $v_t^{b_{t,m}}$ represents the t-th interacted item in session s_i with behavior $b_{t,m}$, where $b_{t,m}$ represents the m-th behavior performed by item v_t, indicating a single behavior user can perform, such as searching, buying, etc.

Prediction Process: Our recommendation system is dedicated to making the next-item prediction for the session. However, due to the diversity of behavioral patterns, there may be multiple consecutive identical items in a session. For example, suppose a session $s_{eg} = [v_1^{b_{1,1}}, v_1^{b_{1,2}}, v_2^{b_{2,1}}, \ldots, v_n^{b_{n,1}}, v_n^{b_{n,2}}]$. Its last two items are all v_n, with only different behaviors being performed. In this case, the predicted next item is likely to still be v_n, which leads to information leakage. To avoid this situation, we define consecutive identical items in a session as a macro item and reconstruct new sessions containing macro items for the original session set $\mathcal{S}$. For session s_{eg} in the previous example, the reconstructed session is $s_{new} = [v_1, v_2, \ldots, v_n]$, and every macro item v_i has a corresponding behavior sequence $h_i = [b_{i,1}, b_{i,2}, \ldots, b_{i,l}]$, where $b_{i,j}$ means the j-th behavior targeting v_i and l represents the length of the behavior sequence.

TABLE I
NOTATIONS DESCRIPTION WITH CLASSIFICATION

Notations	Description	Part
$\mathcal{S}=\{s_1,s_2,\cdots,s_M\}$	The set of all sessions.	**Task Formulation**
$\mathcal{V}=\{v_1,v_2,\cdots,v_N\}$	The set of all items.	
M	The number of sessions.	
N	The number of items.	
$s_i=[v_1^{b_{1,1}},v_1^{b_{1,2}},v_2^{b_{2,1}},\cdots,v_n^{b_{n,m}}]$	The representation of the i-th session.	**Task Formulation**
$v_t^{b_{t,m}}$	The t-th item in the session with behavior $b_{t,m}$.	
$b_{t,m}$	The m-th behavior performed by item v_t.	**Behavior Habits Learning**
$H=[h_1,h_2,\cdots,h_n]$	The list of behavior sequences corresponding to each macro item.	
$h_i=[b_{i,1},b_{i,2},\cdots,b_{i,l}]$	The behavior sequence corresponding to item v_i.	
$s'=[v_1,v_2,\cdots,v_n]$	The reconstructed session that includes macro items.	**Macro Item Representation**
$\mathbf{z}_{i,j}\in\mathbb{R}^d$	The hidden state of behavior $b_{i,j}$.	
$\mathbf{e}_{i,j}^b\in\mathbb{R}^d$	The embedding representation of behavior $b_{i,j}$.	
$\mathbf{Z}=[\mathbf{z}_1,\mathbf{z}_2,\cdots,\mathbf{z}_n]$	The behavior temporal feature of all macro items.	
$\mathbf{e}_s^I\in\mathbb{R}^d$	The intention representation of the session.	
$\mathbf{e}_s^B\in\mathbb{R}^d$	The session representation fused after integrating behavior habits into items.	
$\mathbf{G}=\{\mathcal{V},\mathcal{E}\}$	The global item transition Graph.	**Global Item Transition Capturing**
$\mathcal{E}=\{(v_i,v_j)\|v_i\epsilon\mathcal{V},v_j\epsilon\mathcal{N}(v_i)\}$	The interactive edge set at global-level.	
$\mathcal{N}(v_i)$	The subsequent neighbors set of item v_i.	
$\mathbf{A}^g\in\mathbb{R}^{N\times N}$	The adjacency matrix of global graph $\mathbf{G}$.	
a_{ij}^g	The element of $\mathbf{A}^g$.	
$\mathbf{E}_g^{(l)}\in\mathbb{R}^{N\times d}$	The item embedding that aggregates neighborhood information of layer l.	
$\mathbf{W}^{(l)}\in\mathbb{R}^{d\times d}$	The parameter matrix in the l-th layer.	
$\mathbf{D}\in\mathbb{R}^{N\times N}$	The degree matrix.	
$\mathbf{e}^g$	The initial item embedding.	
$\mathbf{A}_{in},\mathbf{A}_{out}\in\mathbb{R}^{n\times n}$	The in degree matrix and out degree matrix corresponding to the current session.	**Local Graph Structure**
$\mathbf{P}\in\mathbb{R}^{n\times 2d}$	The representation that aggregates the behavioral temporal features of macro items with the item representation.	
$\mathbf{I}_{in},\mathbf{I}_{out}\in\mathbb{R}^{n\times d}$	The high-order fused representation for the in and out views respectively.	
$\mathbf{W}_{in},\mathbf{W}_{out}\in\mathbb{R}^{2d\times d},\mathbf{b}_{in},\mathbf{b}_{out}\in\mathbb{R}^{n\times d}$	The learnable parameters.	**Intention Learning**
$\mathbf{I}\in\mathbb{R}^{N\times 2d}$	The fusion intention representation that combines two views.	**Gate Unit**
$\mathbf{U}\in\mathbb{R}^{n\times d}$	The hidden state of the update gate.	
$R\in\mathbb{R}^{n\times d}$	The hidden state of the reset gate.	
$\mathbf{E}'\in\mathbb{R}^{n\times d}$	The update gate's representation to be updated.	
$\mathbf{X}\in\mathbb{R}^{n\times d}$	The final macro item intention representation in the session.	
$\mathbf{W}_{zi},\mathbf{W}_{ri},\mathbf{W}_{ni}\in\mathbb{R}^{2d\times d}$	The trainable parameters.	
$\mathbf{W}_{zh},\mathbf{W}_{rh},\mathbf{W}_{nh}\in\mathbb{R}^{d\times d}$	The trainable parameters.	
α_i	The attention weight of macro item v_i in the session.	**Session Intention Representation**
$\mathbf{G}^h=\{\mathcal{V}^h,\mathcal{E}^h\}$	The behavior habit hypergraph.	**Behavior Habits Hypergraph**
Q	The correlation matrix of $\mathbf{G}^h$.	
β	The residual.	
$\mathbf{D}_e,\mathbf{D}_v$	The degree matrices normalized in vertex and edge dimensions.	
γ	The gate that controls the weight for updating information.	**Mutual Information Fusion**
$\mathbf{e}_s\in\mathbb{R}^d$	The final session representation.	
$\hat{y}_i,y_i$	The probability and the ground truth of item v_i being predicted respectively.	**Prediction and Optimization**
$\mathcal{L}$	The loss of the framework.	

Our goal is to utilize the above information to predict the next macro item v_{n+1}. We formally set out the problem as follows:

- *Inputs:* The original session $s=[v_1^{b_{1,1}},v_1^{b_{1,2}},v_2^{b_{2,1}},\ldots,v_n^{b_{n,m}}]$, the reconstructed session $s'=[v_1,v_2,\ldots,v_n]$ and the behavior sequence list $H=[h_1,h_2,\ldots,h_n]$ related to the macro items in s'.
- *Outputs:* The predicted next macro item v_{n+1} for this session from the item set.

The main notations appearing in this article are listed in Table 1.

IV. Methodology

A. Workflow Overview

Our proposed framework of **BHSBR** is illustrated in Fig. 3. We design a globally directed graph to extract semantic relationships in the item space, and utilize GRU to capture the behavioral sequential relationship of each macro item, in order to obtain the intention representation of behavior-aware fused semantics. Meanwhile, for the input session, we construct a hypergraph to capture the behavior habits in the session and update the item representation through a hypergraph neural network to enhance user intention. Finally, we fuse the two types of mutual information to obtain the final session representation.

B. Intention Learning

1) Global Item Transition Capturing: Currently, most multi-behavior recommendation methods only encode the items in the current session, ignoring the global complex high-order item transition relationships, which is beneficial to learning item embedding in the session. We construct a global item interaction graph to aggregate global-level neighborhood information.

Define item set as $\mathcal{V}$ and the number of all items is N. If in the session set $\mathcal{S}$ there exist $v_i,v_j\in\mathcal{V}$ such that $v_i\rightarrow v_j$, which means that the user interacts with v_j after interacting with v_i. We define v_j as the subsequent neighbor of v_i, and thus obtain the set of all items' subsequent neighbors $\mathcal{N}$. So the global graph we construct can be represented as $\mathbf{G}=\{\mathcal{V},\mathcal{E}\}$, where $\mathcal{E}=\{(v_i,v_j)|v_i\in\mathcal{V},v_j\in\mathcal{N}(v_i)\}$ represents the interactive edge set at global-level and $\mathcal{N}(v_i)$ is the subsequent neighbors set of item v_i. In this construction, the nodes represent items, and edges represent co-occurrences of items across all

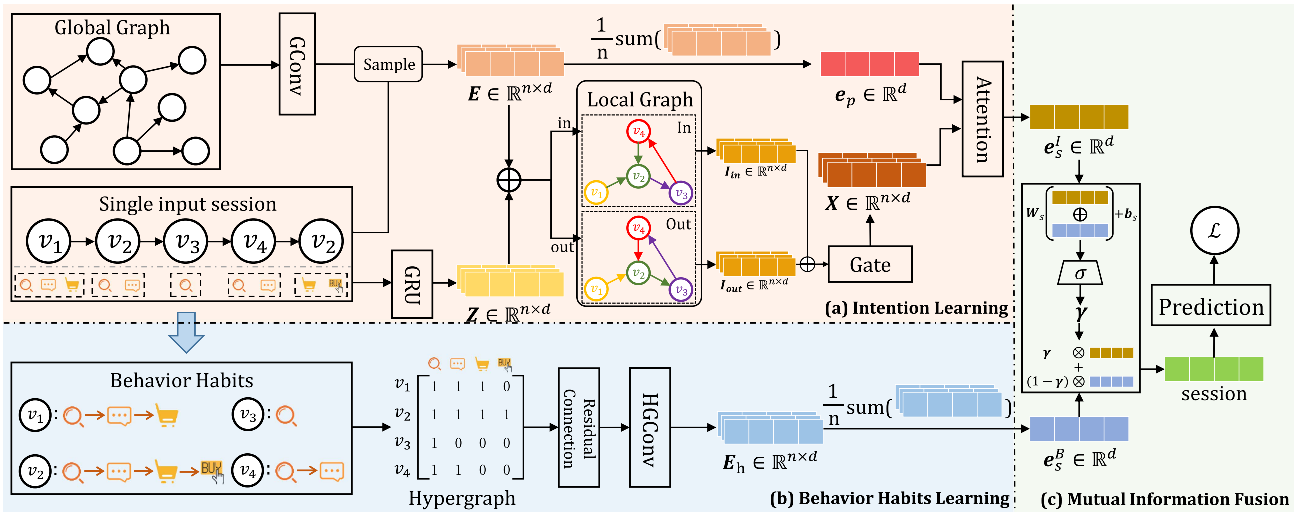

Fig. 3. The overview of the proposed BHSBR.

sessions. According to $\mathbf{G}$, we generate a normalized adjacency matrix $\mathbf{A}^g \in \mathbb{R}^{N\times N}$ as follows:

$$a_{ij}^g = \frac{g_{ij}}{\sum_{k=1}^{N} g_{ik}}, \tag{1}$$

where a_{ij}^g and g_{ij} respectively represent elements from $\mathbf{A}^g$ and $\mathbf{G}$. Whereupon we design a simplified graph convolution layer to globally encode the items:

$$\mathbf{E}_g^{(l+1)} = \mathbf{D}^{-1}\mathbf{A}^g\mathbf{E}_g^{(l)}\mathbf{W}^{(l)}, \tag{2}$$

where $\mathbf{E}_g^{(l)} \in \mathbb{R}^{N\times d}$ and $\mathbf{W}^{(l)} \in \mathbb{R}^{d\times d}$ means the item embedding and parameter matrix in the l-th layer respectively, and d represents the embedding dimension size as well as $\mathbf{D} \in \mathbb{R}^{N\times N}$ is the degree matrix. After L_g-layer propagation, we obtained an item representation $\mathbf{E}_g^{(L_g)}$ that integrates global high-order neighborhood information.

2) Macro Item Representation Learning: According to section III-C, for original session $s = [v_1^{b_{1,1}}, v_1^{b_{1,2}}, v_2^{b_{2,1}}, \ldots, v_n^{b_{n,m}}]$, we can obtain the reconstructed session $s' = [v_1, v_2, \ldots, v_n]$ and behavior sequence list $H = [h_1, h_2, \ldots, h_n]$, where $h_i = [b_{i,1}, b_{i,2}, \ldots, b_{i,l}]$ represents the behavior sequence of macro item v_i. After obtaining the item embedding representation that aggregates global neighborhood information, we design a macro item representation generator to capture the personalized features of each macro item in the session.

In order to extract the behavioral sequence characteristics of macro item v_i, we send its corresponding behavior sequence h_i to GRU:

$$\mathbf{z}_{i,j} = GRU(\mathbf{e}_{i,j}^b, \mathbf{z}_{i,j-1}, \Phi_{GRU}), \tag{3}$$

where $\mathbf{z}_{i,j} \in \mathbb{R}^d$ and $\mathbf{e}_{i,j}^b \in \mathbb{R}^d$ indicate the hidden state and the embedding representation of the behavior $b_{i,j}$ respectively, and Φ_{GRU} is all parameters of GRU. After processing by GRU, the hidden state $\mathbf{z}_{i,l}$ of the last behavior in h_i contains the aggregated information throughout the entire behavior sequence. As a consequence, we consider it as the behavioral feature $\mathbf{z}_i$ of the macro item v_i and we can get the behavior temporal feature of all macro items $\mathbf{Z} \in \mathbb{R}^{n\times d}$ in the session:

$$\mathbf{Z} = [\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_n]. \tag{4}$$

Although sessions are anonymous, learning their items and behaviors from both structural and temporal perspectives can still capture the unique preferences of each session. Unlike the global graph constructed in Section IV-B1, for each session, we construct a local graph to achieve personalized recommendation. In every session, we treat the items as nodes in the local graph and establish directed edges based on the temporal sequence of user interactions. Specifically, an edge is created from item v_i to item v_j if item v_i is interacted with before item v_j in the session. For each session's corresponding local graph, we construct adjacency matrices $\mathbf{A}_{in}, \mathbf{A}_{out} \in \mathbb{R}^{n\times n}$ that separately only include in degree and out degree for the macro item sequence in chronological order, where n is the length of macro item session. These adjacency matrices represent the directed edges between items based on their temporal order in the session. $\mathbf{A}_{in}$ captures the incoming edges, and $\mathbf{A}_{out}$ captures the outgoing edges in the session. Meanwhile, we aggregate item embeddings with behavioral features and propagate information through these two adjacency matrices to capture the preference features of the session:

$$\mathbf{P} = \mathbf{E} \oplus \mathbf{Z}, \tag{5}$$

where $\mathbf{P} \in \mathbb{R}^{n\times 2\,d}$ is the representation that aggregates the behavioral temporal features of macro items with the item representation and $\mathbf{E}$ is the embedding of all macro items in the session sampled from $\mathbf{E}_g^{(L_g)}$. $\oplus$ is the concatenation operation. Then, we propagate the aggregated information into two adjacency matrices, $\mathbf{A}_{in}$ and $\mathbf{A}_{out}$, respectively, to obtain higher-order fused representations for the two views:

$$\mathbf{I}_{in} = \mathbf{A}_{in}\mathbf{W}_{in}\mathbf{P} + \mathbf{b}_{in},$$
$$\mathbf{I}_{out} = \mathbf{A}_{out}\mathbf{W}_{out}\mathbf{P} + \mathbf{b}_{out}, \tag{6}$$

where $\mathbf{W}_{in}, \mathbf{W}_{out} \in \mathbb{R}^{2\,d\times d}$ and $\mathbf{b}_{in}, \mathbf{b}_{out} \in \mathbb{R}^{n\times d}$ are learnable parameters. After local neighborhood propagation, we extract the intention representation of items hidden by users in their behavior. However, the extracted information may contain noise. To improve model performance and generalization ability, we introduce a gating mechanism that aggregates the original item representations through the selection of information by the reset gate and update gate with the learned intent representations. The formula is as follows:

$$\mathbf{I} = \mathbf{I}_{in} \oplus \mathbf{I}_{out},$$
$$\mathbf{U} = \sigma(\mathbf{W}_{zi}\mathbf{I} + \mathbf{W}_{zh}\mathbf{E}),$$
$$\mathbf{R} = \sigma(\mathbf{W}_{ri}\mathbf{I} + \mathbf{W}_{rh}\mathbf{E}),$$
$$\mathbf{E}' = \tanh(\mathbf{W}_{ni}\mathbf{I} + \mathbf{W}_{nh}(\mathbf{R} \otimes \mathbf{E})),$$
$$\mathbf{X} = (\mathbf{1} - \mathbf{U}) \otimes \mathbf{E} + \mathbf{U} \otimes \mathbf{E}', \tag{7}$$

where $\mathbf{I} \in \mathbb{R}^{n\times 2\,d}$ represents the fusion intention representation that combines two views. $\mathbf{W}_{zi}, \mathbf{W}_{ri}, \mathbf{W}_{ni} \in \mathbb{R}^{2\,d\times d}$ and $\mathbf{W}_{zh}, \mathbf{W}_{rh}, \mathbf{W}_{nh} \in \mathbb{R}^{d\times d}$ are trainable parameters. $\sigma(\cdot)$ means *sigmoid* function and $\otimes$ denotes the element level multiplication operator. $\mathbf{U}$ denotes the update gate, which decides how much information from the original item embeddings to pass through. $\mathbf{R}$ represents the reset gate, which determines how much information from the original item embeddings needs to be forgotten. By using the gating mechanism to filter and update item embeddings and behavioral intention information, we have obtained the final macro item intention representation $\mathbf{X} \in \mathbb{R}^{n\times d}$ in the session.

3) Session Intention Representation Generation: In the previous section, we obtain the behavior-aware semantic representation for each macro item in the session, and the intention of the entire session can be learned based on them. In general, the user is always driven by one type of intention in a single session. However, the intentions of items within a session may be diverse, so there is inevitably noise. In order to capture the intention representation of the session more accurately, we utilize attention gating mechanism to determine the weight of each item in the session.

First, we construct the intention prototype of the session. To aggregate all item information in the session, we adopt an average pooling operation:

$$\mathbf{e}_p = \frac{1}{n}\sum_{i=1}^{n}\mathbf{e}_i, \tag{8}$$

where $\mathbf{e}_p \in \mathbb{R}^d$ is the intention prototype of the session. Then we design a gated attention layer to transfer the session intent prototype into the item representation as follows:

$$\alpha_i = \frac{(\mathbf{W}_q\mathbf{x}_i)^T\mathbf{W}_k\mathbf{e}_p}{\sqrt{d}},$$
$$\hat{\mathbf{x}}_i = (\mathbf{1} - \alpha_i)\mathbf{x}_i + \alpha_i\mathbf{e}_p, \tag{9}$$

where α_i is the attention weight of macro item v_i in the session and $\mathbf{W}_q, \mathbf{W}_k \in \mathbb{R}^{d\times d}$ are learnable parameters. In this way, we obtain the final fused item representation. Due to the attention method and GRU's information transmission mechanism, the last item in the session contains information from all previous items. Therefore, we use the representation of the last item v_n as the overall intention representation of the session:

$$\mathbf{e}_s^I = \hat{\mathbf{x}}_n, \tag{10}$$

where $\mathbf{e}_s^I \in \mathbb{R}^d$ denotes the intention representation of the session, which integrates the sequential information of items in the session, global-level neighborhood relationships, and fine-grained behavioral operations within the items.

C. Behavior Habits Learning

The purpose of this module is to capture the complete behavioral process of each item in a session to model user-level behavioral habits and guide item representation learning based on this information to achieve enhancement of user intention. In Section IV-B, although we use GRU to learn the behavioral intentions corresponding to macro items, the information learned by this module is still limited to the sequential patterns in the session and cannot capture the user's whole behavior habits. For example, a user first clicks, browses, and adds items v_1 to the shopping cart, and then discovers items v_2 and v_3 that he may be interested in and browses them separately, and then he makes a purchase of v_1. In that way, the macro item sequence corresponding to this user is $[v_1, v_2, v_3, v_1]$. Our intention learning module can only learn v_1 at two time positions separately, and cannot utilize a complete behavior habit for v_1, thus unable to capture the user's potential behavioral preferences.

Since hypergraphs allow a hyperedge to connect more than two vertices, they can represent more complex relationships compared to general graph structures. In the context of behavior habits, a single item may correspond to multiple types of behaviors, making the hypergraph structure highly suitable for capturing behavior habits. Therefore, we construct a behavior habits hypergraph for each session to capture all item behavior habits within it. For the constructed hypergraph $\mathbf{G}^h = \{\mathcal{V}^h, \mathcal{E}^h\}$, $\mathcal{V}^h$ represents the set of behaviors in the behavior space and $\mathcal{E}^h$ represents the macro item set in the session. For a given session, if an item v_i is associated with a sequence of behaviors (e.g., [view, cart, buy]), then the corresponding hyperedge for v_i connects to all behavior nodes representing the behaviors that the user performed on v_i during that session. The elements in the correlation matrix $\mathcal{Q}$ corresponding to the hypergraph are defined as follows:

$$q_{ij} = |b_j^i| + \beta, \tag{11}$$

where $|b_j^i|$ means the number of behavior b_j performed on item v_i in the session and β is residual to prevent gradient explosion. Different users generate behavior habits hypergraphs that are also not identical, hence personal user behavior preferences can be extracted from them.

After obtaining the behavior habits hypergraph of the session, we utilize the hypergraph graph convolution method for message passing to integrate the behavior habits into the item representation. Inspired by [49], we design our hypergraph learning network as follows:

$$\mathbf{E}_h = \mathbf{D}_v^{-1} \mathcal{Q} \mathbf{D}_e^{-1} \mathcal{Q}^T \hat{\mathbf{E}}, \tag{12}$$

where $\hat{\mathbf{E}} \in \mathbb{R}^{n \times d}$ represents the initial item embedding in the session. $\mathbf{D}_e$ and $\mathbf{D}_v$ are degree matrices normalized in vertex and edge dimensions, respectively. Through hypergraph learning network, we achieve information propagation from nodes to hyperedges, and from hyperedges back to nodes, thereby refining the representation of items. After processing with hypergraph convolution, we obtain the item embedding representation $\mathbf{E}_h$ that integrates behavior habits information. Ultimately, we adopt the average pooling method to calculate the session representation:

$$\mathbf{e}_s^B = \frac{1}{n} \sum_{i=1}^{n} \mathbf{e}_i^h, \tag{13}$$

where $\mathbf{e}_s^B \in \mathbb{R}^d$ means the final session representation fused after integrating behavior habits into items.

The session-level behavior habits hypergraph is constructed as an offline preprocessing step before training, which avoids additional computation during the training phase. The hypergraph construction complexity is linear with respect to the number of sessions and their average length. The main computational overhead during training arises from message passing within the hypergraph convolution layers, which scales with the number of items in a session and the model's embedding dimension. These computations are parallelizable and well-suited for mini-batch processing. For larger datasets or longer sessions, we plan to explore clustering-based methods to further improve scalability by grouping similar behavioral patterns and reducing the number of hyperedges processed during training. Since the session-level hypergraph and the global item graph are constructed in advance, their costs can be amortized over many training epochs, improving overall efficiency.

D. Mutual Information Fusion

In our framework, we encode the session information from two perspectives, the session intention and the behavior habits. To generate the final session representation, we design a gated attention layer to achieve this mutual information fusion with trade-offs. The specific implementation process is as follows:

$$\begin{aligned} \boldsymbol{\gamma} &= \sigma(\mathbf{W}_s(\mathbf{e}_s^I \oplus \mathbf{e}_s^B) + \mathbf{b}_s), \\ \mathbf{e}_s &= \boldsymbol{\gamma} \otimes \mathbf{e}_s^I + (1 - \boldsymbol{\gamma}) \otimes \mathbf{e}_s^B, \end{aligned} \tag{14}$$

where $\mathbf{W}_s \in \mathbb{R}^{2\, d \times d}$ and $\mathbf{b}_s \in \mathbb{R}^d$ are both learnable parameters and $\mathbf{e}_s$ is the final session representation. In this way, we successfully integrate the mutual information between the user's intention and behavior habits in the session, and we can utilize this feature vector for recommendation prediction.

E. Model Optimization and Prediction

To make the prediction process more stable, we perform L2 regularization on the session representation and item embedding separately:

$$\begin{aligned} \hat{\mathbf{e}}_s &= L_2Norm(\mathbf{e}_s), \\ \hat{\mathbf{E}}_g &= L_2Norm(\mathbf{E}_g), \end{aligned} \tag{15}$$

where $\mathbf{E}_g \in \mathbb{R}^{N \times d}$ is the initial global item embedding. Subsequently, we calculate the final prediction results and adopt cross-entropy loss as the optimizer for model training:

$$\begin{aligned} \hat{y}_i &= softmax(\hat{\mathbf{e}}_s^T \hat{\mathbf{E}}_g), \\ \mathcal{L} &= -\sum_{i=1}^{N} y_i log(\hat{y}_i), \end{aligned} \tag{16}$$

where $\hat{y}_i, y_i$ means the probability and the ground truth of item v_i being predicted, respectively.

V. Experiments

In this section, we conduct extensive experiments and analyze the experimental results in detail to answer the following questions:

- *RQ1:* Can the proposed multi-behavior session-based recommendation model BHSBR outperform other baselines?
- *RQ2:* Does each module of the BHSBR model have a positive impact on performance?
- *RQ3:* What is the effect of key parameters to the BHSBR model performance?
- *RQ4:* Does the BHSBR model capture users' intentions accurately?

We first introduce the experimental settings, including the datasets, baselines, evaluation metrics, and parameter settings. Subsequently, we analyze each experimental result to answer the questions asked above.The source code is publicly available at https://github.com/Qin-lab-code/BHSBR.

A. Experimental Settings

1) Datasets: We conduct extensive experiments on three real-world datasets: *Appliances*, *Computers*, and *Trivago*, which contain rich session behavior interaction data. Among them, *Appliances* and *Computers* datasets are from China's huge e-commerce platform JD.com,[1] while the *Trivago* dataset is provided by the global hotel search platform trivago in the RecSys Challenge 2019:[2]

- *Appliances:* This dataset contains session interaction records for "Appliances". It has 10 different types of behaviors, such as "add to cart", "purchase", and so on.
- *Computers:* This dataset, along with *Appliances*, belongs to JD's dataset and contains historical interaction data about "Computers". Similarly, it also includes 10 different types of behaviors.
- *Trivago:* This dataset contains interaction data about hotels, with a total of six types of operation objects being the behavior types of items. During the experiment, we only

[1] https://tinyurl.com/ybo8z4yz
[2] http://www.recsyschallenge.com/2019/

TABLE II
DATASETS

Dataset Name	Appliances	Computers	Trivago
# behavior types	10	10	6
# items	77,258	96,290	183,561
# train sessions	1,294,157	416,664	260,877
# validation sessions	184,399	59,518	37,027
# test sessions	358,795	118,434	74,770

use the training set of the original dataset to repartition and conduct the experiment.

We chose these three datasets to ensure coverage of diverse recommendation scenarios. Appliances and Computers are from e-commerce with rich multi-behavior interactions such as view, cart, and purchase, while Trivago represents hotel booking with distinct user behaviors. This diversity enables evaluation of our framework in different domains. The datasets also vary in size and complexity, testing our model's scalability and robustness. Finally, these datasets are widely used benchmarks in session-based recommendation research, supporting fair and meaningful comparison with prior work. During the dataset processing phase, based on the number of sessions, we divide the training set, validation set, and testing set in the proportions of 70%, 10%, and 20% respectively. For the *Trivago* dataset, we filter out items with fewer than 5 occurrences and sessions with only one interactive item. Unlike the processing of the *Trivago* dataset, for the other two datasets, we filter out items that appeared less than 50 times. The relevant information of these datasets is shown in Table II.

2) Baselines: We compare our **BHSBR** with various recommendation model baselines to verify the effectiveness and superiority of our model.

General Session-based Recommendation Methods (SBRs):

- *GRU4Rec [6]:* It introduced GRU for the first time in session recommendation to learn user preferences.
- *SASRec [50]:* It leveraged attention mechanism to model user historical behavior and extracted session preferences.
- *STAMP [20]:* It designed a long-term and short-term memory network to capture users' long-term memory and recent interests separately.
- *SRGNN [22]:* It built a item level interaction directed graph for each session to capture the complex transformation relationships of items.
- *SGNN-HN [23]:* It utilized star graph neural networks to capture complex transformation relationships in items.
- *COTREC [25]:* It designed a self-supervised contrastive learning paradigm to achieve item-level information enhancement.
- *MCLRec [26]:* It utilized meta-learning to guide parameter updates of model enhancers, in order to improve contrast quality without increasing input data volume.

Hypergraph-enhanced SBRs:

- *DHCN [27]:* It modeled each session as a hyperedge of a hypergraph, with items as vertices, to capture high-order interactions between items.
- *HIDE [28]:* It constructed a hypergraph for each session and decoupled the click intention of the item from both macro and micro perspectives.

Multi-behavior SBRs:

- *MBHT [13]:* It designed a multi-scale transformer to capture sequential patterns and combined it with hypergraph networks to learn behavioral dependencies.
- *EMBSR [14]:* It adopted an extended self-attention network to utilize pairwise relational patterns of micro behavior to capture fine-grained user preferences.
- *PBAT [46]:* It captured user features through dynamic representation encoding and personalized pattern learning, and designed a behavior-aware collaborative extractor to extract the semantics of behavior transitions.
- *MiaSRec [47]:* It derives multiple session representations centered on each item to capture diverse user intents and enhance repeated item information.
- *MGCOT [30]:* It integrates the current session graph, similar session graphs, and a global item graph with multi-head attention and contrastive learning to improve session representation.

3) Evaluation Metrics: To more comprehensively evaluate the performance of our model, we employ several commonly used evaluation metrics in the session-based recommendation domain to assess all models. In order to make the experiment more convincing, we select three evaluation metrics: top-K Recall (Recall@K), top-K Mean Reciprocal Rank (MRR@K), and top-K Normalized Discounted Cumulative Gain (NDCG@K). We set the value of K to $[10, 20]$. Recall@K measures whether the ground-truth item is within the top-K recommended items. MRR@K evaluates the reciprocal of the rank of the ground-truth item, and NDCG@K takes the ranking position into account with position-based discounts. These metrics provide a comprehensive view of ranking quality. Note that all baseline models, including EMBSR, were evaluated under this unified evaluation protocol. Their original model structures were preserved, and only the metric computation procedure was aligned for consistency.

4) Parameter Settings: For all baseline methods, we follow the relevant configurations described in their corresponding papers, preprocess the data, and set parameters to ensure they achieve optimal results. In order to ensure the fairness and effectiveness of the experiment, we unify the prediction goal as the next item of the session without considering whether it is the target behavior. The parameter values used in our framework are set as follows. We configure our model with a batch size of 128 and set the training epochs to 30. For the *Trivago* dataset, we establish a maximum sequence length of 50, while the other two datasets are 20. Additionally, We set the embedding dimension value to 100. We adopt the Adam optimizer and apply dropout with rates of 0.1 (*Appliances, Computers*) and 0.5 (*Trivago*). The learning rate and the layer number of the global semantic capturing process are further analyzed in section V-D.

B. Overall Comparison (RQ1)

We conduct performance evaluations on all baselines and our **BHSBR** in the task for predicting the next item. The evaluation results of the three datasets are shown in Table III. From the research results, we have the following summary:

TABLE III
PERFORMANCE COMPARISON OF DIFFERENT METHODS ON THREE DATASETS

Metric	General Session-based Recommendation							Hypergraph-based		Multi-behavior Methods					Ours	Impr.
	GRU4Rec	SASRec	STAMP	SRGNN	SGNN-HN	COTREC	MCLRec	DHCN	HIDE	MBHT	EMBSR	PBAT	MiasRec	MGCOT	BHSBR	
Appliances																
Recall@10	25.25	27.56	25.58	31.09	31.53	30.25	24.93	31.08	29.20	0.46	34.40	27.28	33.61	34.28	**36.82**	7.03%
MRR@10	13.15	14.61	14.29	16.71	16.75	10.34	11.46	12.45	11.10	0.15	17.21	10.53	16.68	17.15	**18.53**	7.67%
NDCG@10	16.01	17.67	16.97	20.11	20.24	15.00	14.63	16.82	15.37	0.23	21.26	14.42	19.49	20.69	**22.82**	7.34%
Recall@20	31.48	33.83	30.53	38.19	38.70	40.07	32.47	40.77	37.95	0.68	42.75	38.25	40.95	41.71	**45.90**	7.37%
MRR@20	13.59	15.03	14.63	17.20	17.24	11.05	11.98	13.12	11.71	0.16	17.79	11.29	17.01	17.51	**19.13**	7.53%
NDCG@20	17.58	19.22	18.22	21.90	22.05	17.58	16.52	19.26	17.58	0.28	23.38	17.19	21.29	22.56	**25.06**	7.19%
Computers																
Recall@10	5.42	7.82	7.59	11.86	12.53	16.57	13.11	17.12	10.27	2.24	15.05	3.76	13.61	16.54	**18.31**	6.95%
MRR@10	2.49	4.25	4.72	5.43	5.79	5.90	4.88	6.16	3.79	0.84	5.68	1.13	5.92	6.44	**7.69**	19.41%
NDCG@10	3.17	5.09	5.40	6.93	7.37	8.64	6.82	8.76	5.30	1.16	7.83	1.73	8.03	8.72	**10.18**	16.21%
Recall@20	7.74	9.76	8.85	16.20	16.76	23.79	17.91	23.85	14.68	3.61	20.64	6.55	17.78	21.33	**25.06**	5.07%
MRR@20	2.63	4.38	4.80	5.73	6.08	6.39	5.21	6.68	4.10	0.93	6.06	1.32	6.57	7.06	**8.15**	15.44%
NDCG@20	3.70	5.56	5.72	8.03	8.43	10.45	8.03	10.67	6.42	1.51	9.22	2.43	9.11	9.92	**11.86**	11.15%
Trivago																
Recall@10	15.33	17.53	17.65	11.90	9.96	21.26	20.87	23.56	21.93	12.50	22.78	6.51	22.59	23.72	**25.38**	7.00%
MRR@10	7.52	8.38	9.26	5.60	4.92	5.75	7.75	8.97	7.78	5.38	9.93	1.42	7.38	9.57	**11.04**	11.17%
NDCG@10	9.34	10.54	11.26	7.08	6.10	9.35	10.82	12.50	11.08	7.05	12.93	3.14	9.12	12.39	**14.15**	9.43%
Recall@20	19.90	22.53	21.21	15.77	12.94	31.82	28.55	33.56	30.86	17.27	31.10	13.99	28.36	30.05	**34.27**	2.12%
MRR@20	7.81	8.72	9.51	5.87	5.13	6.47	8.28	9.67	8.40	5.71	10.51	1.88	7.62	9.73	**11.56**	9.99%
NDCG@20	10.46	11.81	12.16	8.06	6.86	12.02	12.76	15.05	13.34	8.25	15.04	5.00	10.15	13.42	**16.39**	8.90%

- Our proposed **BHSBR** achieves the best performance on three datasets compared to all baseline methods. On average, it improved by 8.21% on *Trivago*, 14.37% on *Computers*, and 7.36% on *Appliances*, verifying the superiority of our model.
- Among three datasets, the improvement on the *Computers* dataset is more significant. A possible reason is that a significant number of behavior habits in the *Computers* dataset are "search", which means that there may be a multitude of items in each session that have only "search" interactions. Consequently, this results in a substantial amount of noise. Compared to other methods, our proposed model exhibits greater robustness and is capable of effectively mitigating the impact of noise.
- GNN-based methods generally perform better than RNN-based methods, indicating that GNN plays a significant role in modeling session structures. Among them, DHCN performs outstandingly on three datasets, proving the superiority of hypergraphs in capturing high-order information.

C. Ablation Study (RQ2)

In order to verify the rationality and effectiveness of each module design of our **BHSBR**, we conduct sufficient ablation experiments. Specifically, we set the following variants:

- **BHSBR_NH** removes the hypergraph-based behavior habit learning layer and only encodes the session representation through the behavior-aware intention learning layer.
- **BHSBR_NI** removes the entire intention learning module and only retains the behavior habit learning layer, capturing only all behavior habits in the session to encode the final representation.
- **BHSBR_NG** removes the global semantic capture process in the intent learning module, only aggregating behavioral information under macro items without utilizing global semantic information for intent enhancement.

The results of the ablation experiment are shown in Table IV. From the experimental results, we can draw the following conclusions:

TABLE IV
ABLATION STUDY ON THREE DATASETS

Dataset	Metric	BHSBR_NH	BHSBR_NI	BHSBR_NG	BHSBR
Appliances	Recall@10	34.74	35.10	36.73	**36.82**
	MRR@10	16.98	17.53	18.19	**18.53**
	NDCG@10	21.08	21.67	22.56	**22.82**
	Recall@20	43.52	44.24	45.46	**45.90**
	MRR@20	17.57	18.16	18.81	**19.13**
	NDCG@20	23.24	23.98	24.83	**25.06**
Computers	Recall@10	16.76	15.20	17.12	**18.31**
	MRR@10	6.45	6.95	7.23	**7.69**
	NDCG@10	8.82	8.88	9.42	**10.18**
	Recall@20	23.16	20.40	23.01	**25.07**
	MRR@20	6.87	7.30	7.62	**8.15**
	NDCG@20	10.44	10.18	10.84	**11.86**
Trivago	Recall@10	23.29	23.10	23.98	**25.38**
	MRR@10	9.34	10.63	10.23	**11.04**
	NDCG@10	12.59	13.53	13.38	**14.07**
	Recall@20	32.17	30.56	32.38	**34.27**
	MRR@20	9.95	11.13	10.81	**11.56**
	NDCG@20	14.83	15.39	15.52	**16.29**

- Among the three datasets, **BHSBR_NH** generally performs the worst, especially in the MRR and NDCG metrics, indicating that capturing user behavior habits plays a profoundly significant role in enhancing user intention and reducing noise issues in sessions.
- The performance degradation of **BHSBR_NI** is also significant, demonstrating that there is a deep correlation between user intent and the temporal relationship between items in session recommendation. Consequently, integrating behavioral information and semantic information into the session temporal relationship can better capture the user's true intention.
- The performance of **BHSBR_NG** has also decreased on three datasets. Notably, the decrease is particularly significant on the *Trivago* dataset. One possible reason is that in the *Trivago* dataset, most of the predicted items in each session have not appeared in previous interactions. Thus, after removing the global semantic capture layer, the model tends to recommend items that have already been interacted with by users, while lacking the ability to discover new items.

D. The Impact of Hyperparameters (RQ3)

To investigate the impact of different hyperparameter values on model performance, we conduct sufficient hyperparameter

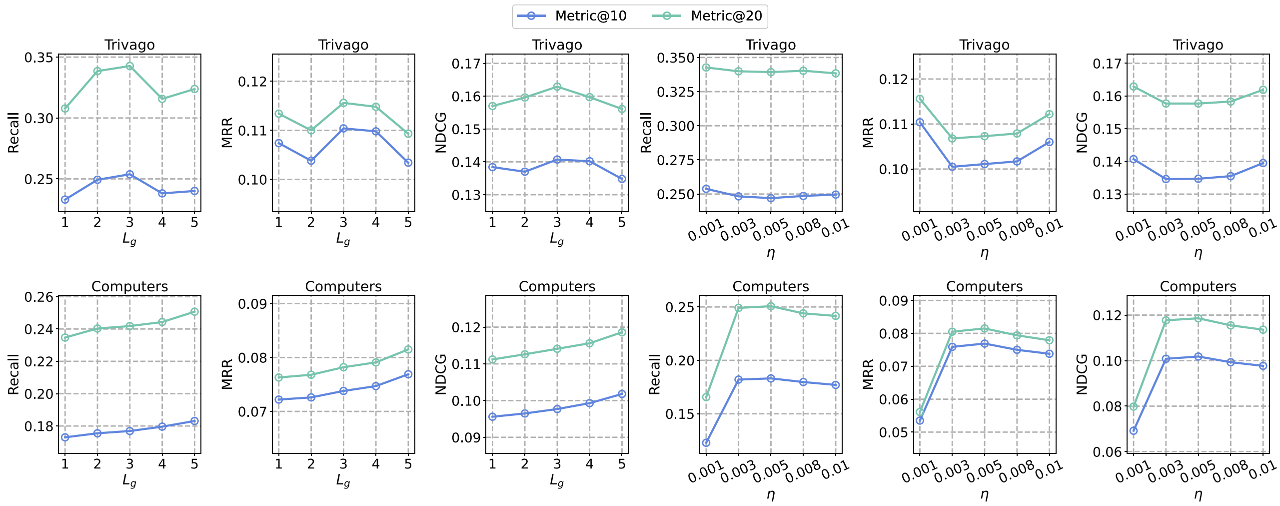

Fig. 4. The impact of hyperparameters L_g and η on *Trivago* and *Computers* datasets.

experiments on the layer number of global semantic capturing process L_g and the learning rate η of Adam loss optimizer on the *Trivago* and *Computers* datasets. The experimental results are shown in Fig. 4.

- We respectively set the value of L_g to $\{1, 2, 3, 4, 5\}$. From Fig. 4, it can be observed that as L_g increases, each item can aggregate richer global neighborhood information, which better captures user intent during the intention encoding process. Consequently, the model's performance improves with the elevation of L_g. However, in *Trivago* dataset, when L_g exceeds 3, the model performance actually decreases. This is because too many iterations of global graph convolution can cause items to aggregate too much neighborhood information, resulting in over-smoothing issues, and leading to a decrease in model performance. In summary, the optimal L_g value varies for different datasets, and appropriate sizes should be selected for each dataset to achieve optimal model performance.
- The values of η is examined in $\{0.001, 0.003, 0.005, 0.008, 0.01\}$. Generally speaking, the learning rate η of the optimizer controls the step size of model optimization. Setting the learning rate too low can easily slow down the convergence speed of the model and drag down its efficiency. If the learning rate is set too high, however, it may lead to missing the optimal solution. Therefore, it is necessary to make a trade-off and set an appropriate learning rate size to achieve relative optimization. From Fig. 4, we can observe that the optimal η value in the *Trivago* dataset is 0.001, while in the *Computers* dataset it is 0.005. Therefore, the optimal η values corresponding to different datasets also vary.

E. Case Study (RQ4)

To verify that our BHSBR can accurately capture user intentions, we conduct a case study and randomly sample a session 435 from the *Computers* dataset. We analyze its behavior

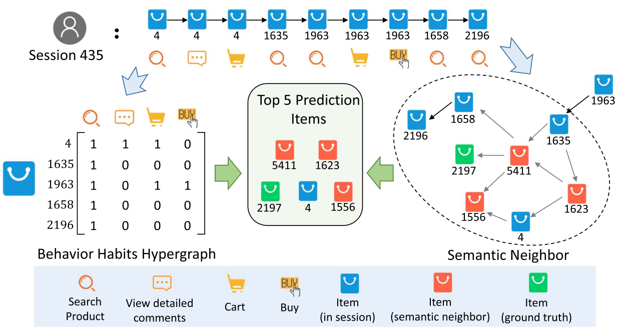

Fig. 5. A case for BHSBR Prediction on *Computers* dataset.

habits hypergraph structure and semantic neighbors in detail and present its prediction results, as shown in Fig. 5. From the graph, we can observe that by capturing the behavior habits of this session, the user has engaged in a lot of behavior towards items 4 and 1963, which means that the user's main intention should be focused on these two items. Meanwhile, our designed intention learning module mines semantic information of items from a global perspective, and it can be observed that in this session, except for item 1963, the other four items can be connected through semantically related items. By combining these two pieces of information, our BHSBR is able to capture that the user's intention should be associated with item 4. We provide the top-5 prediction results of this session, where we successfully predict target item 2197, and all predicted items have strong semantic connections with item 4, which proves that BHSBR can accurately capture user intentions and provide new item recommendations for users.

VI. Conclusion

In this work, we conduct data analysis on multi-behavior scenarios in recommendation models to explore how to use behavioral information to learn behavioral dependencies and better capture user intentions, and explore task settings in multi-behavior scenarios. We propose a new session-based recommendation framework BHSBR. Specifically, we adopt global GNN and GRU to learn semantic relationships and behavioral information of items in a session, and fuse them using gating mechanisms to obtain intent representations. Meanwhile, we design a hypergraph for each session to capture user-level behavior habits and achieve intention enhancement through HGNN. Extensive experiments demonstrate our model's superiority over SOTA approaches, with ablation studies confirming the efficacy of each module. While our approach has demonstrated strong performance across diverse datasets, there remain several directions for future improvement. For instance, we aim to further enhance the model's efficiency for longer sequences and large-scale datasets. Additionally, we plan to explore more adaptive strategies to better capture complex user behavior patterns and improve continuity in longer sessions. We believe addressing these aspects will further strengthen the generalizability and practicality of our framework.

REFERENCES

[1] S. Qiao, W. Zhou, J. Wen, H. Zhang, and M. Gao, "Bi-channel multiple sparse graph attention networks for session-based recommendation," in *Proc. 32nd ACM Int. Conf. Inf. Knowl. Manag.*, Birmingham, U.K., 2023, pp. 2075–2084.

[2] Q. Chen, J. Li, Z. Guo, G. Li, and Z. Deng, "Attribute-enhanced dual channel representation learning for session-based recommendation," in *Proc. 32nd ACM Int. Conf. Inf. Knowl. Manag.*, Birmingham, U.K., 2023, pp. 3793–3797.

[3] D. Yu, Q. Li, H. Yin, and G. Xu, "Causality-guided graph learning for session-based recommendation," in *Proc. 32nd ACM Int. Conf. Inf. Knowl. Manag.*, Birmingham, U.K., 2023, pp. 3083–3093.

[4] S. Rendle, C. Freudenthaler, and L. Schmidt-Thieme, "Factorizing personalized Markov chains for next-basket recommendation," in *Proc. 19th Int. Conf. World Wide Web*, 2010, pp. 811–820.

[5] W. Gu, S. Dong, and Z. Zeng, "Increasing recommended effectiveness with Markov chains and purchase intervals," *Neural Comput. Appl.*, vol. 25, no. 5, pp. 1153–1162, Oct. 2014.

[6] B. Hidasi, A. Karatzoglou, L. Baltrunas, and D. Tikk, "Session-based recommendations with recurrent neural networks," Nov. 2016, *arXiv:1511.06939*.

[7] J. Li, P. Ren, Z. Chen, Z. Ren, and J. Ma, "Neural attentive session-based recommendation," in *Proc. 26th ACM Int. Conf. Inf. Knowl. Manag.*, 2017, pp. 1419–1428.

[8] Y. K. Tan, X. Xu, and Y. Liu, "Improved recurrent neural networks for session-based recommendations," in *Proc. 1st Workshop Deep Learn. Recomm. Syst.*, 2016, pp. 17–22.

[9] S. Lai, E. Meng, F. Zhang, C. Li, B. Wang, and A. Sun, "An attribute-driven mirror graph network for session-based recommendation," in *Proc. 45th ACM SIGIR Conf. Res. Dev. Inf. Retrieval*, 2022, pp. 1674–1683.

[10] D. Jin et al., "Dual intent enhanced graph neural network for session-based new item recommendation," in *Proc. ACM Web Conf.*, 2023, pp. 684–693.

[11] J. Su, C. Chen, W. Liu, F. Wu, X. Zheng, and H. Lyu, "Enhancing hierarchy-aware graph networks with deep dual clustering for session-based recommendation," in *Proc. ACM Web Conf.*, 2023, pp. 165–176.

[12] L. Xia, C. Huang, Y. Xu, and J. Pei, "Multi-behavior sequential recommendation with temporal graph transformer," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 5, pp. 6099–6112, Jun., 2023.

[13] Y. Yang, C. Huang, L. Xia, Y. Liang, Y. Yu, and C. Li, "Multi-behavior hypergraph-enhanced transformer for sequential recommendation," in *Proc. 28th ACM SIGKDD Conf. Knowl. Discov. Data Min.*, 2022, pp. 2057–2067.

[14] J. Yuan, W. Ji, D.-F. Zhang, J. Pan, and X. Wang, "Micro-behavior encoding for session-based recommendation," in *Proc. 38th IEEE Int. Conf. Data Eng*, 2022, pp. 2886–2899.

[15] L. Xia, C. Huang, Y. Xu, P. Dai, B. Zhang, and L. Bo, "Multiplex behavioral relation learning for recommendation via memory augmented transformer network," in *Proc. 43rd ACM Int. Conf. Inf. Retrieval*, Jul. 2020, pp. 1–10.

[16] L. Guo, L. Hua, R. Jia, B. Zhao, X. Wang, and B. Cui, "Buying or browsing?: Predicting real-time purchasing intent using attention-based deep network with multiple behavior," in *Proc. 25th ACM Int. Conf. Knowl. Discov. Data Mining*. New York, NY, USA:ACM, 2019, pp. 1984–1992.

[17] A. Li, Z. Cheng, F. Liu, Z. Gao, W. Guan, and Y. Peng, "Disentangled graph neural networks for session-based recommendation," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 8, pp. 7870–7882, Aug., 2023.

[18] W. Wang et al., "Incorporating link prediction into multi-relational item graph modeling for session-based recommendation," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 3, pp. 2683–2696, Mar., 2023.

[19] A. Vaswani et al., "Attention is all you need," in *Proc. 31st Conf. Neural Inf. Process. Syst.*, 2017, pp. 6000–6010.

[20] Q. Liu, Y. Zeng, R. Mokhosi, and H. Zhang, "Stamp: Short-term attention/memory priority model for session-based recommendation," in *Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2018, pp. 1831–1839.

[21] F. Sun et al., "BERT4Rec: Sequential recommendation with bidirectional encoder representations from transformer," in *Proc. 28th ACM Int. Conf. Inf. Knowl. Manag.*, 2019, pp. 1441–1450.

[22] S. Wu, Y. Tang, Y. Zhu, L. Wang, X. Xie, and T. Tan, "Session-based recommendation with graph neural networks," in *Proc. AAAI Conf. Artif. Intell.*, 2019, vol. 33, no. 1, pp. 346–353.

[23] Z. Pan, F. Cai, W. Chen, H. Chen, and M. de Rijke, "Star graph neural networks for session-based recommendation," in *Proc. 29th ACM Int. Conf. Inf. Knowl. Manag.*, 2020, pp. 1195–1204.

[24] Z. Wang, W. Wei, G. Cong, X.-L. Li, X.-L. Mao, and M. Qiu, "Global context enhanced graph neural networks for session-based recommendation," in *Proc. 43rd Int. ACM SIGIR Conf. Res. Dev. Inf. Retrieval*, Jul., 2020, pp. 169–178.

[25] X. Xia, H. Yin, J. Yu, Y. Shao, and L. Cui, "Self-supervised graph co-training for session-based recommendation," in *Proc. 30th ACM Int. Conf. Inf. Knowl. Manag.*, 2021, pp. 2180–2190.

[26] X. Qin et al., "Meta-optimized contrastive learning for sequential recommendation," in *Proc. 46th ACM Int. Conf. Res. Dev. Inf. Retrieval*, 2023, pp. 89–98.

[27] X. Xia, H. Yin, J. Yu, Q. Wang, L. Cui, and X. Zhang, "Self-supervised hypergraph convolutional networks for session-based recommendation," in *Proc. AAAI Conf. Artif. Intell.*, vol. 35, no. 5, 2021, pp. 4503–4511.

[28] Y. Li, C. Gao, H. Luo, D. Jin, and Y. Li, "Enhancing hypergraph neural networks with intent disentanglement for session-based recommendation," in *Proc. 45th ACM Int. Conf. Res. Dev. Inf. Retrieval*, 2022, pp. 1997–2002.

[29] Y. Qin, Y. Zhang, Y. Liu, Y. Wang, and Y. Li, "Multi-behavior session-based recommendation via graph reinforcement learning," in *Proc. 41st Int. Conf. Mach. Learn.*, 2024, pp. 1119-1134.

[30] Z. Yang, Y. Wang, Y. Li, and Y. Zhang, "Multi-graph co-training for capturing user intent in session-based recommendation," in *Proc. 31st Int. Conf. Comput. Linguistics Assoc. Comput. Linguistics*, 2025, pp. 1377–1386.

[31] J. Ding, Z. Tan, G. Lu, and J. Wei, "Hypergraph denoising neural network for session-based recommendation," *Appl. Intell.*, vol. 55, no. 6, pp. 1–18, 2025.

[32] J. Sohafi-Bonab, M. H. Aghdam, and K. Majidzadeh, "DCARS: Deep context-aware recommendation system based on session latent context," *Appl. Soft Comput.*, vol. 143, 2023, Art. no. 110416.

[33] H. Vaghari, M. H. Aghdam, and H. Emami, "HAN: Hierarchical attention network for learning latent context-aware user preferences with attribute awareness," *IEEE Access*, vol. 13, pp. 49030–49049, 2025.

[34] A. P. Singh and G. J. Gordon, "Relational learning via collective matrix factorization," in *Proc. 14th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2008, pp. 650–658.

[35] A. Krohn-Grimberghe, L. Drumond, C. Freudenthaler, and L. Schmidt-Thieme, "Multi-relational matrix factorization using Bayesian personalized ranking for social network data," in *Proc. 5th ACM Int. Conf. Web Search Data Mining*, 2012, pp. 173–182.

[36] C. Chen et al., "Graph heterogeneous multi-relational recommendation," in *Proc. AAAI Conf. Artif. Intell.*, 2021, vol. 35, no. 5, pp. 3958–3966.

[37] B. Jin, C. Gao, X. He, D. Jin, and Y. Li, "Multi-behavior recommendation with graph convolutional networks," in *Proc. 43rd Int. ACM SIGIR Conf. Res. Dev. Inf. Retrieval*, 2020, pp. 659–668.

[38] Z. Zhao, Z. Cheng, L. Hong, and E. H. Chi, "Improving user topic interest profiles by behavior factorization," in *Proc. 24th Int. Conf. World Wide Web*, 2015, pp. 1406–1416.

[39] W. Zhang, J. Mao, Y. Cao, and C. Xu, "Multiplex graph neural networks for multi-behavior recommendation," in *Proc. 29th ACM Int. Conf. Inf. Knowl. Manag.*, 2020, pp. 2313–2316.
[40] L. Xia, Y. Xu, C. Huang, P. Dai, and L. Bo, "Graph meta network for multi-behavior recommendation," in *Proc. 44th Int. ACM SIGIR Conf. Res. Dev. Inf. Retrieval*, 2021, pp. 757–766.
[41] C. Meng et al., "Hierarchical projection enhanced multi-behavior recommendation," in *Proc. 29th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2023, pp. 4649–4660.
[42] W. Wei, C. Huang, L. Xia, Y. Xu, J. Zhao, and D. Yin, "Contrastive meta learning with behavior multiplicity for recommendation," in *Proc. 15th ACM Int. Conf. Web Search Data Mining*, 2022, pp. 83–91.
[43] J. Xu et al., "Multi-behavior self-supervised learning for recommendation," in *Proc. 46th Int. ACM SIGIR Conf. Res. Dev. Inf. Retrieval*, 2023, pp. 1234–1243.
[44] M. Zhou, Z. Ding, J. Tang, and D. Yin, "Micro behaviors: A new perspective in E-commerce recommender systems," in *Proc. 11th ACM Int. Conf. Web Search Data Mining*, 2018, pp. 727–735.
[45] W. Meng, D. Yang, and Y. Xiao, "Incorporating user micro-behaviors and item knowledge into multi-task learning for session-based recommendation," in *Proc. 43rd Int. ACM SIGIR Conf. Res. Dev. Inf. Retrieval*, 2020, pp. 1091–1100.
[46] J. Su, C. Chen, Z. Lin, X. Li, W. Liu, and X. Zheng, "Personalized behavior-aware transformer for multi-behavior sequential recommendation," in *Proc. 31st ACM Int. Conf. Multimedia*, 2023, pp. 6321–6331.
[47] M. Choi, H. Kim, H. Cho, and J. Lee, "Multi-intent-aware session-based recommendation," in *Proc. 47th ACM SIGIR Conf. Res. Dev. Inf. Retrieval*, 2024, pp. 2532–2536.
[48] L. Guo, S. Zhou, H. Tang, X. Zheng, and Y. Luo, "Multi-behavior hypergraph contrastive learning for session-based recommendation," *IEEE Trans. Knowl. Data Eng.*, vol. 37, no. 3, pp. 1325–1338, Mar., 2025.
[49] Y. Feng, H. You, Z. Zhang, R. Ji, and Y. Gao, "Hypergraph neural networks," in *Proc. 33rd AAAI Conf. Artif. Intell.*, 2019, pp. 3558–3565.
[50] W.-C. Kang and J. McAuley, "Self-attentive sequential recommendation," in *Proc. 18th IEEE Int. Conf. Data Min.*, 2018, pp. 197–206.

Haoyao Zhang is currently working toward the degree in artificial intelligence with the Beijing Institute of Technology, Beijing, China. His current research interests include recommendation systems and graph neural networks.

Shixiao Yang received the bachelor's degree from the Beijing Institute of Technology, Beijing, China, where he is currently working toward the master's degree. His research interests include recommender systems and airspace intelligent networking.

Enjun Du is currently working toward the undergraduation degree with the Beijing Institute of Technology. He is a visiting research student with the Hong Kong University of Science and Technology and a research intern with Tencent, specializing in large language models and graph neural networks. He has authored or coauthored multiple first-authored papers at prestigious venues including NeurIPS (spotlight) and EMNLP (oral presentation).

Zhida Qin received the BS degree in electronic and information engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2014, and the PhD degree in electronic engineering from Shanghai Jiao Tong University, Shanghai, China. He is currently an assistant professor with the School of Computer Science and Technology, Beijing Institute of Technology, Beijing, China. His research interests include recommendation systems, graph neural networks, and online learning.

Wenhao Xue received the bachelor's degree and the master's degree in computer science from the Beijing Institute of Technology, Beijing, China, in 2022 and 2025, respectively. His current research interests include recommendation systems and data mining.

Haotian He is currently working toward the doctoral degree in computer science and technology with the Beijing Institute of Technology, Beijing, China. His research interests include recommendation systems and graph neural networks.

Tianyu Huang received the BE degree in computer science from Jilin University, Changchun, China, in 2002, and the PhD degree in computer science from the Beijing Institute of Technology, Beijing, China, in 2007. She joined as the faculty with the Beijing Institute of Technology, in 2007. She was a visiting scholar with the Department of Computer and Information Science, University of Pennsylvania, Philadelphia, PA, USA, from 2012 to 2013. She is currently a professor with the School of Computer Science and Technology, Beijing Institute of Technology. Her research interests include human motion modeling and simulation, crowds simulation, and virtual reality.

John C.S. Lui (Fellow, IEEE) received the PhD degree in computer science from UCLA. He joined the IBM Laboratory and participated in research and development projects on file systems and parallel I/O architectures. He later joined the CSE Department, The Chinese University of Hong Kong (CUHK). He is currently the Choh-Ming Li chair professor with the Department of Computer Science and Engineering, CUHK. His current research interests include quantum internet and distributed quantum algorithms, online learning algorithms and applications, machine learning on network sciences and networking systems, large-scale data analytics, large-scale storage systems, and performance evaluation theory. He was with the IEEE Fellow Review Committees. He is an elected member of the IFIP WG 7.3, fellow of ACM, senior research fellow of the Croucher Foundation, and a fellow of Hong Kong Academy of Engineering Sciences.