Self-supervised enhancement method for multi-behavior session-based recommendation

IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 6 November 2024, accepted 19 November 2024, date of publication 22 November 2024, date of current version 3 December 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3504496

RESEARCH ARTICLE

Self-Supervised Enhancement Method for Multi-Behavior Session-Based Recommendation

ZHEN ZHANG[1], JINGAI ZHANG[2], JINTAO CHEN[2], YUZHAO HUANG[3], AND XIAOYANG HUANG[3]

[1]Department of Computer Science, University of Reading, RG6 6AH Reading, U.K.
[2]Guangxi Key Laboratory of Trusted Software, Guilin University of Electronic Technology, Guilin, Guangxi 541004, China
[3]Guangxi Key Laboratory of Automatic Detecting Technology and Instruments, Guilin University of Electronic Technology, Guilin, Guangxi 541004, China

Corresponding author: Jingai Zhang (zjazq9@foxmail.com)

ABSTRACT Session-Based Recommendation(SBR) aims to capture users' short-term and dynamic preferences through anonymous sessions. Most existing SBR methods neglect the collaborative information between multiple behaviors in a session when modeling user preferences, and they often struggle to capture the complex correlations of contextual information across sessions, which can lead to poor recommendation performance. To address these issues, this paper proposes a Self-Supervised enhancement method for Multi-Behavior Session-based Recommendation(SSMB-SR). SSMB-SR represents the session sequence using a heterogeneous graph to capture intricate behavior interactions and a hypergraph for contextual information integration. Specifically, we designed a heterogeneous enhancement module that deeply understands the intrinsic connections of behaviors and the interdependencies between different behavior types by enhancing the behavioral information of the central node, effectively capturing the complex dynamic interactions between nodes within the session to obtain accurate item embeddings. Concurrently, we propose a self-supervised training method for the module that mitigates location bias and minimizes the impact of noisy behaviors. For cross-session, we combine relevant contextual information through a hypergraph to achieve accurate recommendation results. Experimental results show that our proposed self-supervised enhancement method significantly improves recommendation performance and has a better performance compared to recommendation methods that only consider a single behavior.

INDEX TERMS Session-based recommendation, graph neural networks, heterogeneous graph, hypergraph, self-supervised learning.

I. INTRODUCTION

In the era of big data, recommendation systems have increasingly addressed the challenge of information overload by delivering relevant content to users. Session-based recommendation, a burgeoning field within recommendation systems, has gained prominence for its ability to discern users' fleeting and dynamic preferences through the analysis of session dependencies, both within and across sessions. This method has captured considerable interest across various domains, including e-commerce, music streaming, and social media platforms. Multi-behavioral sessions are a prevalent occurrence, characterized by their complex and nuanced interdependencies. Notably, these interdependencies manifest not only within the same interaction types (e.g., consecutive product clicks) but also across diverse interaction types (such as clicks, browsing, and purchases) [1]. Presently, the majority of session-based recommendation systems fail to delve into the nuances of behavioral data, thereby missing the intricate relationships within sessions and the cross-behavioral interdependencies. This oversight leads to suboptimal recommendation accuracy. Therefore, one of the challenges currently faced by session-based recommendation systems is how to effectively and accurately understand the interdependencies within and among types of behaviors.

To confront this challenge, our research focuses on demonstrating the application of our model on two real-world datasets, namely from the e-commerce and music

The associate editor coordinating the review of this manuscript and approving it for publication was Zhengmao Li.

streaming domains. These examples serve to illustrate how our approach can navigate the current obstacles in user recommendation. We are persuaded that these practical demonstrations will not only clarify the architecture and operations of our model but also underscore its capacity to effect substantial improvements in various real-world scenarios. Through this, we intend to offer a more tangible understanding of how our methodology can be employed to refine the user experience and to advance the precision and personalization of recommendations within these fast-paced industries.

Certainly, within the scope of this study, it is imprudent to train session-based recommendation systems (SBRS) without accounting for the influence of contextual supplementary information inherent in the sessions. Integrating additional contextual data from session interactions can more effectively mitigate the interference caused by noisy user behavior. To enable a more thorough analysis, we initially constructed a heterogeneous graph, as depicted in Figure. 1(b), which elucidates the correlations among various behavior types within the current SBRS framework. This stands in contrast to the single-behavior representation of users and their music preferences shown in Figure. 1(a), where a user's history is limited to classical music, despite their genuine affinity with pop and rock genres. Considering only a single behavior, such as play, risks recommending items akin to classical music, thus neglecting the user's broader musical tastes. Consequently, in such scenarios, reliance on a single behavior can negatively affect the precision of the recommendations.

The fundamental concept of data augmentation revolves around crafting diverse enhancement techniques to bolster the volume or integrity of the available data. In our method, we leverage a heterogeneous graph to delve into and examine the intricate correlations that exist among various user behaviors. Our objective is to enrich the information pertaining to user actions and the data associated with adjacent nodes in the central node, thereby ascertaining the authentic connections between items. Consequently, (Problem 1) if the interdependencies among multiple behavior types are not taken into account, the Session-Based Recommendation System (SBRS) may inadvertently learn misleading associations stemming from individual behavior types, potentially resulting in imprecise recommendations.

A promising method to address this issue involves building a heterogeneous graph that encapsulates various user behaviors, where nodes represent individual items and edges signify the interconnections between different behavioral patterns. Utilizing methods such as Graph Neural Networks (GNNs) to model the graph structure allows for the consideration of complex dependencies among different behavior nodes. Nonetheless, many conventional models focus solely on user behaviors within the ongoing session to build preference models, overlooking the continuity that spans across sessions. This inter-session influence encompasses vital information that is essential for a precise comprehension of user preferences within the current session. For instance, as depicted in Figure. 1(c), the detection of genuine dependency relationships among items necessitates a thorough examination of users' historical multi-behavior data, rather than isolated or random actions. Furthermore, by delving into potential contextual elements within each session, we can gain deeper insights into the motivations behind users' searches for particular music, allowing for more precise and targeted recommendations of similar tracks. Despite this, the intricate dependencies across sessions remain challenging to unravel due to the presence of noisy behaviors. Consequently, (Problem 2) disregarding the complex interplay of session context will markedly compromise the efficacy of recommendation systems.

In this paper, we tackle Problem 1 by introducing a novel framework for SBRS that accounts for the interdependencies among various behavior types within a session to enhance next-item recommendations. Our proposed model is termed the Self-Supervised Enhanced Multi-Behavior Session Recommendation Method. Drawing inspiration from data augmentation techniques [2] and session-based collaborative filtering methods [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], our method initializes the session sequence and integrates a heterogeneous enhancement module. This module is designed to overcome the limitations of graph neural networks in capturing the relationships between session nodes, thereby facilitating a more nuanced understanding of the dependencies both within and across different behavior types. Following this, we utilize neighboring sessions with analogous behaviors to bolster user choices at the focal node. In the training phase, we adopt a self-supervised learning strategy to train the recommendation model, which aims to alleviate position bias and minimize the distraction caused by noisy behaviors. To address Problem 2, we represent the session sequence as a hypergraph, effectively extracting higher-order information from the contextual correlations that span across sessions, thereby achieving more precise recommendations.

- We utilize a behavior-aware graph neural network that leverages a self-attention mechanism to adeptly capture the intrinsic relationships among behaviors within a session, as well as the interdependencies across various behavior types. Through the application of self-supervised learning at both the intra-behavioral and inter-behavioral levels, our objective is to eliminate position bias and consequently diminish the adverse effects of noisy behaviors.
- We represent the session sequence as a hypergraph, harnessing its ability to encapsulate intricate dependencies that extend beyond pairwise interactions. This method enables us to extract precise contextual information across sessions.
- To assess the efficacy of our model, we have conducted comprehensive experiments on two real-world datasets spanning various domains. We compare the model against representative and state-of-the-art session-based

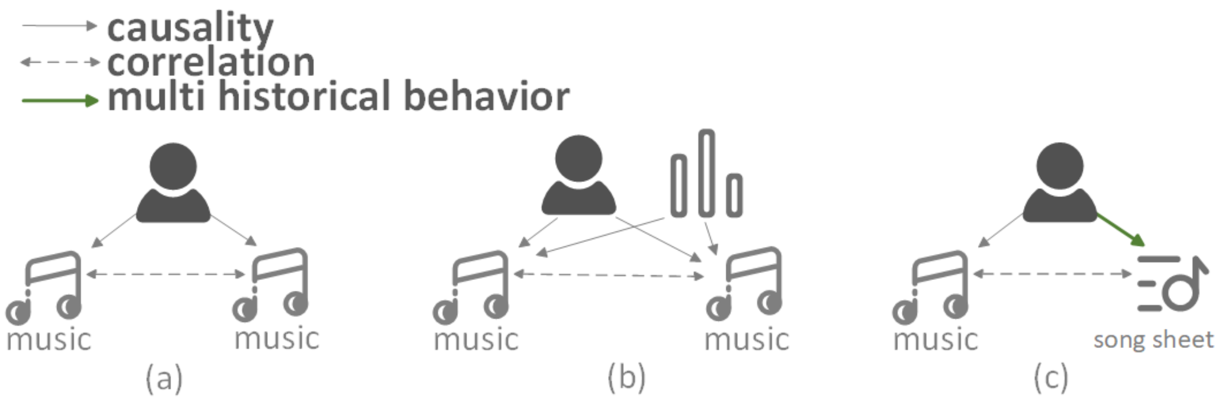

FIGURE 1. (a) The current SBRS generates recommendations by analyzing the co-occurrence patterns between users and music. (b) However, the reliance on co-occurrence between users and music can be misleading. For instance, a user's choice to listen to a specific song might not genuinely reflect an affinity for its genre but could be a result of recent exposure that has temporarily popularized the track. (c) In actuality, there are authentic dependency relationships among items within the SBRS dataset. For example, a user might play a song because their playlist is curated with tracks of a consistent style.

recommendation systems, effectively demonstrating the innovativeness of the proposed model.

II. RELATED WORK

In the context of session-based recommendation, a session denotes a sequence of actions undertaken during an event or over a defined time span (such as a succession of songs streamed). The objective is to predict unseen items within identified sessions by discerning the interdependencies either within a single session or across multiple sessions, which in turn yields more prompt and precise recommendations. This method is geared toward capturing the transient yet evolving preferences of users. In the following section, we provide an overview of pertinent research in the field of session-based recommendation systems.

A. TRADITIONAL RECOMMENDATION METHODS

Before the introduction of matrix factorization, the realm of recommendation systems predominantly relied on collaborative filtering. This method leverages users' historical behavior to identify other users or items that share similar preferences with the target user, consequently suggesting items that might pique the user's interest. Collaborative filtering, however, suffers from an intrinsic limitation: it struggles with sparse data and is particularly susceptible to the "cold start" issue. The advent of matrix factorization has effectively addressed these challenges, offering benefits such as robust generalization, scalability, flexibility, and reduced space complexity. The FPMC [12] model, which combines matrix factorization with first-order Markov chain techniques for recommendations, was an early method to handling sequential data. Nevertheless, Markov chain-based technologies often prioritize the immediate transition between consecutive items, failing to capture the more intricate patterns within sequences.

B. DEEP LEARNING-BASED METHODS

In recent years, the advancement of deep learning has led to notable achievements in modeling sequence data for recommendation systems using neural networks. The CNN model, in particular, has seen significant success, particularly in the field of natural language processing, including applications such as machine translation and dialogue systems. GRU4REC [14] stands out as a seminal RNN model applied to session-based recommendations, being the first to introduce RNNs into this domain. It marked a substantial improvement over traditional methods like KNN and matrix factorization. NARM [20] leverages the attention mechanism of RNNs to capture more meaningful information about item transitions. Similarly, STAMP [3] underscores the significance of the most recent user interaction by incorporating an attention mechanism. Despite the enhancements brought by RNN-based methods to recommendation performance, these methods primarily focus on modeling the sequential transitions between adjacent items and inferring user preferences through given temporal sequences. As a result, they often struggle to effectively model the more complex patterns of item transitions.

C. GRAPH NEURAL NETWORK-BASED METHODS

Graph neural networks represent a category of deep learning models designed for handling graph-structured data. They aggregate information from neighboring nodes through a message-passing mechanism, updating the feature representations of the nodes in the process. This iterative procedure typically spans multiple rounds to encapsulate information from more distant nodes within the graph. The SR-GNN, proposed by the Chinese Academy of Sciences, is a pioneering recommendation system that models session sequences and was the first to apply GNNs to session-based recommendations. The GCE-GNN introduced a globally context-enhanced GNN architecture capable of learning item representations at both global and local scales. HGNN [21] and HyperGCN [22] were among the first to extend graph convolution to hypergraphs. HyperRec [23] constructs a hypergraph based on user interaction sequences to encapsulate user interests. DHCN [24] broke new ground by integrating hypergraph convolutional networks with self-supervised learning for session-based recommendations.

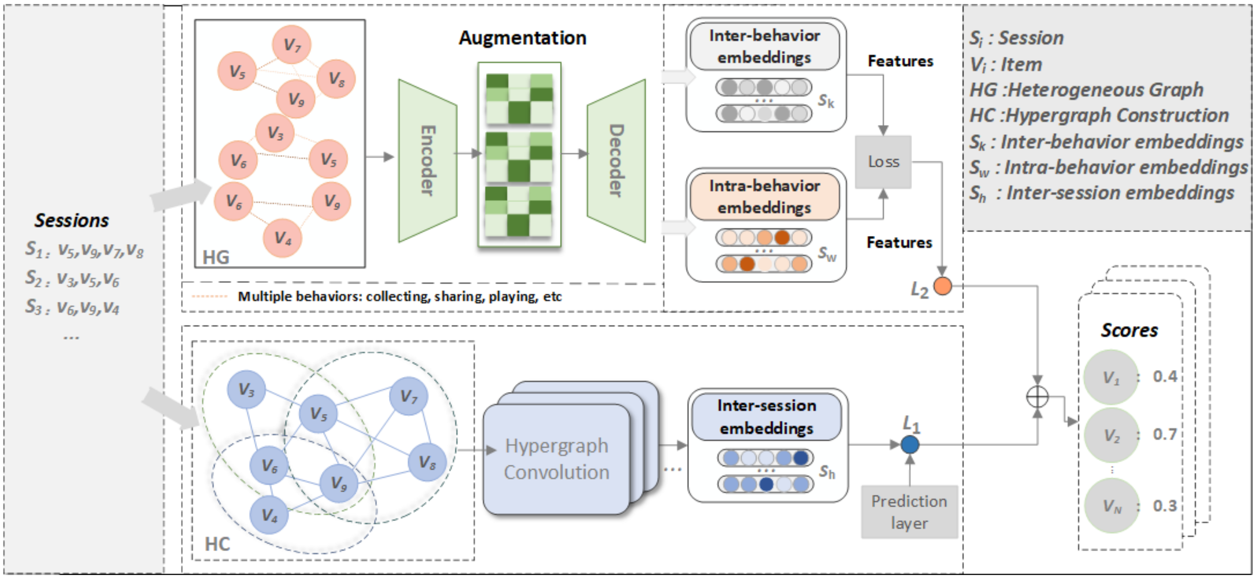

FIGURE 2. The proposed model is structured into two primary components: an upper segment dedicated to the heterogeneous enhancement module and a lower segment comprising the hypergraph module. The heterogeneous enhancement module employs advanced techniques to encapsulate behavioral data, utilizing self-supervised learning to explore the complex interplay of dependencies within and across behaviors. This process integrates diverse behavioral insights, yielding a refined and comprehensive representation of nodes. Additionally, the model incorporates a hypergraph structure, where hypergraph convolutions are applied to capture and exploit the intricate, high-order relationships among items. In the final stage, the model funnels these insights into a specialized prediction module, which produces probabilistic predictions for candidate items.

While these methods have yielded promising results, the structural variations among different graphs lead to distinct item transition relationships being learned, which in turn results in varying representations of user preferences.

The aforementioned methods have indeed enhanced the performance of recommendation systems to a certain degree. However, they predominantly focus on the current user's session data, neglecting the rich item information available within other users' historical sessions. This oversight excludes the collaborative insights from multiple behaviors and the intricate correlations found in cross-session contextual information, which are often instrumental in enhancing session personalization. Our model addresses this by representing the original session sequence as both a multi-behavior heterogeneous graph and a hypergraph. Specifically, the heterogeneous enhancement module effectively augments behavior information to the central node, capturing the complex dynamic interactions among nodes within a session. It delves into the intrinsic connections between behaviors and the interdependencies across different behavior types, yielding precise item embeddings. Concurrently, we introduce a self-supervised training method to mitigate position bias and minimize the disruption caused by noisy behaviors. For recommendations that span sessions, we integrate hypergraphs with pertinent contextual information to deliver accurate and personalized recommendations.

III. PREPARATION

In this section, we commence by formally delineating the key concepts of heterogeneous graphs and sessions that are pertinent to session-based recommendation systems. Then, we propose a multi-behavior heterogeneous graph data enhancement method for session-based recommendations.

Definition 1: A heterogeneous graph is characterized by the presence of multiple types of nodes and edges. A prototypical example is the scholarly data domain, where papers, authors, and research fields are all represented as nodes within the graph. V denote the set of vertices, A represents the categories to which vertices belong, ϵ denote the set of edges, and R represents the categories to which edges belong. A graph is considered heterogeneous when the number of distinct categories A and R exceeds two, indicating the existence of more than one type of vertex, edge, or both. In this work, the vertex category A corresponds to items, while the edge types represent the diverse impacts of various behaviors on items, encompassing the implicit relationships between items.

Definition 2: In a given session S, which encapsulates a sequence of actions performed by the user, we have a set $V = \{v_1, v_2, \ldots, v_n\}$ encompassing all items involved in the interaction. The objective is to utilize the top K items from the recommended set V to forecast the user's subsequent clicks. The sequence of actions is denoted as $S = \{v_1^s, v_2^s, \ldots, v_l^s\}$, with v_i^s signifying the item that user $v_i \in V$ has clicked within session S. The duration of the session is denoted by L. The variable $y\prime = \{y_1\prime, y_2\prime, \ldots, y_N\prime\}$ represents the probability of occurrence for all potential items. To generate recommendations, we select the n items with the highest probabilities from the set $y\prime$.

IV. METHODOLOGY

This paper proposes a model, and Figure. 2 illustrates the overall structure of our model, which mainly consists of a heterogeneous enhancement module and a hypergraph module. Below is a detailed introduction to each part.

A. HETEROGENEOUS ENHANCEMENT MODULE

This section is primarily divided into three modules: 1) Path Enhancement: Project relevance is manifested through co-occurrence behavior in sessions. We generate different weights for each neighboring edge of a project node to distinguish the influence of different neighbor nodes on the central node; 2) Node Enhancement: Referring to MBSSL [25], we utilize an attention mechanism-based behavior-aware graph neural network to capture the dependency relationships of actions within a session, embedding action information to enhance the central node, achieving node enhancement; 3) Self-Supervised Module: Considering the different dependency relationships between behaviors and interactions with the target, learning is conducted at both the intra-behavioral and inter-behavioral levels to obtain multi-behavioral view node representations.

1) HETEROGENEOUS GRAPH PATH ENHANCEMENT

In the multi-behavior heterogeneous graph, we first construct each heterogeneous behavior subgraph and learn each subgraph independently. To accurately model the feature distribution of the central node and its neighbors, we generate weights for the neighboring edges of each project node vi to differentiate the importance of vi neighbors. Specifically, we use co-occurrence information between projects to construct paths and the relevance between two projects is manifested through frequent co-occurrence behaviors in different sessions. Therefore, for two co-occurring projects vi and vj, we define the path (vi ,vi ,$rsimilar$), and the co-occurrence frequency between projects vi and vj is calculated as follows:

$$f(vi, vj) = \frac{\sum_{s \in N(vi) \cap N(vj)} \frac{1}{|N(s)|}}{\sqrt{|N(vi)||N(vj)|}} \tag{1}$$

where $N(vi)$ represents a set of sessions where project vi exists, because there may be multiple edges rin and $rsimilar$ between two projects. To avoid introducing too much noise while capturing the relevance between the central node and its neighbors, for each project node vi, we use the top-K adjacent interaction items. Therefore, the path $K(vi)$ enhancement for project node vi is:

$$K(vi) = \min\{K|N(vi)\} \tag{2}$$

$$N(vi) = \{vj|(vi, vj, r) \ where \ r \in \{rin, rout\}\} \tag{3}$$

2) HETEROGENEOUS GRAPH NODE ENHANCEMENT

Different behavioral type features of nodes are located in different feature spaces. Therefore, to accurately model the feature distribution of behaviors, it is necessary to consider both the features of the central node and the types of behaviors simultaneously. Specifically, we embed actions into the central node and merge the representation of each action into the message passing paradigm, as shown below:

$$s_w^{(l+1)} = \sigma(W^{(l)} \bullet mean(\{S_{i,k}^{(l)} \odot S_k^{(l)} : i \in N_{u,k}\})) \tag{4}$$

Here, we choose $LeakyRelu$ as the activation function, where $s_w^{(l)}$ represents the propagation of node u embedding in behavior type k at the l layer, $s_w^{(l+1)}$ is the embedding of behavior type k at the $l+1$ propagation layer, $S_k^{(l)}$ represents the embedding of action k at the l layer, $N_{u,k}$ is a set of neighbors of vi under behavior type k ; $W^{(l)}$ is the parameter matrix, $\bullet$ represents element-wise multiplication, and W_b is updated by multiplying another specific parameter W :

$$S_k^{(l+1)} = W_b^{(l)} S_k^{(l)} \tag{5}$$

We reflect the dependency between actionsk and the implicit relationships between other user behaviors through the coefficient F_k , thus, the enhanced embedding of behavior type k at the node can be calculated as:

$$F_k = softmax(W_2^k)^T \, tanh((S_e W_1^k)^T) \tag{6}$$

$$S_k^{(l+1)} = W_b^{(l)} S_k^{(l)} \tag{7}$$

where $W_1 \in R^{d \times d'}$, $W_2 \in R^{d'}$ are two behavior-specific parameters, and d' is the size of the output dimension. Since different layers of embeddings represent different connections, we use mean pooling to integrate all layers of embeddings, with the final representation being:

$$S_w = \frac{1}{L} \sum\nolimits_{l=0}^{L-1} S_w^{(l)} \tag{8}$$

$$S_k = \frac{1}{L} \sum\nolimits_{l=0}^{L-1} S_k^{(l)} \tag{9}$$

3) SELF-SUPERVISED MODULE

Considering the different dependency relationships between behaviors and interactions with the target, we introduce a multi-behavior self-supervised learning module. After enhancing the heterogeneous graph encoding in the previous section, we obtained multi-behavioral view node representations, denoted as S_w , S_k . We contrast the positively paired views at the node scale as $\{(S_{u,k}, S_{u,w})\|u \in U\}$ and the negatively paired views as $\{(S_{u,k}, S_{v,w})\|u, v \in U, u \neq v\}$, and the optimization objective is to conduct self-discrimination contrast learning at both the inter-behavioral and intra-behavioral levels through the S_w loss:

$$L_1 = \sum\nolimits_{u \in K} -log(\frac{exp(\phi(S_{u,k}, S_{u,w})/\iota}{\sum_{v \in K} exp(\phi(S_{v,k}, S_{v,w})/\iota}) \tag{10}$$

B. HYPERGRAPH MODULE

Items in the session context exceed pairwise associations and thus contain complex implicit relationships, which hypergraphs model flexibly and naturally. Specifically, we capture the high-order relationships between items by constructing a hypergraph and applying hypergraph convolution. For W_{hh} , we set the hyperedge weights to be the same value of 1, and define hypergraph convolution as follows:

$$x_i^{(l+1)} = \sum\nolimits_{j=1}^{M} \sum\nolimits_{h=1}^{M} H_{ih} H_{jh} W_{hh} x_j^{(l)} \tag{11}$$

Writing the above equation in matrix form is as follows:

$$x_m^{(l+1)} = D^{-1}HWB^{-1}H^TX_m^l \tag{12}$$

Hypergraph convolution aggregates nodes to form hyperedge features and then aggregates the hyperedge features through convolution to obtain new node features, thus forming an aggregation process that transforms item information into nodes hyperedge nodes. Among them, $x_h^{(l)}$ represents the aggregation of information from nodes to hyperedges, and before that, the information from hyperedges to nodes is aggregated by multiplying H . After X^0 undergoes L-layer hypergraph convolution, we sum up the embeddings of each layer and take the average as the final item embedding:

$$X_m = \frac{1}{L+1}\sum\nolimits_{l=0}^{L} X_m^l \tag{13}$$

To enable graph neural networks to better understand the sorting and positional dependencies in projects, we incorporate the project's positional information into the project embedding through positional embedding. $P = [p_1, p_2, \ldots, p_k]$ is the learnable positional embedding matrix, and k is the length of the current session. The project embedding with location information is as follows:

$$x_i^* = tanh\left(W_1\left[x_i \parallel p_{k-i+1}\right] + b\right) \tag{14}$$

where $W_1 \in R^{d\times 2d}$ and $b \in R^d$ are learnable parameters. By aggregating the item representations in the session to generate the session $S = \{v_1^s, v_2^s, \ldots, v_k^s\}$ embedding, we enhance the session embedding representation according to the strategy in SR-GNN:

$$\alpha_i = f^T\sigma\left(W_2x_s^* + W_3x_i^* + c\right) \tag{15}$$

where $f \in R^d$, $W_2 \in R^{d\times d}$, and $W_3 \in R^{d\times d}$ are attention parameters used for learning project weights. S_h is the session representation generated by hypergraph convolution, $x_s^h = \frac{1}{k}\sum_{i=1}^{k} x_i$ is the average embedding representation of all items in the session, and x_s^h is the embedding of the i-th item in session S after adding item information.

C. PREDICTION MODULE

This model harnesses the combined strengths of the heterogeneous graph data augmentation and hypergraph modules to predict item probabilities. The former captures in-session node interactions, while the latter uncovers cross-session relationships, together forming a precise final prediction. In the hypergraph module, perform dot product operation on the embeddings of each initial candidate item and the general user session embeddings obtained in the previous section, and obtain the probability $y\prime$ of a certain item vi being selected next in the target session through *softmax* function operation:

$$y\prime = softmax({S_h}^T h_{vi}) \tag{16}$$

Ultimately, the model is trained by optimizing the following objective function:

$$L_2 = -\sum_{i=1}^{m} y_i \log\left(y_i\prime\right) + (1-y_i)\log\left(1-y_i\prime\right) \tag{17}$$

The final objective function to be minimized through learning is $L = L_1 + \lambda L_2$. At the same time, we use Bayesian optimization techniques to automatically select the optimal λ.

V. EXPERIMENTS AND ANALYSIS

A. EXPERIMENTAL SETUP

In comparative experiments, this paper selects ten classic models to be evaluated on two datasets and two evaluation metrics. We begin by selecting four conventional recommendation algorithms (Item-KNN, FPMC, GRU4REC, and STAMP), followed by four graph neural network-based session recommendation methods (SR-GNN, GCE-GNN, SIR-GNN, and MGIR), and two methods that leverage hypergraphs to encapsulate session context dependencies (SHARE and DHCN). This array of comparative experimental methods encompasses a range of techniques, including collaborative filtering, RNNs, attention-based neural networks, graph neural networks, and hypergraph neural network.

- Item-KNN [26]: This method generates recommendations by leveraging the similarity between items clicked within the current session and other items, with similarity being computed based on session data.
- FPMC [12]: The FPMC model blends matrix factorization with a first-order Markov chain to facilitate recommendation generation.
- GRU4Rec [13]:GRU4Rec is a session-based recommendation system that employs Recurrent Neural Networks (RNNs), utilizing a parallel mini-batch training method specifically tailored for session data.
- STAMP [3]: STAMP incorporates both users' long-term and current preferences, and introduces an attention-based model that focuses on short-term user behavior to predict subsequent item interactions.

The following are five session-based recommendation methods that consider not only the internal session but also the session context.

- SR-GNN [9]: Pioneering the use of graph neural networks in recommendation systems, SR-GNN transforms session data into graph structures, extracting item transition information via GNNs.
- GCE-GNN [11]: This method learns item embeddings at two levels—session graphs and global graphs—to model the movement of items within sessions.
- SIR-GNN [27]: Utilizing an innovative graph neural network, SIR-GNN captures intricate transitions between items and constructs rich representations of interaction recommendation states.
- MGIR [28]: MGIR introduces a novel multi-faceted model of global item relationships, enhancing session representations by incorporating both positive and negative relational dynamics.

The following are two session-based recommendation methods that use hypergraphs to capture session context.

- SHARE [29]: This method integrates sliding window technology with hypergraph attention networks to

TABLE 1. Performance comparison on session-based recommendation.

Methods	Tmall				Nowplaying			
	P@10	P@20	MRR@10	MRR@20	P@10	P@20	MRR@10	MRR@20
Item-KNN	6.65	9.15	3.11	3.31	10.96	15.94	4.55	4.91
FPMC	13.10	16.06	7.12	7.32	5.28	7.36	2.68	2.82
GRU4Rec	9.47	10.93	5.78	5.89	6.74	7.92	4.40	4.48
STAMP	22.63	26.47	13.12	13.36	13.22	17.66	6.57	6.88
SR-GNN	23.41	27.57	13.45	13.45	14.17	18.87	7.15	7.47
GCE-GNN	28.01	33.42	15.08	15.42	16.94	21.23	8.03	8.41
SIR-GNN	28.67	34.93	15.32	16.27	14.40	19.68	6.64	7.01
MGIR	29.43	35.41	15.88	16.98	17.21	21.64	7.64	8.06
SHARE	25.14	30.46	14.13	14.57	16.47	22.42	6.27	7.56
DHCN	26.22	31.42	14.60	15.05	17.35	23.50	7.87	8.18
SSMB-SR	**32.25**	**37.56**	**16.92**	**18.32**	**17.96**	**23.85**	**8.23**	**9.25**

TABLE 2. Statistics of the experimental datasets.

Dataset	Tmall	Nowplaying
Clicks	818,479	1,367,963
Training	351,268	825,304
Testing	25,898	89,824
Items	40,728	60,417
Average Session Length	6.69	7.42

extract evidence of user intent from the contextual text windows.

- DHCN [23]: The DHCN model is founded on self-supervised hypergraph convolution, adept at capturing both item-level high-order relationships and session-level interactions. It employs hypergraphs to model the sequence of sessions effectively.

Datasets: We assessed the proposed method using two real-world datasets, Tmall and Nowplaying.

Tmall: The Tmall dataset is sourced from the IJCAI-15 competition and includes the shopping histories of anonymous users on the Tmall e-commerce platform. Each behavior record in this dataset contains details such as user ID, item ID, timestamp, and the type of action, among others. The recorded behavior categories span a range of actions, including clicks, additions to cart, purchases, and item collections.

Nowplaying: This dataset comes from [13], and it is a collection of 49 million listening event data from Twitter, describing users' music listening behavior.

Data Preparation: This dataset comes from [13], and it is a collection of 49 million listening event data from Twitter, describing users' music listening behavior.

- Delete sessions with a length of 1.
- Delete items that appear less than 5 times.
- The data is divided into training data and test data, with the most recent week's data set aside as test data, and the remaining historical data set as training data.
- Furthermore, we segmented the sessions to create a series of sequences and corresponding labels. The sequences were then generated accordingly. Table 1 summarizes the statistics of the two datasets after our preprocessing.

Evaluation Scheme and Metrics: We used two widely used ranking-based indicators, P@K and MRR@K, to evaluate the performance of each SBRS. Specifically, for each test session, we iteratively select each item as the target item, and the items before this item are the context of the session. For each algorithm, we rank all items according to the recommendation scores output by the model, and calculate P@K and MRR@K from the final sorted scores.

P@K: P@K weighs the precision of the model in the top K predictions. Specifically, it is calculated using the following formul

$$P@K = \frac{n_{hit}}{N} \tag{18}$$

MRR@K: MRR@K is used to weigh the ranking accuracy of the model in the top K predictions. It is specifically calculated using the following formula:

$$M@K = \frac{1}{N}\sum_{t\in L}\frac{1}{Rank(t)} \tag{19}$$

As with the method in [14], we set the dimension of the latent vectors to 100, and the mini-batch size during training to 100. To ensure a fair comparison, the hyperparameters of all models were adjusted to be consistent. The initial learning rate of the model was set to 0.001, and it was decayed by a factor of 0.1 at the end of every three training epochs. We set the L2 penalty to 10^{-5} , and searched for the dropout value in the range 0.1, 0.2, …, 0.9.

It is worth noting that the range of P@K and MRR@K is 0 to 1, and the higher the value, the better the model's accuracy and ranking quality in the top K predictions.

B. RECOMMENDATION PERFORMANCE EVALUATION AND ANALYSIS

In this section, we conducted comparative experiments aimed at evaluating the performance of our developed model against

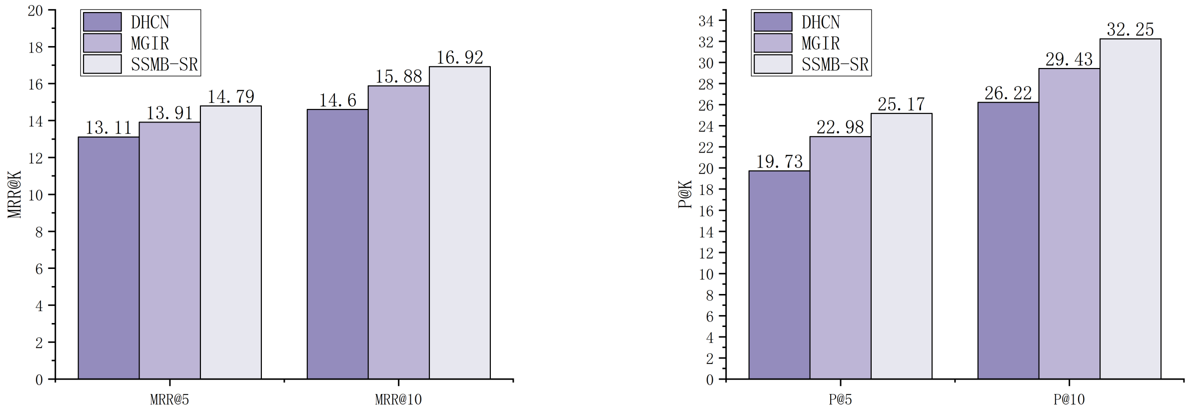

FIGURE 3. Optimized performance comparison for small k values on Tmall.

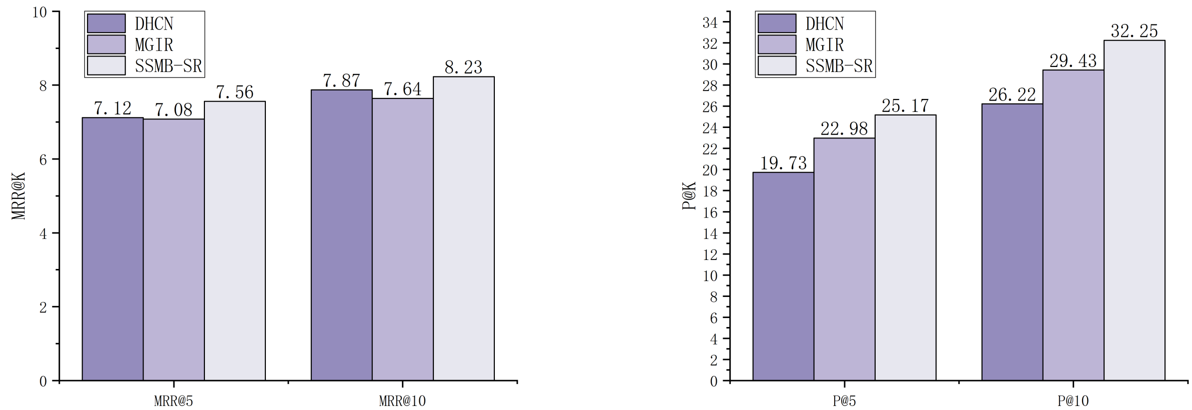

FIGURE 4. Optimized performance comparison for small k values on Nowplaying.

all baseline algorithms. We aim to answer the critical query: ''How does the performance of our proposed model measure up against the established SBRS baseline algorithm?''

As depicted in Table 2, the model proposed in this paper was compared with 10 other baseline models on two datasets for two performance metrics, with the best results highlighted in bold. A review of the data in Table 2 leads to the following insights. Conventional recommendation techniques (Item-KNN, FPMC) exhibit pronounced disparities when contrasted with deep learning-based methods (e.g., GRU4REC, STAMP). This discrepancy arises because traditional methods generally rely on straightforward feature representations, such as item similarity or frequent sequences, whereas deep learning methods are capable of extracting and integrating sophisticated features from the data via multi-layered network architectures and activation functions.This enables a more nuanced understanding of the interactions between users and items. These findings underscore the pivotal role of deep learning in enhancing the effectiveness of session-based recommendation systems.

The session-based recommendation model grounded in graph neural networks notably surpasses its counterparts that rely on recurrent neural networks and attention mechanisms. In particular, the GCE-GNN model outshines the SR-GNN in terms of performance metrics, owing to its incorporation of positional and reverse positional information, along with an attention mechanism that adeptly extracts contextually relevant information from the global graph for the current session. This outcome underscores the criticality of harnessing information across various levels to more effectively model user preferences.

The DHCN model has demonstrated impressive advancements in recommendation quality by leveraging hypergraph structures. This is largely due to its pioneering incorporation of self-supervised learning into the training process for recommendation systems. Self-supervised learning serves to augment the model's expressive capabilities, thereby enhancing its performance in recommendation tasks. Significantly, in datasets characterized by shorter average session lengths, the role of self-supervised learning becomes even more

TABLE 3. Ablation experiment.

Methods	Tmall				Nowplaying			
	P@10	P@20	MRR@10	MRR@20	P@10	P@20	MRR@10	MRR@20
SSMB-SR-G	25.85	30.26	14.80	15.16	17.59	22.34	7.63	8.57
SSMB-SR-H	31.41	35.90	16.36	17.76	16.30	21.11	7.02	8.06
SSMB-SR	**32.25**	**37.56**	**16.92**	**18.32**	**17.96**	**23.85**	**8.23**	**9.25**

critical. This aligns with our hypothesis that the sparsity of session data may limit the advantages of hypergraph modeling. Despite SHARE's enhancement of performance through the integration of a sliding window technique with a hypergraph attention network, its singular modeling method, as revealed by comparative experimental data, also imposes constraints on recommendation performance. Therefore, simply constructing session sequences as hypergraphs to capture dependencies between items may not yield optimal results.

In conclusion, our proposed model exhibits substantial enhancement over traditional baseline methods across both the P@K and MRR@K evaluation metrics. This advancement is attributed to our model's nuanced comprehension of the complex interdependencies among various behaviors within sessions. The model's heterogeneous data augmentation module overcomes graph neural network limitations, enhancing the precision of session node relationships and robustness against data sparsity and noise. Utilizing self-supervised training reduces position bias and integrates relevant context for accurate recommendations. SSMB-SR excels in capturing the intricate connections within sessions and across behavior types, outperforming existing methods in handling complexity, enabling real-time recommendations, and ensuring robustness against noisy behaviors, as evidenced by its superior performance on the Tmall dataset.

C. ABLATION ANALYSIS

To assess the influence of individual modules on the performance of SSMB-SR, we developed two variants: SSMB-SR-G and SSMB-SR-H. SSMB-SR-G is a version of the SSMB-SR model that excludes the heterogeneous enhancement module, whereas SSMB-SR-H omits the hypergraph component. Table 3 presents the performance outcomes for these two abridged variants, as well as the complete SSMB-SR model, on the Tmall and Nowplaying datasets. The data reveal thatSSMB-SR-G yields the least favorable results, affirming the efficacy of the heterogeneous enhancement module. This suggests that our self-attention mechanism-driven behavior-aware graph neural network is adept at capturing the intrinsic connections between behaviors within sessions and the interdependencies across various behavior types. Through self-supervised learning applied at both the intra-behavior and inter-behavior levels, positional biases are mitigated, thereby reducing the impact of noise from behaviors. SSMB-SR-H also exhibits a decline in performance, indicating the value of hypergraphs in capturing intricate dependencies between sessions by facilitating the modeling of relationships that extend beyond pairwise interactions, effectively encoding higher-level information within session contexts. When SSMB-SR integrates both modules, it achieves a comprehensive and effective capture of session dependencies, culminating in optimal recommendation performance.

VI. CONCLUSION

This paper navigates the intricacies of session-based recommendation systems, analyzing the interdependencies between various behavioral types within session data and proposing a self-supervised enhancement method for the complex task of recommending within multi-behavior sessions. By modeling the session sequence as a multi-behavioral heterogeneous graph and hypergraph, we have developed an enhancement module that captures the nuanced interactions among nodes, resulting in precise item embeddings. Our self-supervised training counters position bias and reduces noise interference, while the integration of hypergraphs with contextual information aims to refine cross-session recommendations. The adaptability of our method suits real-time recommendation scenarios, such as online shopping or music streaming. As we progress, industry collaboration will refine our model for reduced inference time and practical deployment. While our work advances SBR theory and offers practical insights, it is crucial to acknowledge its limitations. The model's computational intensity may hinder scalability, and its reliance on high-quality session data could limit performance in scenarios with incomplete or less detailed information. The assumption of equal behavior relevance may not hold for all sessions, and the self-supervised learning could introduce biases if not carefully designed. The model's generalizability across domains and cultural contexts, as well as its adaptability to long-term user behavior, are areas for further investigation. Addressing these limitations will be a focus of our ongoing research to enhance the robustness and applicability of our recommendation system.

REFERENCES

[1] S. Wang, L. Cao, Y. Wang, Q. Z. Sheng, M. A. Orgun, and D. Lian, "A survey on session-based recommender systems," *ACM Comput. Surv. (CSUR)*, vol. 54, no. 7, pp. 1–38, 2021.

[2] K. Ding, Z. Xu, H. Tong, and H. Liu, "Data augmentation for deep graph learning: A survey," *ACM SIGKDD Explor. Newslett.*, vol. 24, no. 2, pp. 61–77, Nov. 2022.

[3] 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. Discovery Data Mining*, Jul. 2018, pp. 1831–1839.

[4] A. Vaswani, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017.

[5] 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. Manage.*, Oct. 2020, pp. 1195–1204.

[6] R. Qiu, J. Li, Z. Huang, and H. Yin, "Rethinking the item order in session-based recommendation with graph neural networks," in *Proc. 28th ACM Int. Conf. Inf. Knowl. Manage.*, 2019, pp. 579–588.

[7] B. Hidasi, M. Quadrana, A. Karatzoglou, and D. Tikk, "Parallel recurrent neural network architectures for feature-rich session-based recommendations," in *Proc. 10th ACM Conf. Recommender Syst.*, Sep. 2016, pp. 241–248.

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

[9] 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.*, vol. 33, 2019, pp. 346–353.

[10] 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.

[11] 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. Develop. Inf. Retr.*, Jul. 2020, pp. 169–178.

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

[13] D. Jannach and M. Ludewig, "When recurrent neural networks meet the neighborhood for session-based recommendation," in *Proc. 11th ACM Conf. Recommender Syst.*, Aug. 2017, pp. 306–310.

[14] B. Hidasi, A. Karatzoglou, L. Baltrunas, and D. Tikk, "Session-based recommendations with recurrent neural networks," 2015, *arXiv:1511.06939*.

[15] J. Song, H. Shen, Z. Ou, J. Zhang, T. Xiao, and S. Liang, "ISLF: Interest shift and latent factors combination model for session-based recommendation," in *Proc. 28th Int. Joint Conf. Artif. Intell.*, Aug. 2019, pp. 5765–5771.

[16] K. Marino, R. Salakhutdinov, and A. Gupta, "The more you know: Using knowledge graphs for image classification," 2016, *arXiv:1612.04844*.

[17] P. Zhang, J. Guo, C. Li, Y. Xie, J. B. Kim, Y. Zhang, X. Xie, H. Wang, and S. Kim, "Efficiently leveraging multi-level user intent for session-based recommendation via Atten-mixer network," in *Proc. 16th ACM Int. Conf. Web Search Data Mining*, Feb. 2023, pp. 168–176.

[18] Y. Chen, Y. Tang, and Y. Yuan, "Attention-enhanced graph neural networks with global context for session-based recommendation," *IEEE Access*, vol. 11, pp. 26237–26246, 2023.

[19] W. Song, S. Wang, Y. Wang, K. Liu, X. Liu, and M. Yin, "A counterfactual collaborative session-based recommender system," in *Proc. ACM Web Conf.*, Apr. 2023, pp. 971–982.

[20] J. Li, P. Ren, Z. Chen, Z. Ren, T. Lian, and J. Ma, "Neural attentive session-based recommendation," in *Proc. ACM Conf. Inf. Knowl. Manage.*, Nov. 2017, pp. 1419–1428.

[21] Y. Feng, H. You, Z. Zhang, R. Ji, and Y. Gao, "Hypergraph neural networks," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, 2019, pp. 3558–3565.

[22] N. Yadati, M. Nimishakavi, P. Yadav, V. Nitin, A. Louis, and P. Talukdar, "HyperGCN: A new method for training graph convolutional networks on hypergraphs," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019.

[23] J. Wang, K. Ding, L. Hong, H. Liu, and J. Caverlee, "Next-item recommendation with sequential hypergraphs," in *Proc. 43rd Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Jul. 2020, pp. 1101–1110.

[24] R. Ye, Q. Zhang, and H. Luo, "Cross-session aware temporal convolutional network for session-based recommendation," in *Proc. Int. Conf. Data Mining Workshops (ICDMW)*, Nov. 2020, pp. 220–226.

[25] J. Xu, C. Wang, C. Wu, Y. Song, K. Zheng, X. Wang, C. Wang, G. Zhou, and K. Gai, "Multi-behavior self-supervised learning for recommendation," in *Proc. 46th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Jul. 2023, pp. 496–505.

[26] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-based collaborative filtering recommendation algorithms," in *Proc. 10th Int. Conf. World Wide Web*, Apr. 2001, pp. 285–295.

[27] L. Shi, Z. Zhang, S. Wang, Q. Zhang, M. Wu, C. Yang, and S. Li, "Session-based interactive recommendation via deep reinforcement learning," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Dec. 2023, pp. 1319–1324.

[28] Q. Han, C. Zhang, R. Chen, R. Lai, H. Song, and L. Li, "Multi-faceted global item relation learning for session-based recommendation," in *Proc. 45th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Jul. 2022, pp. 1705–1715.

[29] J. Wang, K. Ding, Z. Zhu, and J. Caverlee, "Session-based recommendation with hypergraph attention networks," in *Proc. SIAM Int. Conf. Data Mining (SDM)*. Philadelphia, PA, USA: SIAM, 2021, pp. 82–90.

ZHEN ZHANG received the Bachelor of Engineering degree in network engineering from Liaocheng University, in 2022. He is currently pursuing the Master of Science degree in data science and advanced computing with the University of Reading. His research interests include session-based recommendation and sequential recommendation.

JINGAI ZHANG received the Graduate degree from Qingdao University of Science and Technology. She is currently pursuing the master's degree in computer and information security with Guilin University of Electronic Science and Technology, China. She has published two articles and one invention patent. Her main research interest includes session-based recommendation systems.

JINTAO CHEN was born in 1999. He received the B.Eng. degree from Guilin University of Electronic Technology, China, in 2022, where he is currently pursuing the master's degree. His primary research interests include blockchain technology and session-based recommendation systems.

YUZHAO HUANG received the bachelor's degree in mechanical and electronic engineering from Guangdong University of Technology. He is currently pursuing the master's degree with the School of Electronic Engineering and Automation, Guilin University of Electronic Science and Technology. His research interest includes recommendation systems.

XIAOYANG HUANG received the bachelor's degree in automation from Guilin University of Electronic Science and Technology, in 2023, where she is currently pursuing the Master of Engineering degree in control engineering. From 2019 to 2023, she participated in the National Intelligent Car Competition for college students and won the Second Prize of the National Double-Car Relay in the Undergraduate Group.