Graph Contrastive Learning for Multi-Behavior Recommender System

Zhidong Zhao, *Member, IEEE*, Jianing Liu, Nana Huang, Pengfei Jiao, *Member, IEEE*, and Huaming Wu, *Senior Member, IEEE*

***Abstract*—Multi-behavior recommender system aims to model user preference representation based on multiple types of user-item interactions (e.g., viewing, adding to favorites, adding to the cart, and purchasing). However, existing works have two limitations in general: 1) Most of them only concern the sparse observed user-item interactions (explicit interaction) and ignore the huge amount of unobserved user-item interactions (implicit interaction), which are incapable of fully capturing user preference in recommender systems. 2) Previous works typically tend to only extract valuable information by distinguishing target and auxiliary behaviors to model user reference representation, and they fail to explore the fine-grained commonality between target and auxiliary behaviors. To tackle these limitations, we propose a new model named Graph Contrastive learning with Multi-Behavior (GCMB) for a multi-behavior recommender system. Specifically, we utilize Randomized Singular Value Decomposition (rSVD) to inject implicit interaction into the model, and then combine explicit interaction and implicit interaction to learn user preference by graph contrastive learning. Furthermore, we consider the multi-level commonality between target and auxiliary behaviors to capture the fine-grained commonality and then model high quality of user preference representation. Extensive experiments on two real-world datasets demonstrate that our method consistently outperforms various state-of-the-art recommender methods.**

***Index Terms*—Multi-behavior recommender system, graph contrastive learning, explicit collaborative relation learning, implicit collaborative relation learning, rSVD.**

I. Introduction

With the explosive growth of online information, recommender systems play a key role to alleviate such information overload [1]. The recommender system aims to learn user preference and predict the items that he/she will be interested in based on the observed historical interactions between users and items. Due to the important application value of recommender systems, they have become indispensable tools for online applications, e.g., online e-commerce platform [2], [3], [4], tourism services platforms [5] and online video [6] or music platform [7].

Recently, the development of recommender systems has advanced rapidly in both academia and industry. However, existing works typically concentrate on single-type interactions between users and items (e.g., click records), overlooking the diverse array of behaviors that can occur between them. For example, in an online e-commerce platform, users can engage with items in multiple ways, such as viewing, adding to favorites, adding to the cart, and purchasing. Typically, we suppose purchase behavior is the primary target behavior, while assume other types of behaviors are auxiliary behavior. These diverse behaviors offer valuable signals for constructing a comprehensive user preference representation, which proves beneficial in mitigating the challenge posed by significant data sparsity within the context of the target behavior [8].

To leverage these different types of behaviors, several efforts on methods of the multi-behavior recommender system have been made. For example, some methods attempt to capture the relationship between auxiliary and target behaviors while modeling the representation of nodes. These methods use attention scores to assign weights to the nodes of different behaviors, integrating them to derive the node representation for the target behavior [9], [10]. To further capture the implicit relationship between auxiliary and target behaviors, both the [11] and [12] utilize a relation-aware encoder to capture the hidden dependencies between auxiliary and target behaviors under a message-aggregation architecture. Other methods [13], [14], [15], [16] propose that these behaviors often follow certain ordinal relations, then they utilize the relationships between different behaviors to learn the node representation. In addition to distinguishing the semantic information of various types of behaviors, [17] and [18] use meta-paths or second-order neighbors to capture the item-item correlations reflected in different types of behaviors.

Despite the effectiveness of the existing methods above, it is nontrivial to effectively model user preference representation through multiple behaviors. **Firstly**, most existing methods only utilize the sparse explicit interaction and ignore the huge amount of implicit interaction. However, in the huge amount of implicit interaction data, the most of non-interaction phenomena are not caused by user's dislike, but by the limitation of platform push opportunities, which makes users

Received 28 June 2024; revised 29 October 2024 and 19 November 2024; accepted 30 November 2024. Date of publication 4 December 2024; date of current version 14 August 2025. This work was supported in part by the Zhejiang Provincial Natural Science Foundation of China under Grant LDT23F01012F01; in part by the Research Project of the Zhejiang Provincial Department of Education under Grant Y202455123; and in part by the National Natural Science Foundation of China under Grant 62372146.
(Corresponding authors: Nana Huang; Huaming Wu.)

Zhidong Zhao, Nana Huang, and Pengfei Jiao are with the School of Cyberspace and the Data Security Governance Zhejiang Engineering Research Center, Hangzhou Dianzi University, Hangzhou 310018, China (e-mail: zhaozd@hdu.edu.cn; nanahuang@hdu.edu.cn; pjiao@hdu.edu.cn).

Jianing Liu is with the School of Intelligent Manufacturing, Hebi Automobile Engineering Professional College, Hebi 458030, Henan, China (e-mail: ljn2000@hdu.edu.cn).

Huaming Wu is with the Center for Applied Mathematics, Tianjin University, Tianjin 300072, China (e-mail: whming@tju.edu.cn).
Digital Object Identifier 10.1109/TCE.2024.3510862

unable to interact with these items and give feedback. Sparse exolicit interaction data will seriously affect the quality of user preference representation. Although contrastive learning [19] has been used to enhance the performance of graph-based recommender methods and can alleviate the data sparsity problem [20], [21], these methods often generate contrastive views through random perturbations, which may lead to the loss of valuable structural information, potentially misleading the learning of node representation. Besides these methods also ignore the implicit interaction. **Secondly**, the certain user could connect to items through different behaviors, so we suspect that there are some commonalities (e.g., brand, price, color) between items. And these commonalities can reflect the key reasons why users interact with items; these commonalities are the overall user preferences under different behaviors. But the above works tend to overlook the commonality between target and auxiliary behaviors, which is a crucial factor in determining whether users generate target behaviors with the item. Although [10] utilizes the commonality between target and auxiliary behaviors, it only focuses on the commonality between the final preference representation of target and auxiliary behaviors, neglecting the commonality between other layers of user preference representation. The commonality between the final preference representation of target and auxiliary behaviors is broad and coarse-grained rather than fine-grained.

In light of the aforementioned limitations, we revisit the graph contrastive learning paradigm for multi-behavior recommenders, introducing an effective augmentation method known as GCMB. This approach takes into account the multi-level commonality between target and auxiliary behaviors during the preference representation learning process. Specifically, in order to take into account both explicit and implicit collaboration relationship when modeling user preference, we first utiliz rSVD to prefill the implicit collaboration relationship under multi-bahavior and generate a contrastive view. Then, the implicit collaboration relationship is injected into the model by aligning the original view and the contrastive view under different behaviors. This allows our model to extract additional information from the contrastive view to enhance the preference representation in the original view. Furthermore, our model considers the multi-level commonality between target and auxiliary behaviors to capture fine-grained similarities. This capability allows our model to refine reference representations further and achieve high-quality learning outcomes.

In summary, we have made the following contributions to this work:

- We emphasized the importance of implicit interaction, and proposed to explore the relationship of implicit multi-behavior interaction. And then we utilize implicit interaction to improve and enhance the quality of user preference representation.
- In our model, we explored the fine-grained commonality to refine and improve the quality of user preference representation.
- The effectiveness of our GCMB model is demonstrated on two real-world datasets, showing improved recommender performance compared to baselines.

II. Related Work

A. Graph-Based Recommender Models

Recently, Graph Neural Network (GNN) techniques have been widely utilized in recommender systems [1]. Some works [22], [23] apply random walk method in graph-based recommenders systems. Reference [22] constructs an item-item similarity graph based on the user-item graph and runs the item-item graph on a variant of the [24] algorithm called [25]. Graph Convolutional Networks (GCN) [26], [27], [28] have demonstrated significant advantages in graph representation learning, due to the essence of data in recommender system is graph structure, so GCN has been widely applied in recommender systems [29], [30], [31]. [31] remove non-linear activation functions and feature transformations to simplify GCN.

B. Multi-Behavior Recommender Systems

Multi-behavior recommender systems aim to enhance the recommender performance of the target behavior by utilizing multiple types of auxiliary behaviors. Existing works can be classified into two categories based on the relationships between behaviors.

Firstly, some works consider that different types of interaction behaviors often follow certain orders (e.g., click >add to cart >purchase). For example, [13] proposes a model that associates each behavior type's predictions in a cascading manner. Reference [14] associates the predictions of each behavior in a transitive manner. Reference [15] builds on previous works and utilizes GCN to capture higher-order information in the graph. Reference [32] incorporates the cascading relationships between behaviors into the learning process of embedding representation. Reference [16] is an extension of [32] that designs feature transformation modules to avoid misleading embedding learning. Then it aggregates the learned embeddings of different types of behaviors for final prediction.

Secondly, some works treat some behaviors as strong signals, while others may be regarded as weak signals, so they use attention scores to represent the weights of different behaviors. Reference [17] utilizes attention mechanisms in the propagation layer to learn behavior strength while capturing behavior semantics through item-item propagation layers to aid in better learning of embedding representation. References [11] and [12] further consider the dependency between different types of behavior embeddings in the learning process.

C. Contrastive Learning for Multi-Behavior Recommender Systems

Reference [33] proposed a new graph contrastive learning based framework by coupling with hyper metapaths to learn embeddings of user behavior patterns adaptively. Reference [34] proposed a contrastive meta network to capture the diverse multi-behavior patterns. Reference [10] further employed contrastive learning modules to capture the commonality between behaviors. While these three papers utilized comparative learning to capture commonalities in multi-behavioral patterns, they differ from our proposed approach

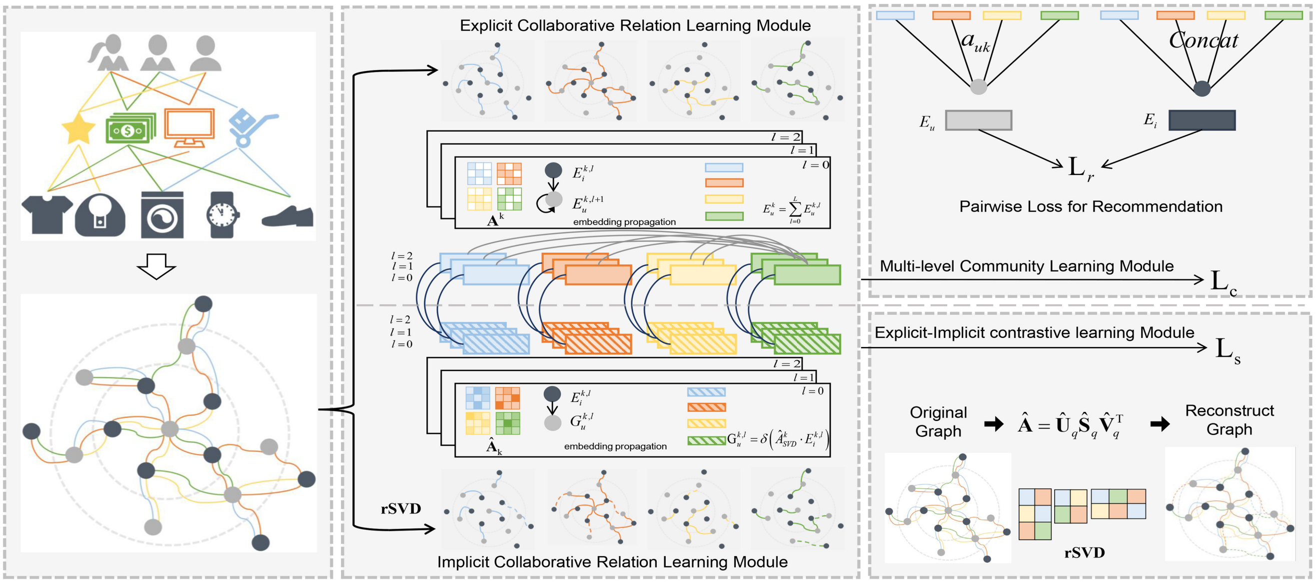

Fig. 1. The model architecture of GCMB. i) Explicit Collaborative Relation Learning Module learns the node representation under multiple behaviors interaction. ii) Implicit Collaborative Relation Learning Module extracts implicit collaborative signals by rSVD. iii) Multi-level Community Learning Module captures the commonality between target and auxiliary behaviors. iv) Explicit-Implicit Contrastive Learning Module combine explicit and implicit interaction to learn user preference by graph contrastive learning.

in that this paper integrates explicit and implicit interactions between users and items, and considers multi-level commonalities between target and ancillary behaviors to enhance user preference representations.

III. Preliminaries

A. Problem Definition

Firstly, we define the interaction between nodes as graph $G = (V, E)$, where nodes V consist of the node of users $u \in U$ and items $i \in I$. And the edge E consists of K ($K \geq 2$) different types of user-item interaction edges. Besides the user-item interaction edges under the k_{th} ($1 \leq k \leq K$) behavior is defined as E^k, E^k together with all nodes (users and items) can be defined as a subgraph $G_k = (V, E^k)$, which also can be expressed as an interaction adjacent matrix $A^k \in \mathcal{R}^{|U|*|I|}$. We hypothesize the first behavior is target behavior, and other $k-1$ behaviors are auxiliary behaviors. Usually, the target behavior is purchase, and it is the prediction objective, the other behaviors are regarded as auxiliary behavior, and assist the target behavior to complete recommend tasks.

The research problem in our study is defined as follows **Input**: consists of user-item interactions across various types of behavior, labeled as $\{G_1, \ldots, G_k, \ldots, G_K\}$. **Output**: a predictive function that estimates the likelihood of user u will interact with item i under the target behavior.

B. Overall Framework

The overall framework of our GCMB model is illustrated in Figure 1 and consists of four key modules. First, in the Explicit Collaborative Relation Learning Module, we utilize GCN to learn node embedding representations from user-item explicit interaction subgraphs under different types of behaviors. These embeddings are then integrated using automatically learned weight coefficients. Second, in the Implicit Collaborative Relation Learning Module, we employ rSVD to extract implicit collaborative signals from a global perspective. This step pre-populates implicit collaborative relationships under multiple behaviors and injects them into the GCMB model, thereby enhancing the user preference representation. Third, in the Multi-level Commonality Learning Module, we capture the multi-level commonalities between target behaviors and auxiliary behaviors through a multi-level aggregation layer learning approach. Specifically, we learn the commonalities between these behaviors at different aggregation layers to achieve finer-grained and high-quality node representations. Finally, in the Explicit-Implicit Contrastive Learning Module, we combine explicit and implicit interactions through graph contrastive learning to better learn user preferences and improve the overall performance of the model. In summary, by introducing rSVD to pre-populate implicit collaborative relationships and using a multi-level commonality learning module to capture the commonalities between different behaviors, our GCMB model performs well in recommendation tasks.

C. Explicit Collaborative Relation Learning

First, we derive the embedding vectors $E_u \in R^d$ and $E_i \in R^d$ for user u and item i through initialization, where d is the embedding dimension. Next, we perform multi-layer GCN to aggregate neighborhood information for each node in multiple types of behavior subgraphs. The process of aggregation is represented as follows:

$$Z_u^{k,l} = \sigma\left(p\left(A^k\right) \cdot E_i^{k,l}\right), \quad (1)$$

$$Z_i^{k,l} = \sigma\left(p\left(A^{k\top}\right) \cdot E_u^{k,l}\right), \quad (2)$$

where $Z_u^{k,l}$ and $Z_i^{k,l}$ represent the aggregated embedding for user u and item i in the l_{th} layer under the k_{th} behavior. We apply the activation function $\sigma(\cdot)$ using a LeakyReLU with a negative slope of 0.5. A^k denotes the normalized adjacency matrix, and the application of edge dropout is indicated as $p(\cdot)$, aiming to mitigate overfitting concerns. Furthermore, to preserve the inherent node information, we incorporate residual connections between each layer:

$$E_u^{k,l+1} = Z_u^{k,l+1} + E_u^{k,l}, \tag{3}$$

$$E_i^{k,l+1} = Z_i^{k,l+1} + E_i^{k,l}. \tag{4}$$

The final embedding for a node is the sum of its embeddings across all layers,

$$E_u^k = \sum_{l=0}^{L} E_u^{k,l}, \tag{5}$$

$$E_i^k = \sum_{l=0}^{L} E_i^{k,l}. \tag{6}$$

Inspired by [10], the final node embedding is expressed as follows:

$$a_{uk} = \frac{exp(W_k * n_{uk})}{\sum_{m=1}^{K} exp(W_m * n_{um})}, \tag{7}$$

$$E_u = W_u\left(\sum_{k=0}^{K} a_{uk} \cdot E_u^k\right) + b_u, \tag{8}$$

$$E_i = W_i\left(Concat\left(E_i^k\right)\right) + b_i, \tag{9}$$

where W_k is considered a strength weight for behavior k, which remains constant for all users. n_{uk} represents the relative number of interactions edges under behavior k of user u. Additionally, W_u, b_u, W_i, b_i are the weight and bias of neural network. The final step involves calculating the inner product between the ultimate embedding E_u and E_i to predict user u's preference towards item i. In order to optimize this module, we employ the pairwise loss.

$$L_r = \sum_{(u,i,j \in R)} max\left(0, 1 - E_u^\top E_i + E_u^\top E_j\right), \tag{10}$$

where $R = (u, i, j) \mid (u, i) \in R_+, (u, j) \in R_-$, and R_+ is the observed interactions, R_- is the unobserved interactions.

D. Implicit Collaborative Relation Learning

To further alleviate the issue of sparse explicit multi-behavior data, we propose to use rSVD as guidance to generate the contrastive view under different types of behaviors and align the node representation between the original view and the contrastive view under different types of behaviors. By doing so, we effectively extract supplementary information from the contrastive view, thereby enhancing the quality of node representation in the original view. Specifically, we begin by performing SVD [35] on the adjacency matrix A_k, $A^k = U^k S^k V^{k\top}$. However, conducting SVD on large matrices is computationally expensive. Therefore, drawing inspiration from the rSVD algorithm [8], [36], [37], we opt to approximate the range of the input matrix through a lower-rank orthogonal matrix. Subsequently, we apply the SVD procedure to this reduced matrix, effectively mitigating the computational challenges associated with larger matrices.

$$\hat{U}_q^k, \hat{S}_q^k, \hat{V}_q^{k\top} = rSVD\left(A^k, q\right), \tag{11}$$

$$\hat{A}_{SVD}^k = \hat{U}_q^k \hat{S}_q^k \hat{V}_q^{k\top}, \tag{12}$$

where q is the required rank for the decomposed matrices, and $\hat{U}_q^k \in \mathcal{R}^{U\times q}$, $\hat{S}_q^k \in \mathcal{R}^{q\times q}$, $\hat{V}_q^k \in \mathcal{R}^{I\times q}$ are the approximated versions of U^k, S^k, V^k. The global aggregation process is expressed as follows:

$$G_u^{k,l} = \sigma\left(\hat{A}_{SVD}^k \cdot E_i^{k,l}\right), \tag{13}$$

$$G_i^{k,l} = \sigma\left(\hat{A}_{SVD}^k{}^\top \cdot E_u^{k,l}\right). \tag{14}$$

E. Multi-Level Commonality Learning

In traditional multi-behavior recommendation methods, the commonalities between target and auxiliary behaviors are often overlooked, although these commonalities are crucial for predicting whether user-item interactions will occur. Existing methods typically rely on the final node representations to learn the commonalities between behaviors, which can result in overly broad and coarse-grained commonality information.

To address this issue, we propose a Multi-level Commonality Learning Module that captures the commonalities between target behavior k and auxiliary behavior k' across different aggregation layers, thereby achieving finer-grained and higher-quality node representations. The specific steps are as follows:

First, Multi-layer Aggregation Learning. In the multi-layer aggregation learning, we use GCN to learn node embedding representations across multiple propagation layers. Each propagation layer captures different levels of local structural information. Through multi-layer aggregation, we can obtain node representations at various levels of granularity, from coarse to fine. Specifically, the node embedding representation $Z_{k,l}^u$ at layer l can be calculated using the following formula:

$$Z_{k,l}^u = \sigma\left((A_k) \cdot E_{k,l}^i\right), \tag{15}$$

where A_k is the adjacency matrix for behavior, $E_{k,l}^i$ is the node embedding at layer l, and σ is the activation function. This multi-layer aggregation approach captures local structural information at different levels, thereby providing richer node representations.

Second, multi-level commonality extraction. After each propagation layer, we compute the similarity between the node representations of the target behavior k and the auxiliary behavior k'. Specifically, we use cosine similarity or other similarity measures to quantify the commonalities between these node representations. By doing so, we can capture the multi-level commonalities between the target and auxiliary behaviors across different propagation layers. This multi-level commonality extraction helps to more comprehensively understand user preference patterns across different behavior types, thereby improving the model's accuracy.

Next, the contrastive learning loss function. To further improve the quality of node representations, we introduce the InfoNCE loss function. This loss function uses contrastive learning to compare the node representations of the target behavior and the auxiliary behavior. The specific formula is as follows:

$$L_u^{k,k'} = -log \frac{exp\left(sim\left(Z_{k,l}^u, Z_{k',l}^u\right)/\tau\right)}{\sum_{u' \in U} exp\left(sim\left(Z_{k,l}^u, Z_{k',l}^{u'}\right)/\tau\right)}, \tag{16}$$

where $Z_{k,l}^u$ and $Z_{k',l}^u$ represent the node embedding representations of user u at layer l for behavior k and behavior k', respectively. $sim()$ is the similarity function, τ is the temperature parameter, and U is the set of users. Through contrastive learning, we can better distinguish the commonalities between target and auxiliary behaviors, thereby improving the quality of the node representations.

Finally, integration into the model. The multi-level commonalities extracted through the above steps are integrated into the model to optimize the node representations. Specifically, we combine the loss from multi-level commonality learning with the loss from explicit-implicit contrastive learning to jointly optimize the overall performance of the model. The overall loss function is as follows:

$$L_c = \sum_{k'} \left(L_u^{k,k'} + L_i^{k,k'}\right). \tag{17}$$

Through this approach, the multi-level commonality learning module not only captures the multi-level commonalities between target and auxiliary behaviors but also effectively integrates both explicit and implicit collaborative information, thereby enhancing the accuracy and personalization of recommendations.

F. Explicit-Implicit Collaborative Relation Contrastive Learning

Traditional graph-based contrastive learning methods for recommender systems often utilize the three-view paradigm, where generated contrast views are used for contrastive learning, while the original view is not involved in contrastive learning loss. Because traditional methods usually use randomly perturbed generated contrast views and that may mislead the original view.

In contrast, our proposed method takes rSVD as a guide to generate contrast views based on global collaborative relationships. We combine explicit interaction and implicit interaction to learn user preference by graph contrastive learning. Consequently, we directly simplify the contrastive learning in the infoNCE loss by aligning the representation of the original and contrast views for different types of behaviors. The InfoNCE loss L_s^i for the items are defined in the same way.

$$L_s^u = \sum_{k,l,u} -log \frac{exp\left\{s\left(Z_u^{k,l}, G_u^{k,l}/\tau\right)\right\}}{\sum_{u'=0}^{U} exp\left\{s\left(Z_u^{k,l}, G_{u'}^{k,l}/\tau\right)\right\}}. \tag{18}$$

The overall loss function of the Explicit-Implicit contrastive learning module can be obtained as below,

$$L_s = L_s^u + L_s^i. \tag{19}$$

TABLE I
DATASET STATISTICS

Dataset	#User	#Item	#View	#Add-to-cart	#Purchase
Beibie	21,716	7,977	2,412,586	642,622	304,576
Taobao	48,749	39,493	1,548,126	173,747	259,747

G. Joint Optimization

We combine the above two modules to optimize our recommender model, which the overall loss function of model is formalized as,

$$L_s = L_r + \lambda_1 L_s + \lambda_2 L_c + \mu \|\Theta\|_2^2, \tag{20}$$

where λ_1, λ_2 and μ are hyperparameters to control the influence weight of Multi-level Commonality Learning Module and Explicit-Implicit Contrastive Learning Module and L_2 regularization, respectively, Θ represents all trainable parameters in our Modules.

IV. EXPERIMENTS

A. Experimental Settings

1) Dataset: To demonstrate the superior performance of our model GCMB, we conducted experiments on two real-world datasets: BeiBei and Taobao [14] in Tabel I. The BeiBei dataset and the Taobao dataset both consist of three different types of user behaviors: view, add to cart, and purchase. The interaction data in both datasets exhibit high sparsity, with the sparsity level of 1.93% and 0.10%. Specifically, in the BeiBei dataset, the sparsity rates for the various behaviors stand at 1.39%, 0.37%, and 0.17%, respectively. In parallel, the Taobao dataset showcases sparsity rates of 0.08%, 0.01%, and 0.01% for the same behaviors. Besides we observed the purchase behavior data in both datasets is severely sparse. In the BeiBei dataset, around 76% of users have only 0 to 15 purchase behavior interactions, while this percentage is as high as 99% in the Taobao dataset. In summary, both of these datasets are highly sparse.

2) Baseline: We compare our GCMB with the following state-of-the-art methods:

- BPR [38] supposed that observed interactions should have a higher likelihood than unobserved interactions.
- NCF [39] used a multi-layer MLP to enhance the embedding paradigm in Collaborative Filtering (CF), in order to achieve non-linear feature interactions.
- NGCF [40] was an advanced CF model based on GNN, which utilized multi-layer information propagation to capture multi-behavior representations containing high-order semantic information.
- LightGCN [31] performed node embedding by neighborhood aggregation on the graph and removed the transformation and nonlinear activation on the basis of GCN.
- LightGCL [41] designed a view contrastive enhancement strategy guided by rSVD for the single-behavior recommender.

TABLE II
THE PERFORMANCE OF MODEL WITH THE METRICS OF RECALL@K AND NDCG@K (K=10, 20, 40) ON BEIBEI, TAOBAO

Beibei		Recall@10	Recall@20	Recall@40	NDCG@10	NDCG@20	NDCG@40
single-behavior	BPR	0.0315	0.0482	0.0862	0.0204	0.0237	0.0314
	NCF	0.0368	0.0494	0.0931	0.0184	0.0242	0.0321
	NGCF	0.0383	0.0643	0.1068	0.0188	0.0253	0.0339
	LightGCN	0.0389	0.0638	0.1076	0.0192	0.0257	0.0346
	LightGCL	0.0382	0.0604	0.1018	0.0202	0.0257	0.0341
multi-behaviors	NMTR	0.0389	0.0651	0.1092	0.0192	0.0258	0.0348
	EHCF	0.0383	0.0642	0.1084	0.0196	0.0261	0.0351
	GNMR	0.0384	0.0667	0.1173	0.0191	0.0261	0.0381
	MBRec	0.0395	0.0676	0.1196	0.0198	0.0272	0.0394
	S-MBRec	0.0403	0.0677	0.1163	0.0205	0.0273	0.0372
	MB-CGCN	0.0458	0.0726	0.1314	0.0221	0.0295	0.0413
	GCMB	**0.0558**	**0.0941**	**0.1478**	**0.0264**	**0.0360**	**0.0469**
Taobao		**Recall@10**	**Recall@20**	**Recall@40**	**NDCG@10**	**NDCG@20**	**NDCG@40**
single-behavior	BPR	0.0143	0.0211	0.0305	0.0080	0.0097	0.0116
	NCF	0.0182	0.0238	0.0401	0.0101	0.0921	0.0147
	NGCF	0.0206	0.0289	0.0412	0.0115	0.0136	0.0161
	LightGCN	0.0219	0.0291	0.0420	0.0124	0.0143	0.0177
	LightGCL	0.0236	0.0342	0.0477	0.0132	0.0158	0.0186
multi-behaviors	NMTR	0.0258	0.0481	0.0656	0.0157	0.0202	0.0279
	EHCF	0.0276	0.0499	0.0671	0.0153	0.0210	0.0285
	GNMR	0.0319	0.0458	0.0687	0.0164	0.0209	0.0288
	MBRec	0.0327	0.0463	0.0695	0.0177	0.0211	0.0292
	S-MBRec	0.0336	0.0467	0.0664	0.0182	0.0214	0.0252
	MB-CGCN	0.0366	0.0596	0.0881	0.0207	0.0265	0.0324
	GCMB	**0.0744**	**0.0989**	**0.1289**	**0.0451**	**0.0512**	**0.0574**

- NMTR [13] considered the relationship of multiple behaviors as the cascading relationship and optimized the model under the multi-task learning framework.
- EHCF [14] proposed a novel non-sampling transfer learning solution.
- GNMR [11] explicitly modeled the dependencies between different types of user-item interactions under a graph-based message-passing architecture.
- MBRec [12] focused on the collaborative relationship of behavior patterns between cross-layer preference representations.
- S-MBRec [10] used a multi-layer graph convolutional neural network to capture behavior preference representations that contained high-order semantic information and learned the differences and commonalities between different behaviors.
- MB-CGCN [16] utilized behavior dependencies to model the preference representation in the preference representation learning process.

3) Evaluation Metrics: In order to fully evaluate the effectiveness of our model, we adopt two representative evaluation metrics in the field of recommender: Recall@K and NDCG@K [42].

4) Parameters Setting: The implementation environment of our model GCMB is PyTorch. The learning rate is 0.001. The training batch size is 256. The embedding dim is 32. The L_2 regularization coefficient is 0.1 and 0.2 for Beibei and Taobao. The rank of rSVD is 4 and 2 for Beibei and Taobao. λ_1 is searched from $\{0.0001, 0.0005, 0.001, 0.005, 0.01\}$, λ_2 is searched from $\{0.02, 0.025, 0.03, 0.035, 0.04\}$, and the temperature coefficient τ is searched in $\{0.05, 0.1, 0.2, 0.3, 0.4, 0.5\}$.

B. Performance Comparison

To fully demonstrate the superiority of our model, we conducted experiments with different values of $K = \{10, 20, 40\}$. From Table II, we summarized the following observations.

Firstly, compared to all baseline methods, our model achieved the best performance. When compared to the best baseline method, our model showed significant improvements in Recall@20 on both datasets, with an increase of 29.6% and 65.9%. We attribute this improvement to two main reasons: 1) Generating contrastive views guided by rSVD and then directly aligning the node representation between the original view and the contrastive view under different types of behaviors. The model can effectively extract additional information from the contrastive views of multi-behavior to enhance the node representation in the original views. 2) Learning multi-level commonalities between target and auxiliary behaviors enhanced the quality of the node representation under target behavior.

Secondly, the overall performance of the multi-behavior recommender in the baselines outperformed the single-behavior recommender. This reveals that auxiliary behaviors provide valuable and useful information for the multi-behavior recommender system, positively impacting the recommendation

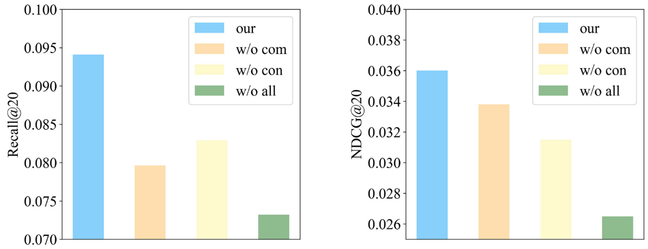

Fig. 2. The ablation study of Beibei.

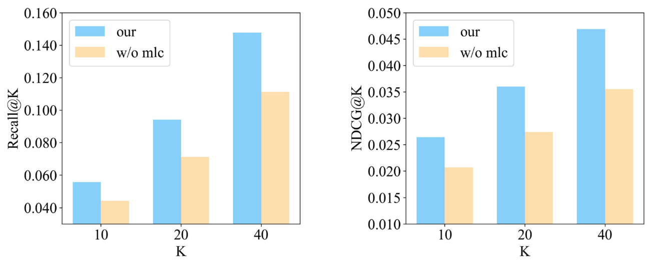

Fig. 3. The ablation study of Beibei about the Multi-level Commonality Learning module, with the metrics of Recall@K and NDCG@K (K=10, 20, 40).

for the target behavior. When compared with the best method in single-behavior recommender method, we discover that our model achieved significant improvements, further validating the above conclusions.

C. Ablation Study

To explore the importance of the Explicit-Implicit Contrastive Learning Module and the Multi-level Commonality Learning Module in our model, we conducted experiments by individually removing each module and comparing the recommender performance. As shown in Figure 2, we use the Beibei dataset as an example. And w/o com means to remove the Multi-level Commonality Learning Module, w/o con means to remove the Explicit-Implicit Contrastive Learning Module.

We observed a significant decrease in experimental results when either of the modules was removed. Therefore, we can conclude that both modules have a positive effect on this model, and the Explicit-Implicit Contrastive Learning Module can alleviate the issue of sparse observed user-item interaction data.

Furthermore, we compared the performance of model when only consider the commonality between final representation and multi-level commonality. As shown in Figure 3, w/o mlc means only consider the commonality. So, we can demonstrate the Multi-level Commonality Learning Module is indispensable.

D. Performance on Sparse Data

In this section, we aim to demonstrate the effectiveness of the Multi-level Commonality Learning and the Explicit-Implicit Contrastive Learning Modules in alleviating data sparsity. We also seek to showcase the superior recommender performance of our model on sparse datasets. To

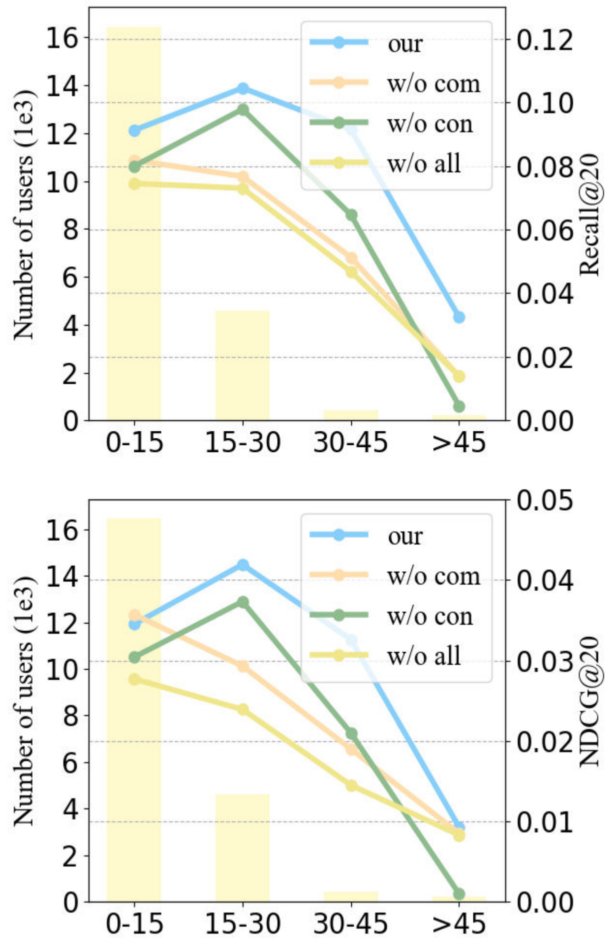

Fig. 4. The performance of sparse data for GCMB.

achieve this, we divided users into four groups based on the number of interactions they have with items under the target behavior, i.e., $\leq$15, 15-30, 30-45, and $\geq$45. Then we compared the Recall@20 of different user groups in the Beibei dataset through ablation experiments. Recall@20 is the average recall value across user groups, as shown in Figure 4.

We observed our model outperforms all ablation experiments across different user groups, and w/o com and w/o con also outperform w/o all across different user groups. This demonstrates our model can alleviate the issue of sparse multi-behavior data, and both the Multi-level Commonality Learning and the Explicit-Implicit Contrastive Learning Modules play a key role. Besides, the performance of w/o com is better than w/o con when the user group is $\leq$15. This also demonstrates that the Explicit-Implicit Contrastive Learning Module is more significant when data is sparse. As the Explicit-Implicit Contrastive Learning Module enhances node representations of different types of behaviors, it provides a certain guarantee for graph learning.

Additionally, we compared the performance of our model with two representative baselines, [10] and [16]. As shown in Figure 5 and Figure 6, our model consistently outperforms [10] and [16] in all user groups, confirming its superior recommender performance on sparse datasets.

TABLE III
PERFORMANCE OF GCMB WITH DIFFERENT LAYER ON BEIBEI

layer	Recall@10	Recall@20	Recall@40	NDCG@10	NDCG@20	NDCG@40
1	0.0070	0.0186	0.0405	0.0030	0.0058	0.0102
2	**0.0558**	**0.0941**	**0.1478**	**0.0264**	**0.0360**	**0.0469**
3	0.0408	0.0704	0.1203	0.0195	0.0269	0.0370

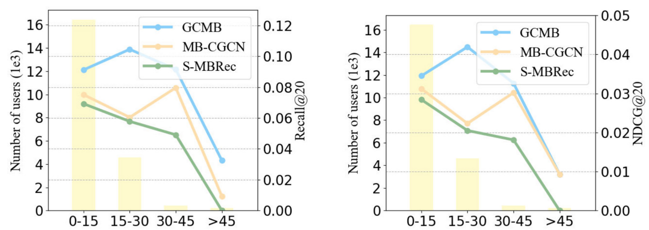

Fig. 5. The performance of sparse data on Beibei.

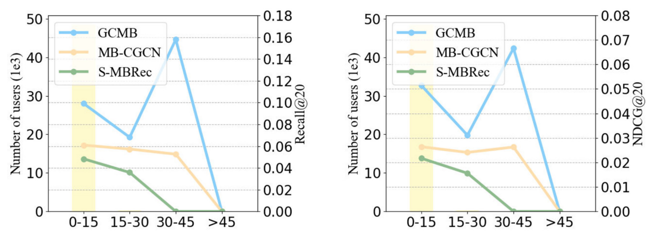

Fig. 6. The performance of sparse data on Taobao.

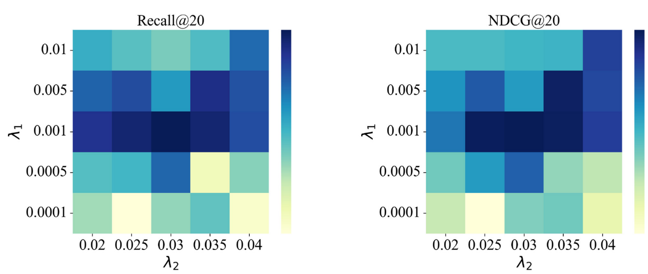

Fig. 7. Performance of GCMB with different λ_1 and λ_2 on Beibei.

E. Hyper-parameter Study

Firstly, we consider various values for λ_1 and λ_2 from the sets $\{0.0001, 0.0005, 0.001, 0.005, 0.01\}$ and $\{0.02, 0.025, 0.03, 0.035, 0.04\}$. Then, we evaluate the recommender performance for all combinations of λ_1 and λ_2 and present the results in the form of a heatmap. The darker the color, the better the corresponding recommender metrics. As can be seen in Figure 7 and Figure 8, when $\lambda_1 = 0.001$ and $\lambda_2 = 0.03$, the recommender performance is optimal.

Secondly, we analyze the effect of parameters τ on recommender performance. By observing Figure 9, we can find when τ is larger than 0.1, the recommender performance shows a downward trend. When $\tau = 0.1$ the model realizes the optimal state, so we can infer that $\tau = 0.1$ is the optimal parameter setting.

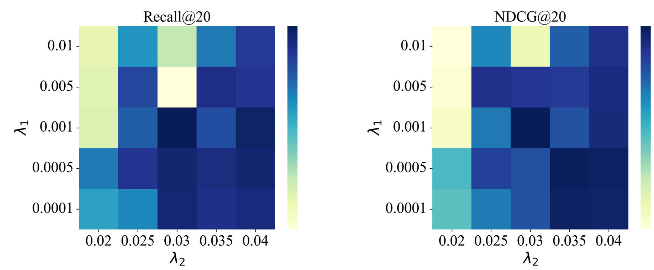

Fig. 8. Performance of GCMB with different λ_1 and λ_2 on Taobao.

Finally, we analyze the influence of different layers on the model. As can be seen in Tabel III, layer = 2 is the optimal choice for recommender.

F. Discussion

1) rSVD for Implicit Interaction: In this paper, we chose to use rSVD to inject implicit interactions, primarily due to its efficiency, scalability, and global signal extraction capability. rSVD reduces computational complexity through random sampling, making it suitable for large-scale datasets. It effectively extracts implicit collaborative signals, supplementing unobserved user-item interactions, and enhances user preference representation by combining explicit and implicit signals, improving recommendation accuracy and personalization.

Compared to traditional methods, rSVD is more efficient than GCN for large, sparse graphs, extracts a broader range of implicit signals than Item-based CF, and provides a lightweight alternative to Autoencoders, especially in data-sparse scenarios. Experiments on two real-world datasets show that rSVD consistently outperforms state-of-the-art recommendation methods, confirming its effectiveness and practicality.

However, rSVD also has limitations: random sampling may introduce uncertainty, affecting the stability of the results; the choice of the rank k and the number of samples significantly impacts performance and requires careful tuning.

2) Enhanced Comparative Analysis: We validate the proposed GCMB method in three ways. First, compared to matrix factorization-based methods [38] and graph-based methods [31], our method uses rSVD to inject implicit interactions, more comprehensively capturing user preferences. Second, unlike commonality analysis methods based on attention mechanisms [14] and meta-path-based methods [19], our multi-level commonality learning module captures commonalities at different levels through multi-layer aggregation and contrastive learning, providing finer-grained and higher-quality node representations. Finally, by incorporating recent

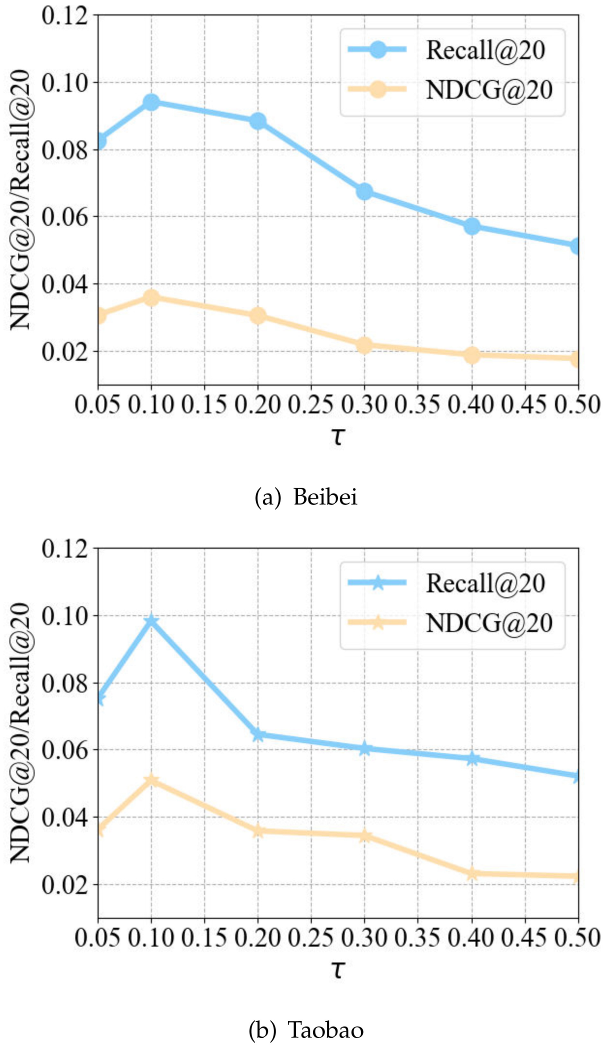

Fig. 9. The performance of different τ on Beibei and Taobao.

multi-behavior recommender system methods, such as MB-CGCN [16] and MBGCN [17], our experiments show that our method outperforms these approaches across multiple evaluation metrics, especially in terms of implicit interaction and commonality analysis.

3) Scalability Analysis: In this paper, the proposed GCMB method demonstrates excellent scalability for large-scale datasets. We address potential computational limitations and bottlenecks by using rSVD to inject implicit interactions. rSVD significantly reduces computational complexity to $O(n \cdot klogk)$, compared to the $O(n^3)$ of traditional SVD, and decreases memory usage through random sampling. It also supports batch processing and parallel computation, enhancing efficiency.

Experiments on two large real-world datasets validate the high efficiency and accuracy of our method. To mitigate potential bottlenecks in data preprocessing and hyperparameter tuning, we employ efficient preprocessing techniques and automatic hyperparameter tuning tools. Overall, the GCMB method exhibits strong scalability and effectively addresses computational limitations.

4) Application Analysis: In e-commerce, GCMB enhances user satisfaction and purchase conversion rates by providing precise personalized recommendations through the combination of explicit and implicit interaction signals. It also comprehensively understands user preferences via multi-behavior analysis. In social media platforms, GCMB optimizes content recommendations, increasing user engagement and retention, and uncovers potential connections and interest similarities among users through global collaborative signals. For travel service platforms, GCMB offers personalized travel routes and attraction recommendations, improving user experience, and predicts future user behavior through historical behavior analysis, enabling the platform to prepare resources and services in advance. In online video and music platforms, GCMB generates personalized playlists, enhancing user experience, and helps users discover more content of interest by capturing behavioral commonalities, thereby increasing user stickiness and activity. These real-world application cases highlight the broad applicability and significant benefits of the GCMB method across various domains.

V. Conclusion

In this paper, we propose a GCMB for the multi-behavior recommender. Specifically, to enhance the node representation of the original view, we devise a graph contrastive learning paradigm guided by rSVD instead of the traditional three-view paradigm. This approach effectively injects implicit collaborative relation learning into the multi-behavior recommender model. Furthermore, we improve the quality of the node representation by focusing on the multi-level commonality between target and auxiliary behaviors. Extensive experiments on two real-world datasets demonstrate that our method consistently outperforms various state-of-the-art recommender methods.

In future work, we will integrate knowledge graphs into the GCMB model to enhance recommendation performance. This includes extracting entities and relationships from public knowledge graphs (e.g., Wikidata and DBpedia), aligning them with user behavior data, and fusing the knowledge graph with the multi-behavior graph. Using GNNs, we will learn embeddings and incorporate contextual information, thereby improving the model's understanding of user preferences. This integration will significantly enhance feature representation, cold-start handling, contextual awareness, and interpretability.

References

[1] S. Wu, F. Sun, W. Zhang, X. Xie, and B. Cui, "Graph neural networks in recommender systems: A survey," *ACM Comput. Surv.*, vol. 55, no. 5, pp. 1–37, 2022.

[2] J. Wu, J. Zhang, M. Bilal, F. Han, N. Victor, and X. Xu, "A federated deep learning framework for privacy-preserving consumer electronics recommendations," *IEEE Trans. Consum. Electron.*, vol. 70, no. 1, pp. 2628–2638, Feb. 2024.

[3] A. Alzahrani and M. Z. Asghar, "Maintaining user security in consumer electronics-based online recommender systems using federated learning," *IEEE Trans. Consum. Electron.*, vol. 70, no. 1, pp. 2657–2665, Feb. 2024.

[4] M. Tariq, B. Tahir, and Z. Zhou, "Secure and enhanced online recommendations: A federated intelligence approach," *IEEE Trans. Consum. Electron.*, vol. 70, no. 1, pp. 2500–2507, Feb. 2024.

[5] A. Dadoun, R. Troncy, O. Ratier, and R. Petitti, "Location embeddings for next trip recommendation," in *Proc. Companion World Wide Web Conf.*, 2019, pp. 896–903.

[6] P. Mondal, P. Kapoor, S. Singh, S. Saha, J. P. Singh, and A. K. Singh, "Genre effect toward developing a multi-modal movie recommendation system in Indian setting," *IEEE Trans. Consum. Electron.*, vol. 70, no. 1, pp. 2517–2526, Feb. 2024.

[7] A. Ferraro, "Music cold-start and long-tail recommendation: Bias in deep representations," in *Proc. 13th ACM Conf. Recomm. Syst.*, 2019, pp. 586–590.

[8] O. A. Arqub and Z. Abo-Hammour, "Numerical solution of systems of second-order boundary value problems using continuous genetic algorithm," *Inf. Sci.*, vol. 279, pp. 396–415, Sep. 2014.

[9] H. Chai, X. Wei, H. Ma, and X. Jiang, "Knowledge-enhanced graph transformer network for multi-behavior and item-knowledge session-based recommendation," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, 2022, pp. 3421–3426.

[10] S. Gu, X. Wang, C. Shi, and D. Xiao, "Self-supervised graph neural networks for multi-behavior recommendation," in *Proc. Int. Joint Conf. Artif. Intell.*, 2022, pp. 2052–2058.

[11] L. Xia, C. Huang, Y. Xu, P. Dai, M. Lu, and L. Bo, "Multi-behavior enhanced recommendation with cross-interaction collaborative relation modeling," in *Proc. 37th IEEE Int. Conf. Data Eng. (ICDE)*, 2021, pp. 1931–1936.

[12] L. Xia, C. Huang, Y. Xu, P. Dai, and L. Bo, "Multi-behavior graph neural networks for recommender system," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 4, pp. 5473–5487, Apr. 2024.

[13] C. Gao et al., "Neural multi-task recommendation from multi-behavior data," in *Proc. IEEE 35th Int. Conf. Data Eng. (ICDE)*, 2019, pp. 1554–1557.

[14] C. Chen, M. Zhang, Y. Zhang, W. Ma, Y. Liu, and S. Ma, "Efficient heterogeneous collaborative filtering without negative sampling for recommendation," in *Proc. AAAI Conf. Artif. Intell.*, 2020, pp. 19–26.

[15] C. Chen et al., "Graph heterogeneous multi-relational recommendation," in *Proc. AAAI Conf. Artif. Intell.*, 2021, pp. 3958–3966.

[16] Z. Cheng, S. Han, F. Liu, L. Zhu, Z. Gao, and Y. Peng, "Multi-behavior recommendation with cascading graph convolution networks," in *Proc. ACM Web Conf.*, 2023, pp. 1181–1189.

[17] 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. Retr.*, 2020, pp. 659–668.

[18] Y. Ma, Z. Wang, M. Tang, W. Qu, M. Deveci, and D. Pamucar, "Personalized recommendation system based on meta-learning for human-centric consumer services in gig economy," *IEEE Trans. Consum. Electron.*, vol. 70, no. 1, pp. 1080–1091, Feb. 2024.

[19] Y. You, T. Chen, Y. Sui, T. Chen, Z. Wang, and Y. Shen, "Graph contrastive learning with augmentations," in *Proc. 34th Int. Conf. Neural Inf. Process. Syst.*, 2020, pp. 1–12.

[20] J. Wu et al., "Self-supervised graph learning for recommendation," in *Proc. 44th Int. ACM SIGIR Conf. Res. Dev. Inf. Retr.*, 2021, pp. 726–735.

[21] J. Yu, H. Yin, X. Xia, T. Chen, L. Cui, and Q. V. H. Nguyen, "Are graph augmentations necessary? Simple graph contrastive learning for recommendation," in *Proc. 45th Int. ACM SIGIR Conf. Res. Dev. Inf. Retr.*, 2022, pp. 1294–1303.

[22] M. Choi, J. Kim, J. Lee, H. Shim, and J. Lee, "S-Walk: Accurate and scalable session-based recommendation with random walks," in *Proc. 15th ACM Int. Conf. Web Search Data Min.*, 2022, pp. 150–160.

[23] B. Perozzi, R. Al-Rfou, and S. Skiena, "DeepWalk: Online learning of social representations," in *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, 2014, pp. 701–710.

[24] M. A. Dayeh, B. Sartawi, and S. Salah, "A bias-free time-aware PageRank algorithm for paper ranking in dynamic citation networks," *Intell. Inf. Manage.*, vol. 14, no. 2, pp. 53–70, 2022.

[25] X. Niu, G. Liu, and Q. Yang, "OpinionRank: Trustworthy website detection using three valued subjective logic," *IEEE Trans. Big Data*, vol. 8, no. 3, pp. 855–866, Jun. 2022.

[26] L. Peng et al., "GRLC: Graph representation learning with constraints," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 6, pp. 8609–8622, Jun. 2024.

[27] X. Luo et al., "Self-supervised graph-level representation learning with adversarial contrastive learning," *ACM Trans. Knowl. Discov. Data*, vol. 18, no. 2, pp. 1–23, 2023.

[28] H. Li, J. Cao, J. Zhu, Q. Luo, S. He, and X. Wang, "Augmentation-free graph contrastive learning of invariant-discriminative representations," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 8, pp. 11157–11167, Aug. 2024.

[29] C. Su, M. Chen, and X. Xie, "Graph convolutional matrix completion via relation reconstruction," in *Proc. 10th Int. Conf. Softw. Comput. Appl.*, 2021, pp. 51–56.

[30] C. Gao et al., "A survey of graph neural networks for recommender systems: Challenges, methods, and directions," *ACM Trans. Recomm. Syst.*, vol. 1, no. 1, pp. 1–51, 2023.

[31] X. He, K. Deng, X. Wang, Y. Li, Y. Zhang, and M. Wang, "LightGCN: Simplifying and powering graph convolution network for recommendation," in *Proc. 43rd Int. ACM SIGIR Conf. Res. Dev. Inf. Retr.*, 2020, pp. 639–648.

[32] M. Yan et al., "Cascading residual graph convolutional network for multi-behavior recommendation," *ACM Trans. Inf. Syst.*, vol. 42, no. 1, pp. 1–26, 2023.

[33] H. Yang, H. Chen, L. Li, P. S. Yu, and G. Xu, "Hyper meta-path contrastive learning for multi-behavior recommendation," in *Proc. IEEE Int. Conf. Data Min. (ICDM)*, 2021, pp. 787–796.

[34] 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 Min.*, 2022, pp. 1120–1128.

[35] R. Chen, D. Pu, Y. Tong, and M. Wu, "Image-denoising algorithm based on improved K-singular value decomposition and atom optimization," *CAAI Trans. Intell. Technol.*, vol. 7, no. 1, pp. 117–127, 2022.

[36] B. Zhang and M. Mascagni, "Pass-efficient randomized lu algorithms for computing low-rank matrix approximation," *Monte Carlo Methods Appl.*, vol. 29, no. 3, pp. 181–202, 2023.

[37] Z. Abo-Hammour, O. Alsmadi, S. Momani, and O. A. Arqub, "A genetic algorithm approach for prediction of linear dynamical systems," *Math. Problems Eng.*, vol. 2013, no. 1, 2013, Art. no. 831657. [Online]. Available: https://doi.org/10.1155/2013/831657

[38] B. Loni, R. Pagano, M. Larson, and A. Hanjalic, "Bayesian personalized ranking with multi-channel user feedback," in *Proc. 10th ACM Conf. Recomm. Syst.*, 2016, pp. 361–364.

[39] X. He, L. Liao, H. Zhang, L. Nie, X. Hu, and T.-S. Chua, "Neural collaborative filtering," in *Proc. 26th Int. Conf. World Wide Web*, 2017, pp. 173–182.

[40] X. Wang, X. He, M. Wang, F. Feng, and T.-S. Chua, "Neural graph collaborative filtering," in *Proc. 42nd Int. ACM SIGIR Conf. Res. Dev. Inf. Retr.*, 2019, pp. 165–174.

[41] X. Cai, C. Huang, L. Xia, and X. Ren, "LightGCL: Simple yet effective graph contrastive learning for recommendation," 2023, *arXiv:2302.08191*.

[42] W. Krichene and S. Rendle, "On sampled metrics for item recommendation," in *Proc. 26th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, 2020, pp. 1748–1757.

Zhidong Zhao (Member, IEEE) received the B.S. and M.S. degrees in mechanical engineering from the Nanjing University of Science and Technology, Nanjing, China, in 1998 and 2001, respectively, and the Ph.D. degree in biomedical engineering from Zhejiang University, Hangzhou, China, in 2004. He is currently a Full Professor with Hangzhou Dianzi University, Hangzhou. His research interests include biomedical signal processing, wireless sensor networks, biometrics, and machine learning.

Jianing Liu graduated from Hangzhou Dianzi University, Hangzhou, China. She is currently a Teacher with the Hebi Automobile Engineering Professional College, Hebi, Henan, China. Her current research interests include recommender system and network embedding.

Nana Huang received the Ph.D. degree in cyber science and engineering from Wuhan University, Wuhan, China, in 2023. She is currently a Lecturer with Hangzhou Dianzi University, Hangzhou, China. Her current research interests include data mining, user behavior pattern analysis, recommendation systems, and imbalanced learning.

Pengfei Jiao (Member, IEEE) received the Ph.D. degree in computer science from Tianjin University, Tianjin, China, in 2018, where he was a Lecturer with the Center of Biosafety Research and Strategy from 2018 to 2021. He is currently a Professor with the School of Cyberspace, Hangzhou Dianzi University, Hangzhou, China. His current research interests include complex network analysis and its applications.

Huaming Wu (Senior Member, IEEE) received the B.E. and M.S. degrees in electrical engineering from the Harbin Institute of Technology, China, in 2009 and 2011, respectively, and the Ph.D. degree (Hons.) in computer science from the Free University of Berlin, Germany, in 2015. He is currently a Professor with the Center for Applied Mathematics, Tianjin University, China. His research interests include wireless networks, mobile edge computing, the Internet of Things, and complex networks.