Applied Intelligence (2024) 54:12332–12346
https://doi.org/10.1007/s10489-024-05825-z

Multi-view denoising contrastive learning for bundle recommendation

Lei Sang[1] · Yang Hu[1] · Yi Zhang[1] · Yiwen Zhang[1]

Accepted: 28 August 2024 / Published online: 12 September 2024

Abstract

The goal of bundle recommendation is to offer users a set of items that match their preferences. Current methods mainly categorize user preferences into bundle and item levels, and then use graph neural networks to obtain representations of users and bundles at both levels. However, real-world interaction data often contains irrelevant and uninformative noise connections, leading to inaccurate representations of user interests and bundle content. In this paper, we introduce a **M**ulti-view **D**enoising **C**ontrastive **L**earning approach for **B**undle **R**ecommendation (**MDCLBR**), aiming to reduce the negative effects of noisy data on users' and bundles' representations. We use the original view, which includes bundle and item levels, to guide data augmentation for creating augmented views. Then, we apply the multi-view contrastive learning paradigm to enhance collaboration within the original view, the augmented views, and between them. This leads to more accurate representations of users and bundles, reducing the impact of noisy data. Our method outperforms previous approaches in extensive experiments on three real-world public datasets.

Keywords Bundle recommendation · Graph neural networks · Constrastive learning

1 Introduction

Bundle recommendation [1–3] centers around the recommendation of a group of items that share common themes to users. This concept is applied in various online platforms, including music streaming services and fashion e-commerce websites. By presenting item bundles instead of individual products, these platforms can significantly enhance the one-stop shopping experience for users. Furthermore, platforms that embrace bundles as part of their marketing strategy often witness increased sales revenue and attract users who appreciate bundle discounts. This trend leads to a growing preference for bundles, such as music playlists and fashion apparel, among both users and platform operators, as opposed to individual items like single songs or standalone clothing pieces. Consequently, the development of efficient bundle recommender systems attracts substantial attention from both academic researchers and industry professionals.

Modeling user preferences presents a central challenge in bundle recommender systems. After an extensive review of prior research [1–5] in the field, as illustrated in Fig. 1, we can analyze user preferences from two levels: (1) Through interactions between users and bundles, represented as the U-B graph, the focus is on users' preferences at the bundle level. (2) By leveraging interactions between users and individual items, depicted as the U-I graph, and incorporating knowledge about the associations between bundles and individual items, illustrated as the B-I graph, the emphasis is on understanding users' preferences at the item level. These two levels offer different angles for understanding users' interest preferences. Early studies [6, 7] predominantly concentrated on modeling user interest preferences by considering interactions between users and items, as well as among items within bundles. However, these studies only focused on the one-order connectivity information among users, bundles, and items.

Amidst the swift ascent of Graph Neural Networks (GNNs) [8, 9], most studies in bundle recommendation are dedicated to harnessing these networks for capturing high-

✉ Yiwen Zhang
zhangyiwen@ahu.edu.cn

Lei Sang
sanglei@ahu.edu.cn

Yang Hu
huyang@stu.ahu.edu.cn

Yi Zhang
zhangyi.ahu@gmail.com

[1] School of Computer Science and Technology, Anhui University, 111 Jiulong Rd, Hefei 230601, Anhui, China

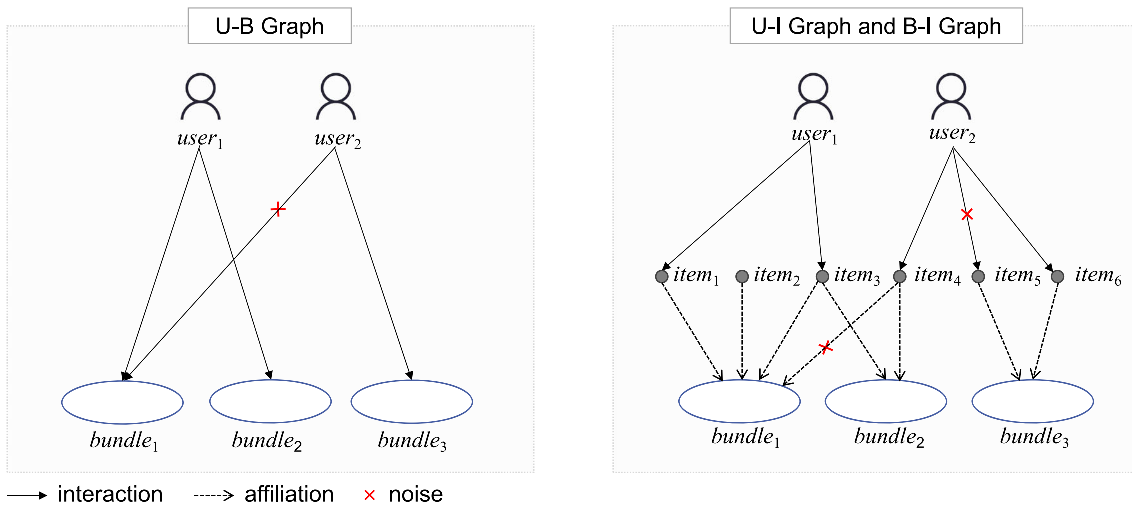


Fig. 1 The description of the user-item-bundle relationship

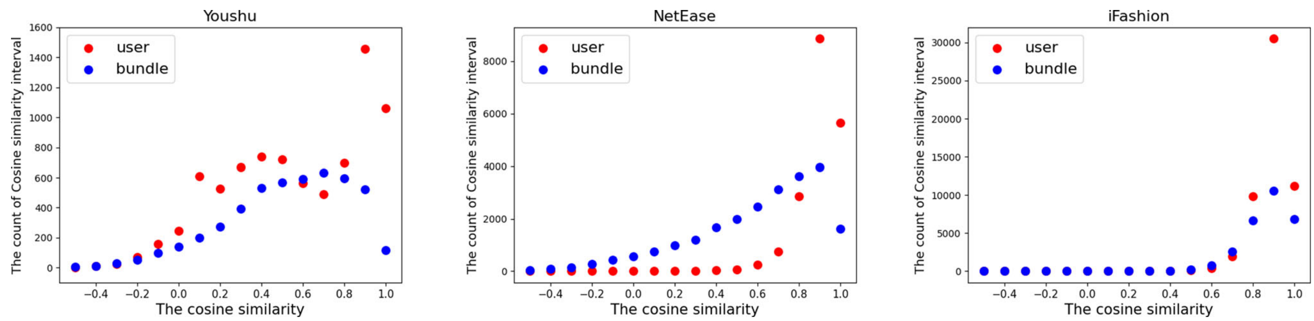
order neighborhood information, consequently enhancing the modeling of user behavior. BundleNet [5] considers the U-B graph, U-I graph, and B-I graph as a unified tripartite graph. It employs GNNs to aggregate neighborhood information and generate representations for users and bundles. Subsequently, BGCN [3] performs representation learning and preference prediction for the U-B graph, as well as combinations of U-I graph and B-I graph at both the bundle and item levels. However, it merely combines predictions related to each level without directly integrating collaborative relationships between levels into the optimized representation. This approach may not guarantee the capture of the complementarity between the two levels. In contrast, CrossCBR [10] leverages contrastive learning to enhance the complementary relationship between two levels, effectively capturing the collaborative relationships and improving the perceptual representations of each level.

Despite the promising performance, however, these GNN-based bundle recommendation methods are vulnerable to unnoticeable noise data attacks on graph.

As illustrated in Fig. 1, noisy data encompasses incorrect interactions between users and bundles or items, as well as irrelevant connections between bundles and items. This noisy data holds the potential to substantially influence the precision of level representation, potentially resulting in inaccurate recommendation outcomes. As GNNs propagate through the graph topology, manipulating a single node affects itself and its neighboring nodes, potentially causing a cascading effect. Therefore, we believe that compared to standard neural networks, GNNs are more susceptible to the influence of noisy data during the training process. Recent studies primarily try to use contrastive learning to enhance the quality of level representation for GNN-based model. However, they does not account for the impact of noisy data on representations during the representation learning stage. As presented in Fig. 2, we analyze the cosine similarity distributions of users' and bundles' representations at both the bundle level and item level in three real datasets [11–13] using the CrossCBR model to illustrate this concern.

Through the utilization of effective data augmentation strategies and integration into the contrastive learning paradigm, it becomes feasible to improve model robustness and more effectively mitigate the potential impact of noise. Most of the research incorporate a simple graph data augmentation technique. This method randomly removes a certain proportion of edges from the original graph based on a specified ratio parameter. The edges that are randomly removed still preserve the essential local structure of the graph. However, purely arbitrary deletion operations may lead to the removal of meaningful interaction records, limiting the effectiveness of contrastive learning in mitigating the effects of noise.

In this paper, We propose a Multi-view Denoising Contrastive Learning method for Bundle Recommendation (MDCLBR) to mitigate the adverse effects of noisy data. The fundamental process behind this approach is to treat U-B graph, U-I graph and B-I graph as original view. We employ graph neural network to capture users' and bundles' representations at both the bundle and item levels, evaluating the collaborative relationships between these levels using the cosine similarity of their representations. These collaborative relationships guide the data augmentation of the U-B graph, leading to the creation of the augmented view. Subsequently, we apply multi-view contrastive learning to capture collaborative relationships within the original view, the augmented view, and between them. Through continuous optimization of user and bundle representations, our aim is to mitigate the detrimental impact of noise in the data, ultimately achieving denoising effects. Thanks to multi-view denoising contrastive learning,



(a) Youshu (b) NetEase (c) iFashion

Fig. 2 The count of cosine similarity distribution intervals for users/bundles at both the bundle level and item level in the original view. (In the Youshu and NetEase datasets, the cosine similarity distributions of users' (bundles') representations at both levels show a loose trend and even include negative cosine similarity values, while in the iFashion dataset, the cosine similarity distribution trends for users' (bundles') representations are notably tight. High-quality interaction records will exhibit a cosine similarity distribution trend similar to that illustrated in iFashion dataset, while interaction records containing noisy data may lead to cosine similarity distribution trends similar to those depicted in the Youshu and NetEase datasets. Noisy data has negative effects on users' and bundles' representations)

MDCLBR exhibits superior performance compared to baselines across the three datasets.

Our main contributions are as follows:

- We propose a novel data augmentation method to mitigate the adverse impact of noisy data by identifying more information-rich interactions within the original view. This method first combines the collaborative relationships at the bundle and item levels in the original view, guiding the data augmentation of the U-B graph to form augmented views.
- We employ a multi-view contrastive learning paradigm to delve into more profound collaborative relationships within the original view, within the augmented views, and between the original and augmented views, further alleviating the adverse effects of noisy data.
- Our model outperforms state-of-the-art (SOTA) baselines on three datasets, showcasing the superiority of MDCLBR. Furthermore, we illustrate how the principles underlying MDCLBR can be extended to a broader range of tasks.

The structure of the remaining sections in this paper is outlined as follows: Section 2 provides an overview of the problem definition. Section 3 delves into the proposed approach, and Section 4 presents the experimental results. Section 5 discusses related work, while Section 6 concludes this work.

2 Problem definition

Given a set of user sets $\mathcal{U} = \{u_1, u_2, \ldots, u_M\}$, bundles sets $\mathcal{B} = \{b_1, b_2, \ldots, b_N\}$, item sets $\mathcal{I} = \{i_1, i_2, \ldots, i_O\}$, node sets $\mathcal{V} = \mathcal{U} \cup \mathcal{B}$, where M, N and O represent the number of users, bundles and items, respectively. We define $\mathbf{X}_{M \times N} = \{x_{ub} | u \in \mathcal{U}, b \in \mathcal{B}\}$, $\mathbf{Y}_{M \times O} = \{y_{ui} | u \in \mathcal{U}, i \in \mathcal{I}\}$ and $\mathbf{Z}_{N \times O} = \{z_{bi} | b \in \mathcal{B}, i \in \mathcal{I}\}$ respectively represent user-bundle interactions, user-item interactions and bundle-item interactions. We use binary variables x_{ub}, y_{ui} and z_{bi}, where a value of 1 indicates the interaction of user u with bundle b or item i, or the membership relationship of item i to bundle b. Our goal is to estimate the probability of interaction between user u with non-interacted bundle b based on available interaction records and membership relationships.

3 Methodology

In this section, we present a comprehensive overview of MDCLBR, with additional details illustrated in Fig. 3. MDCLBR is segmented into three stages: representation learning, data augmentation, and joint optimization. Subsequent sections will furnish comprehensive explanations for each of these stages.

3.1 Representation learning

We primarily use LightGCN as the GNN encoder to capture the representations of users and bundles at both the bundle level and item level in the original view.

A bipartite graph, denoted as U-B, is created at the bundle level based on the user-bundle interaction record matrix $\mathbf{X}$. LightGCN [14] is employed to learn representations of users and bundles in this graph. Notably, the method introduces information propagation on the user-bundle graph with

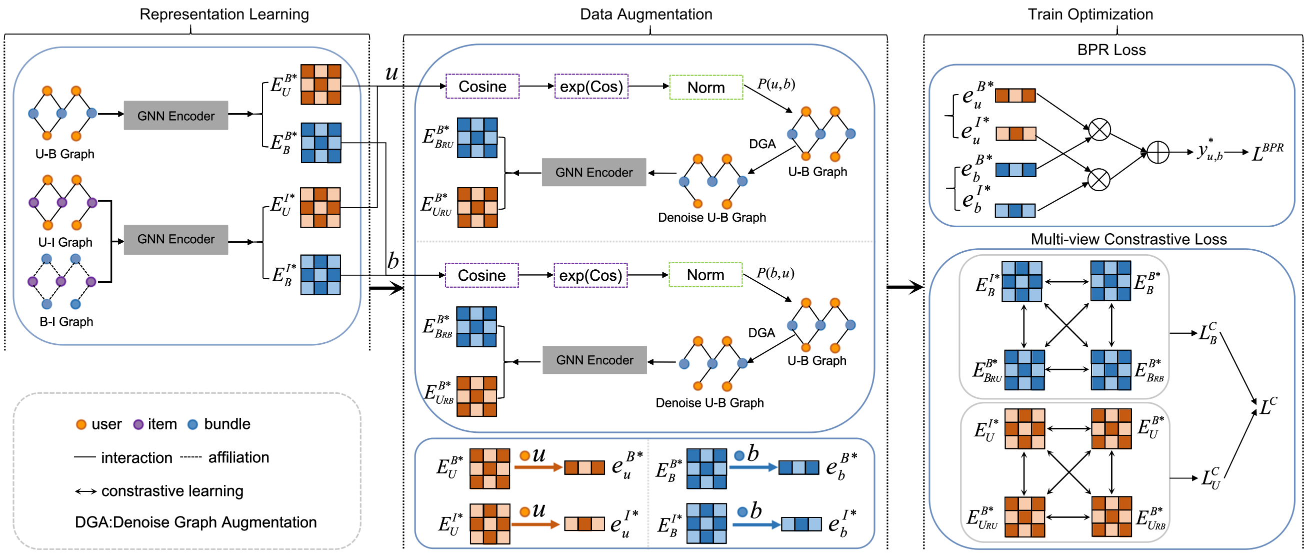


Fig. 3 The proposed MDCLBR model architecture. MDCLBR is divided into three key stages: representation learning, data augmentation, and training optimization. In the representation learning stage, the emphasis is on capturing representation of user and bundle at the bundle and item levels in the original view. Data augmentation stage includes augmenting the U-B graph utilizing collaborative signals between different levels in the original view and capturing the representations of the generated U-B graph after augmentation. Training optimization stage mainly focuses on the joint optimization of these representations using Bayesian personalized ranking loss and multi-view contrastive loss

k layers of information propagation defined as follows:

$$\mathbf{e}_u^{B(k)} = \sum_{b \in N_u} \frac{1}{\sqrt{|N_u|}\sqrt{|N_b|}} \mathbf{e}_b^{B(k-1)}, \quad \mathbf{e}_b^{B(k)} = \sum_{u \in N_b} \frac{1}{\sqrt{|N_b|}\sqrt{|N_u|}} \mathbf{e}_u^{B(k-1)}, \tag{1}$$

$$\mathbf{e}_u^{B*} = \sum_{k=0}^{K} \mathbf{e}_u^{B(k)}, \ \mathbf{e}_b^{B*} = \sum_{k=0}^{K} \mathbf{e}_b^{B(k)}, \tag{2}$$

where $\mathbf{e}_u^{B(k-1)}$ and $\mathbf{e}_b^{B(k-1)}$ represent the user u and bundle b reached by the information propagation in the (k-1)-th layer, N_u and N_b are used to describe the neighbors of user u and bundle b in U-B graph, $\mathbf{e}_u^{B(0)}$ and $\mathbf{e}_b^{B(0)}$ are randomly initialized vectors generated for user u and bundle b.

At the item level, based on user-item interactions $\mathbf{Y}$ and bundle-item affiliations in $\mathbf{Z}$, two bipartite graphs are created, denoted as U-I and B-I graphs. Similar to the learning process for the U-B graph, LightGCN is utilized to learn representations of users and items in U-I graph. The obtained users' representations from this graph are used as users' representations. Information propagation on the user-item graph is conducted, with k layers of information propagation defined as follows:

$$\mathbf{e}_u^{I(k)} = \sum_{i \in N_u} \frac{1}{\sqrt{|N_u|}\sqrt{|N_i|}} \mathbf{e}_i^{I(k-1)}, \quad \mathbf{e}_i^{I(k)} = \sum_{u \in N_i} \frac{1}{\sqrt{|N_i|}\sqrt{|N_u|}} \mathbf{e}_u^{I(k-1)}, \tag{3}$$

$$\mathbf{e}_u^{I*} = \sum_{k=0}^{K} \mathbf{e}_u^{I(k)}, \ \mathbf{e}_i^{I*} = \sum_{k=0}^{K} \mathbf{e}_i^{I(k)}, \tag{4}$$

where $\mathbf{e}_u^{B(k-1)}$ and $\mathbf{e}_i^{B(k-1)}$ represent the user u and item i reached by the information propagation in the (k-1)-th layer. N_u and N_i are used to describe the neighbors of user u and item i in U-I graph, and $\mathbf{e}_i^{I(0)}$ represents randomly initialized vectors generated for item i.

Then, items representations at the item level are obtained through guided bundle-item graph (B-I graph) by performing average pooling to acquire bundles' representations:

$$\mathbf{e}_b^{I*} = \frac{1}{|N_b|} \sum_{i \in N_b}^{K} \mathbf{e}_i^{I*}, \tag{5}$$

where $\mathbf{e}_i^{B(k-1)}$ represent the item i reached by the information propagation in the (k-1)-th layer, N_b is used to describe the number of items contained in bundle b in B-I graph. After conducting representation learning on the original view, we can obtain users' and bundles' representations $\mathbf{E}_U^{B*}, \mathbf{E}_U^{I*} \in R^{M\times d}$ and $\mathbf{E}_B^{B*}, \mathbf{E}_B^{I*} \in R^{N\times d}$ where d represents the embedding dimension.

3.2 Data augmentation module

After the representation learning phase, our data augmentation module leverages the collaborative relationships of user (bundle) representations to guide the augmentation of the U-B graph by fully exploiting the collaborative relationships represented by users (bundles) at both the bundle level and item level in the original view. The basic strategy of data augmentation module is to identify more information-rich interactions by capitalizing on the collaborative relationships of user (bundle) representations in the original view and extracting more valuable insights from the U-B graph.

Following the acquisition of representations of bundles from the original view using GNNs, we employ the bundles content consistency score to measure the degree of collaborative relationships among bundles in the original view. It is defined as:

$$c(b) = s(\mathbf{e}_b^{B*}, \mathbf{e}_b^{I*}). \tag{6}$$

Notably, as the consistency score of bundles content captured from substantial interaction relationships increases, the impact of noisy data on the bundles' representations perceived from the original view decreases. This enables more effective modeling of the authentic content of bundles. We incorporate the bundles content consistency score $c(b)$ into the probability of regenerating the user-bundle interaction graph of $p(b, u)$. The specific formula is as follows:

$$w(b) = \exp(c(b));\ p^*(b, u) = \max(\frac{w(b) - w^{\min}}{w^{\max} - w^{\min}}, p_{\tau b});$$
$$p(b, u) = \lambda_\mu \cdot \mu_{p^*} \cdot p^*(b, u), \tag{7}$$

where $p_{\tau b}$ is a threshold that governs the normalization of the minimum and maximum values of $\lambda(b)$, thus influencing the value of $p^*(b, u)$. μ_{p^*} represents the mean of $p^*(b, u)$ and λ_μ is a weight that controls the impact of μ_{p^*}.

Based on the users' representations acquired from GNNs in the original view, we utilize the user behavior consistency score to gauge the degree of collaborative relationships among user representations in the original view. The score is defined as follows:

$$c(u) = s(\mathbf{e}_u^{B*}, \mathbf{e}_u^{I*}). \tag{8}$$

Likewise, we include the user behavior consistency score $c(u)$ in the probability $p(u, b)$ of regenerating the user-bundle interaction graph under the control of $p_{\tau u}$. Then, based on the Bernoulli distribution, we use the sets $P(b, u)$, $P(u, b)$, generated by two probabilities $p(b, u)$, $p(u, b)$, to generate two mask matrixs $\mathbf{M}_b^1$ and $\mathbf{M}_u^2$, where the values are either 0 or 1. These mask matrixs are then applied to the user-bundle interaction $\mathbf{X}$, resulting in two augmented user-bundle interaction graph, and the process is defined as follows:

$$\mathbf{X}_b = \mathbf{M}_b^1 \odot \mathbf{X}, \mathbf{X}_u = \mathbf{M}_u^2 \odot \mathbf{X}. \tag{9}$$

Taking $c(b)$ as an example, a larger $c(b)$ suggests that the original view reflects more consistent user interests. The probability of noisy data in $\mathbf{X}_b$ generated by our proposed data augmentation module is lower. Therefore, the enhanced $\mathbf{X}_b$ contain more accurate user-bundle interaction records. Based on the U-B diagram of $\mathbf{X}_b$ and $\mathbf{X}_u$, we reuse the LightGCN recommendation framework on two augmented views $\mathcal{G}_b = (\mathcal{V}, \mathbf{X}_b)$ and $\mathcal{G}_u = (\mathcal{V}, \mathbf{X}_u)$ for the purpose of learning the augmented representations of users and bundles, denoted as $\mathbf{E}_{U_{RB}}^{B*}$, $\mathbf{E}_{B_{RB}}^{B*}$, $\mathbf{E}_{U_{RU}}^{B*}$ and $\mathbf{E}_{B_{RU}}^{B*}$ Respectively, where $\mathbf{E}_{U_{RB}}^{B*}, \mathbf{E}_{U_{RU}}^{B*} \in R^{M\times d}, \mathbf{E}_{B_{RB}}^{B*}, \mathbf{E}_{B_{RU}}^{B*} \in R^{N\times d}$.

3.3 Multi-view constrastive learning module

After the formation of augmented views in the data augmentation phase, we employ multi-view contrastive loss to reinforce the collaborative relationships within the original view, within the augmented views, and between the original view and augmented views, enhancing the accuracy of users'(bundles') representations and mitigating the adverse effects of noisy data.

To capture the collaborative relationships between users' and bundles' representations in the original view, we utilizes the InfoNCE [15] loss as the contrastive loss. This contrastive loss simultaneously promotes consistency among similar users (bundles) from different levels in the original view and enforces diversity among dissimilar users (bundles). The equations for the contrastive loss are as follows:

$$L_U^C = \frac{1}{|U|} \sum_{u\in U} -\log \frac{\exp(s(\mathbf{e}_u^{B*}, \mathbf{e}_u^{I*})/\tau)}{\sum_{v\in U} \exp(s(\mathbf{e}_u^{B*}, \mathbf{e}_v^{I*})/\tau)}, \tag{10}$$

$$L_B^C = \frac{1}{|B|} \sum_{b\in B} -\log \frac{\exp(s(\mathbf{e}_b^{B*}, \mathbf{e}_b^{I*})/\tau)}{\sum_{p\in B} \exp(s(\mathbf{e}_b^{B*}, \mathbf{e}_p^{I*})/\tau)}, \tag{11}$$

where $s(\mathbf{e}_1, \mathbf{e}_2)$ represents the cosine similarity between the vector $\mathbf{e}_1$ and $\mathbf{e}_2$, τ represents the temperature as a hyperparameter. By summing the two losses L_U^C and L_B^C, we can derive the original view contrastive loss: $L_{O-O}^C = L_U^C + L_B^C$.

Similarly, we establish the InfoNCE [15] loss within the two augmented views $\mathcal{G}_u$ and $\mathcal{G}_b$ to encourage the proximity of the same users(bundles) within the view and enforce separation of different users(bundles) across different views. The InfoNCE loss formula of users between the two augmented views $\mathcal{G}_u$ and $\mathcal{G}_b$ is as follows:

$$L_{U-R-R}^{C} = \frac{1}{|U|}\sum_{u\in U} -\log \frac{\exp(s(\mathbf{e}_{uru}^{B*}, \mathbf{e}_{urb}^{B*})/\tau)}{\sum_{v\in U}\exp(s(\mathbf{e}_{uru}^{B*}, \mathbf{e}_{vru}^{B*})/\tau)} + \frac{1}{|U|}\sum_{u\in U} -\log \frac{\exp(s(\mathbf{e}_{urb}^{B*}, \mathbf{e}_{uru}^{B*})/\tau)}{\sum_{v\in U}\exp(s(\mathbf{e}_{urb}^{B*}, \mathbf{e}_{vrb}^{B*})/\tau)}. \tag{12}$$

Equally, we obtain the InfoNCE loss of bundles L_{B-R-R}^{C}. Then, the summary of the contrastive loss function between the two augmented views $\mathcal{G}_u$ and $\mathcal{G}_b$ is as follows:

$$L_{R-R}^{C} = L_{U-R-R}^{C} + L_{B-R-R}^{C}. \tag{13}$$

We also use InfoNCE loss between the two augmented views $\mathcal{G}_u$, $\mathcal{G}_b$ and original view to further enhance the collaborative relationships between views. In this context, unlike traditional contrastive learning sampling methods, as depicted in (10), where $\mathbf{e}_u^{I*}$ and $\mathbf{e}_v^{I*}$ represent node embedding vector representations of different users at the bundle and item levels in the original view, while $\mathbf{e}_u^{B*}$ and $\mathbf{e}_{uru}^{B*}$ represent node embedding vector representations of same users within the distinct view. The InfoNCE loss formula for users between $\mathcal{G}_u$, $\mathcal{G}_b$ and at the bundle level in the original view is as follows:

$$L_{U-O-R}^{C} = \frac{1}{|U|}\sum_{u\in U} -\log \frac{\exp(s(\mathbf{e}_{u}^{B*}, \mathbf{e}_{uru}^{B*})/\tau)}{\sum_{v\in U}\exp(s(\mathbf{e}_{u}^{B*}, \mathbf{e}_{v}^{I*})/\tau)} + \frac{1}{|U|}\sum_{u\in U} -\log \frac{\exp(s(\mathbf{e}_{u}^{B*}, \mathbf{e}_{urb}^{B*})/\tau)}{\sum_{v\in U}\exp(s(\mathbf{e}_{u}^{B*}, \mathbf{e}_{v}^{I*})/\tau)}. \tag{14}$$

In the same way, the InfoNCE loss formula for users between the two augmented views and at the item level in the original view is as follows:

$$L_{U-O-R-I}^{C} = \frac{1}{|U|}\sum_{u\in U} -\log \frac{\exp(s(\mathbf{e}_{u}^{I*}, \mathbf{e}_{uru}^{B*})/\tau)}{\sum_{v\in U}\exp(s(\mathbf{e}_{u}^{I*}, \mathbf{e}_{v}^{B*})/\tau)} + \frac{1}{|U|}\sum_{u\in U} -\log \frac{\exp(s(\mathbf{e}_{u}^{I*}, \mathbf{e}_{urb}^{B*})/\tau)}{\sum_{v\in U}\exp(s(\mathbf{e}_{u}^{I*}, \mathbf{e}_{v}^{B*})/\tau)}. \tag{15}$$

Analogously, we derive the two InfoNCE loss of bundles L_{B-O-R}^{C} and $L_{B-O-R-I}^{C}$. The summary of the contrastive loss function between the original view and the two augmented views is as follows:

$$L_{O-R}^{C} = L_{U-O-R}^{C} + L_{B-O-R}^{C} + L_{U-O-R-I}^{C} + L_{B-O-R-I}^{C}. \tag{16}$$

3.4 Model train optimization

After the two stages mentioned above, we utilize Bayesian Personalized Ranking (BPR) [16] and contrastive loss functions are applied to optimize these representations. Following the representation learning phase, where users' and bundles' representations are obtained from the original view at the bundle and item levels, the final prediction is formulated as follows: $y_{u,b}^{*} = \mathbf{e}_u^{B*\,\mathrm{T}}\mathbf{e}_b^{B*} + \mathbf{e}_u^{I*\,\mathrm{T}}\mathbf{e}_b^{I*}$. The primary loss function employed is the BPR loss: $L^{BPR} = \sum_{(u,b,b')\in Q} -\ln\sigma(y_{u,b}^{*} - y_{u,b'}^{*})$. Completing the multi-view contrastive learning module to obtain individual contrastive losses, we combine them to form the joint loss function as follows:

$$L = L^{BPR} + \lambda_1 L_{O-O}^{C} + \lambda_2 L_{R-R}^{C} + \lambda_3 L_{O-R}^{C} + \lambda_4\|\theta\|_2^2, \tag{17}$$

In the above formulas, λ_1 , λ_2 and λ_3 are weights used to balance the multi-view contrastive losses, λ_4 is the parameter that controls the weight of regularization in the joint loss function, and $\theta = \{\mathbf{E}_U^{B(0)}, \mathbf{E}_B^{B(0)}, \mathbf{E}_I^{I(0)}\}$ represents all the model parameters.

3.5 Model discussion

In summary, our model is capable of mitigating the impact of noisy data and achieving denoising effects. Specifically, Data augmentation module enhances the U-B graph by leveraging the consistency scores $c(u)$ and $c(b)$ in the original view, resulting in two more accurate user-bundle interaction graphs, $\mathbf{X}_b$ and $\mathbf{X}_u$. We encode these augmented views $\mathcal{G}_u$ and $\mathcal{G}_b$ using graph neural networks to obtain more accurate representations. Subsequently, multi-view contrastive learning module further strengthens collaboration within the original view, augmented views, and between the original view and the augmented views. This encourages the original view's representations to be closer to the more accurate representations, reducing the impact of noisy data on the representations and achieving denoising effects.

More specifically, taking user u and user v as an example to illustrate the denoising capability of our proposed model. we have the following representations: $\mathbf{e}_u^{B*}$, $\mathbf{e}_u^{I*}$, $\mathbf{e}_{uru}^{B*}$ and $\mathbf{e}_{urb}^{B*}$ represent the representations of user u at both the bundle and item levels in the original view and the two augmented views $\mathcal{G}_u$ and $\mathcal{G}_b$ respectively. $\mathbf{e}_v^{B*}$, $\mathbf{e}_{vru}^{B*}$ and $\mathbf{e}_{vrb}^{B*}$ represent

Fig. 4 The impact of intra-view constrastive learning

(a) Original view contrastive learning (b) Augmented view contrastive learning

the representations of user v at bundle level in the original view and the two augmented views $\mathcal{G}_u$ and $\mathcal{G}_b$ respectively. Our goal is to bring the representations of different views' user u denoted as $(\mathbf{e}_u^{B*}, \mathbf{e}_u^{I*}, \mathbf{e}_{uru}^{B*}, \mathbf{e}_{urb}^{B*})$ closer to each other and separate the representations of user u denoted as $\mathbf{e}_u^{B*}$ and user v denoted as $(\mathbf{e}_v^{B*}, \mathbf{e}_{vru}^{B*}, \mathbf{e}_{vrb}^{B*})$ from each other through multi-view contrastive learning.

We provide detailed explanations on how the representations of the same user converge. As depicted in Fig. 4a, we use λ_1 to control and encourage the representations of the same users in the original view denoted as $(\mathbf{e}_u^{B*}, \mathbf{e}_u^{I*})$ to be closer to each other. On the other hand, as depicted in Fig. 4b, λ_2 is used to bring the representations of the user u denoted as $(\mathbf{e}_{uru}^{B*}, \mathbf{e}_{urb}^{B*})$ closer to each other. Finally, as illustrated in Fig. 5, with the assistance of λ_3, we aim to bring the representations of user u denoted as $\mathbf{e}_{uru}^{B*}$ and $\mathbf{e}_{urb}^{B*}$ closer to the representations of user u denoted as $\mathbf{e}_u^{B*}$ and $\mathbf{e}_u^{I*}$. One aspect involves converging the representations of user u in both the original and augmented views. Simultaneously, we emphasize maintaining consistency between user u's representation in the original view and the more precise representation in the augmented view, improving the accuracy of user u's representation and thus achieving denoising effects within the same user.

Similarly, we illustrate the divergence of representations among different users. As described in Fig. 5, under the control of λ_3, user u's representations $(\mathbf{e}_u^{B*}, \mathbf{e}_{uru}^{B*})$ become closer to each other. Additionally, under the supervision of λ_2, user u's representations $(\mathbf{e}_{uru}^{B*}, \mathbf{e}_{vru}^{B*})$ are separate, as depicted in Fig. 4b. User u's representations $(\mathbf{e}_u^{B*}, \mathbf{e}_{vru}^{B*})$ are also separate. Similarly, $(\mathbf{e}_u^{B*}, \mathbf{e}_{vrb}^{B*})$ are separate because user u's representations $\mathbf{e}_u^{B*}$ and user v representations $\mathbf{e}_v^{B*}$ are separate. This separation results in user u and v representing each other as distinct in all their views, further accentuating the distinctiveness of user v's representations from those of user u.

4 Experiment

In this section, we conduct experiments to evaluate our proposed method using three publicly available datasets: Youshu [11], NetEase [12] and iFashion [13]. Our objective is to address the following research questions:

- **RQ1**: Can our proposed model achieve state-of-the-art performance compared to existing baselines?
- **RQ2**: Are all the key components in our framework effective?
- **RQ3**: Can our model achieve denoising effects?
- **RQ4**: How does our model perform under different crucial parameters?

4.1 Experimental settings

4.1.1 Dataset description and evaluation metrics

We followed previous research and utilized three fundamental bundle datasets:

- **Youshu**[1] is a dataset focused on books, where bundles are represented as collections of book titles.
- **NetEase**[2] is a music-related dataset, where bundles are comprised of sets of individual song titles.
- **iFashion** [3] is a clothing-related dataset, where bundles are characterized by lists of individual clothing items.

To ensure a fair comparison, we keep the same data preprocessing procedures as in the previous work for all three datasets. The dataset descriptions are provided in Table 1. To evaluate the predictive performance of the models, we use Recall@K and NDCG@K [8, 9] as evaluation metrics where K $\in \{20, 40\}$.

4.1.2 Methods comparison

At the baseline stage, we select both the standard user-item (U-I) recommendation model and the specialized bundle recommendation model for comparison with our proposed method. The U-I recommendation model treats bundle sets as a special type of item and does not consider subsidiary items within the bundle set, relying solely on U-B records. We also choose the following state-of-the-art (sota) methods:

[1] http://www.yousuu.com
[2] https://music.163.com
[3] https://github.com/mysbupt/CrossCBR/blob/master/dataset.tgz

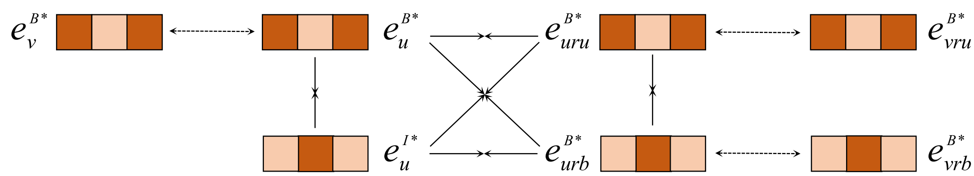


Fig. 5 The impact of multi-view constrastive learning

MFBPR [16]: A matrix factorization optimization method based on Bayesian personalized ranking loss.

LightGCN [14]: A recommendation model based on collaborative filtering and graph neural networks (GNNs). It utilizes a lightweight version of the graph learning kernel.

SGL [17]: By integrating contrastive graph learning into LightGCN, it achieves advanced performance.

The specialized bundle recommendation method involves the design of bundle recommendation and leverages all U-B interactions, U-I interactions and B-I auxiliary data. We consider the following methods:

DAM [11]: It leverages an attention mechanism on subsidiary items to acquire bundle representations and employs multi-task learning for optimizing U-B and U-I interactions.

BundleNet [5]: This approach constructs a tripartite U-B-I graph and employs GCN for representation learning, integrating multi-task learning.

BGCN [3, 4]: It decomposes the U-B-I relationship graph into two independent levels, including the bundle level and item level. It utilizes GCN to obtain representations for these two levels and achieves advanced performance by aggregating predictions from both levels.

CrossCBR [10]: Similar to BGCN, this approach decomposes the U-B-I relationship graph into two independent levels: the bundle level and the item level. By employing GCN for representation learning across these two levels, it further reinforces collaborative relationships between levels through contrastive learning. Ultimately, it achieves state-of-the-art performance by aggregating predictions from both levels.

Some earlier works are not considered in this comparison because they do not demonstrate performance as strong as the methods listed above.

4.1.3 Hyper-parameter and experimental environment

To ensure fairness, we utilize Xavier normal initialization for model initialization. The model is optimized using the Adam optimizer and the learning rates for Youshu, NetEase and iFashion are 0.001. The embedding size is 64, the batch size is 2048, the number of layers in LightGCN, the temperature and data augmentation parameters remain consistent with the settings used in previous research. However, We employ the grid search strategy to determine the parameters $(P_{\tau u}, P_{\tau b})$ and λ_3 with their predefined ranges across all three datasets as {(0.22, 0.21), (0.23, 0.24), (0.26, 0.28), (0.30, 0.32), (0.32, 0.34)},{0.005, 0.01, 0.015, 0.02, 0.025}, {(0.39, 0.22), (0. 40, 0.24), (0.44, 0.26), (0.50, 0.30), (0.58, 0.34)},{0.100, 0 .15, 0.200, 0.25, 0.300}, {(0.37, 0.28), (0.41, 0.33), (0.44, 0.40), (0.50, 0.41), (0.57, 0.49)},{0.200, 0.30, 0.400, 0.50, 0.600}.

All models are trained on PyTorch 1.7 and NVIDIA RTX 3090 GPUs. Each model completes 5 times, and and the evaluation results are based on the averages of these runs.

4.2 Performance larison (RQ1)

Tables 2 and 3 provides an overview of the overall performance of various methods. The best-performing methods

Table 1 Statistics of three experimental datasets

DataSet	Youshu	NetEase	iFashion
$\#U$	8,039	18,528	53,897
$\#B$	4,771	22,864	42,563
$\#I$	32,770	12,3628	27,694
$\#U-B$	51,377	303,303	1,679,708
$\#U-I$	138,515	1,128,065	2,290,645
$\#B-I$	176,667	1,778,838	164,293
$\#Avg$ U-B interactions	6.39	16.32	31.17
$\#Avg$ U-I interactions	17.23	42.50	42.50
$\#Avg$ B-I relations	37.03	77.80	3.86

Table 2 Significant improvements in experimental results on Top-20 predictions compared to other baselines for MDCLBR

Method	Youshu		NetEase		iFashion	
	Rec@20	NDCG@20	Rec@20	NDCG@20	Rec@20	NDCG@20
MFBPR	0.1959	0.1117	0.0355	0.0181	0.0752	0.0542
LightGCN	0.2286	0.1344	0.0496	0.0254	0.0837	0.0612
SGL	0.2568	0.1527	0.0687	0.0368	0.0933	0.069
DAM	0.2082	0.1198	0.0411	0.021	0.0629	0.045
BundleNet	0.1895	0.1125	0.0391	0.0201	0.0626	0.0447
BGCN	0.2347	0.1345	0.0491	0.0258	0.0733	0.0531
CrossCBR	0.2813	0.1668	0.0842	0.0457	0.1173	**0.0895**
MDCLBR	**0.2855**	**0.1696**	**0.0887**	**0.0476**	**0.1174**	0.0888
Improve%	**1.49**	**1.68**	**5.34**	**4.57**	**0.08**	-0.78

$(p - \text{value} << 0.01)$

are indicated in bold, while the second-best performance is underlined. Several observations can be made:

For general user-item (U-I) recommendation tasks, LightGCN consistently outperforms MFBPR, highlighting the strong capabilities of graph neural network-based methods in U-I recommendation. SGL consistently outperforms LightGCN, indicating the effectiveness of contrastive learning in constraining the U-I bipartite graph. In bundle recommendation tasks, BundleNet merges two levels into a ternary graph but inadequately considers collaboration between blurry levels, leading to suboptimal performance. In contrast, BGCN models user behavior as two separate levels, leading to improved performance. Building upon BGCN, CrossCBR leverages contrastive learning to establish levels' cooperative relationships, resulting in significant performance improvements. As shown in Tables 2 and 3, our method demonstrates particularly notable performance on the Youshu and NetEase datasets, achieving state-of-the-art results on specific segments of the iFashion dataset. This phenomenon is intriguing because, as depicted in Fig. 2, the Youshu and NetEase datasets contain more noise, while iFashion exhibits relatively less noise. This serves as evidence of the robust denoising capability inherent in our proposed method. In the subsequent sections, we will conduct a more in-depth discussion to further validate the effectiveness of the introduced key modules.

4.3 Ablation study (RQ2)

We conduct a series of ablation experiments to demonstrate the effectiveness of the key modules proposed, and the results are presented in Table 4. In the table, the term "improve" indicates the performance enhancement achieved by our proposed modules. We select the YouShu and NetEase datasets for these experiments due to their varying noise levels and used evaluation metrics like Recall@20 and NDCG@20, as shown in Fig. 2.

To validate the effectiveness of the data augmentation module we introduce, we utilize the core idea of SGL to randomly generate two augmented views, referred to as MDCLBR random. The results in Table 4 indicate a slight performance improvement compared to CrossCBR on the YouShu dataset. However, on the NetEase dataset, MDCLBR random exhibits a slight decrease in performance compared

Table 3 Significant improvements in experimental results on Top-40 predictions compared to other baselines for MDCLBR

Method	Youshu		NetEase		iFashion	
	Rec@40	NDCG@40	Rec@40	NDCG@40	Rec@40	NDCG@40
MFBPR	0.2735	0.132	0.06	0.0246	0.1162	0.0687
LightGCN	0.319	0.1592	0.0795	0.0334	0.1284	0.077
SGL	0.3537	0.179	0.1058	0.0467	0.1389	0.0851
DAM	0.289	0.1418	0.069	0.0281	0.0995	0.0579
BundleNet	0.2675	0.1335	0.0661	0.0271	0.0A986	0.0574
BGCN	0.3248	0.1593	0.0829	0.0346	0.1128	0.0671
CrossCBR	0.3785	0.1938	0.1264	0.0569	0.1699	**0.1080**
MDCLBR	**0.3867**	**0.1976**	**0.1342**	**0.0595**	**0.1708**	0.1076
Improve%	**2.17**	**1.96**	**6.17**	**4.57**	**0.53**	-0.37

(p-value< <0.01)

Table 4 The results of the ablation experiments

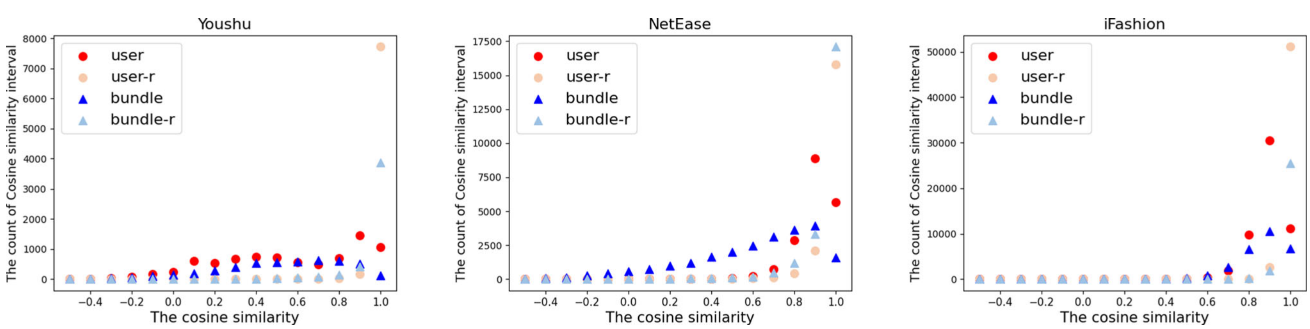
Method	Youshu		NetEase	
	Rec@20	NDCG@20	Rec@20	NDCG@20
CrossCBR	0.2813	0.1668	0.0842	0.0457
MDCLBR	0.2855	0.1696	0.0887	0.0476
MDCLBR random	0.2842	0.1687	0.0836	0.0442
$Improve\%$	1.03	1.14	−0.71	−3.28
$-improve\%$	0.46	0.53	5.75	7.44
MDCLBR w/o L_{O-O}^{C}	0.2831	0.1687	0.0850	0.0461
$Improve\%$	0.64	1.14	0.95	0.88
$-improve\%$	0.84	0.53	4.17	3.15
MDCLBR w/o L_{R-R}^{C}	0.2832	0.1675	0.0879	0.0468
$Improve\%$	0.68	0.42	4.39	2.41
$-improve\%$	0.81	1.24	0.90	1.68
$Improve\%$	1.49	1.68	5.34	4.16

to CrossCBR. This interesting phenomenon can be attributed to the fact that, as depicted in Fig. 2, the NetEase dataset contains a lower proportion of noisy data compared to the YouShu dataset. The augmented views generated using the random method, constrained by the loss function L_{O-R}^{C}, disrupts the accurate representations of bundles and users in the original view, resulting in a reduction in the model's performance. In both the Youshu and NetEase datasets, the final performance show a decline when compared to MDCLBR, demonstrating the effectiveness of the data augmentation approach we propose.

To further validate the effectiveness of our proposed multi-view contrastive learning module while ensuring the effectiveness of our data augmentation module, we exclude the loss functions L_{O-O}^{C} and L_{R-R}^{C} from all loss functions. This is denoted as MDCLBR w/o L_{O-O}^{C} and MDCLBR w/o L_{R-R}^{C}. As shown in Table 4, for both the Youshu and NetEase datasets, we observe a slight decrease in performance relative to MDCLBR. Nevertheless, it still outperformes CrossCBR, demonstrating the necessity of loss functions L_{O-O}^{C} and L_{R-R}^{C} and thereby validating the effectiveness of our proposed multi-view contrastive learning approach.

4.4 Model denoising effect study (RQ3)

To demonstrate the model's denoising effectiveness, we delve into the collaborative relationships between user and bundle representations across various views, using user representations as an illustrative example. In Fig. 6, the red circular dots represent the cosine similarity between user representations at both the bundle and item levels in the original view, denoted as user. The light red circular dots, denoted as user-r, represent the cosine similarity of user representations between the bundle level in the original view and the augmented view. As depicted in Fig. 6, even when our proposed model achieves its best performance, we can observe that, in comparison to the cosine similarity scores of user, the cosine similarity scores of user-r are generally higher and more tightly distributed. Compared to the user representations at the item level in the original view, user representations at the bundle



(a) Youshu (b) NetEase (c) iFashion

Fig. 6 The count of cosine similarity distribution intervals for users/bundles under both the original view and the augmented view level in the original view are closer to user representations of the augmented view. This indicates that our model is capable of capturing more precise representations and is effective in alleviating the impact of noise.

4.5 Hyper-parameter setting (RQ4)

The MDCLBR framework we propose introduces several important hyperparameters that can significantly affect the model's training process. In this section, we explore the impact of these parameters and provide optimal configurations for other hyperparameters.

4.5.1 The impact of hyper-parameter $(P_{\tau u}, P_{\tau b})$

The data augmentation module involves a hyperparameter P_τ. To illustrate this, let's consider the behavior consistency score of users as an example. The parameter $P_{\tau u}$ represents the score threshold between the bundle-level user representations $\mathbf{e}_u^{B*}$ and the item-level user representations $\mathbf{e}_u^{I*}$ for user u in the original view. If $P_{\tau u}$ is set to a larger value, it results in a larger score threshold. However, this may lead to the generation of two augmented user-bundle interactions $\mathbf{X}_b$ and $\mathbf{X}_u$ with fewer high-quality records, which cannot effectively capture the user's genuine behavior interests. On the other hand, setting $P_{\tau u}$ to a smaller value leads to a score threshold, potentially resulting in two augmented user-bundle interactions $\mathbf{X}_b$ and $\mathbf{X}_u$ with many noisy records, which also may not effectively represent user's genuine behavior interests. Therefore, selecting an appropriate value for $P_{\tau u}$ is crucial.

Similarly, setting an appropriate value for $P_{\tau b}$ is equally important in the aspect of the bundles content consistency score. Our approach involves selecting the top$N$ where $N \in \{10\%, 20\%, 30\%, 40\%, 50\%\}$. For example, as introduced in Section 4.1.3, the values of $P_{\tau u}$ and $P_{\tau b}$ are 0.22 and 0.21. They describe the values of $w(b)$ and $w(u)$ in (7) after min-max normalization, corresponding to the top 50% in the Youshu dataset. The values of N corresponding to $P_{\tau u}$ and $P_{\tau b}$ are described in a similar manner across the three real datasets. The values of $P_{\tau u}$ and $P_{\tau b}$ are obtained by calculating the median after multiple training sessions. Figure 7 illustrates the impact of different $(P_{\tau u}, P_{\tau b})$ hyperparameter values on the model's performance. When the top$N$ value of $P_{\tau u}$ and $P_{\tau b}$ is set to $N = 30\%$, the model achieves the best performance in Youshu, NetEase and iFashion.

4.5.2 The impact of hyper-parameter λ_3

The multi-view contrastive learning module involves three contrastive loss weights: λ_1, λ_2 and λ_3. Among these, the λ_3 parameter controls the alignment between the original bundle level user representations $\mathbf{e}_u^{B*}$ for user u and the augmented view user representations $\mathbf{e}_{uru}^{B*}$ for user u. Setting λ_3 to a higher value increases the alignment between $\mathbf{e}_u^{B*}$ and $\mathbf{e}_{uru}^{B*}$ but may neglect the collaboration between e_u^{B*} and e_u^{I*} at different levels in the original view. Conversely, a smaller λ_3 value diminishes the alignment between $\mathbf{e}_u^{B*}$ and $\mathbf{e}_{uru}^{B*}$, which may not effectively mitigate the noise issue. Therefore, selecting an appropriate value for λ_3 is crucial. Figure 8 illustrates the impact of different λ_3 hyperparameter values on the model's performance. The best performance in Youshu, NetEase, and iFashion is achieved when the value of λ_3 is set to $\{0.015, 0.15, 0.4\}$.

In summary, our model not only aligns representations within the same user for improved accuracy but also effectively separates representations between different users, demonstrating robust denoising capabilities that contribute to the overall effectiveness of the proposed approach.

5 Related work

In this section, we provide a brief review of the relevant work from three perspectives: bundle recommendation, graph-based denoising recommendation and contrastive learning in recommendation.

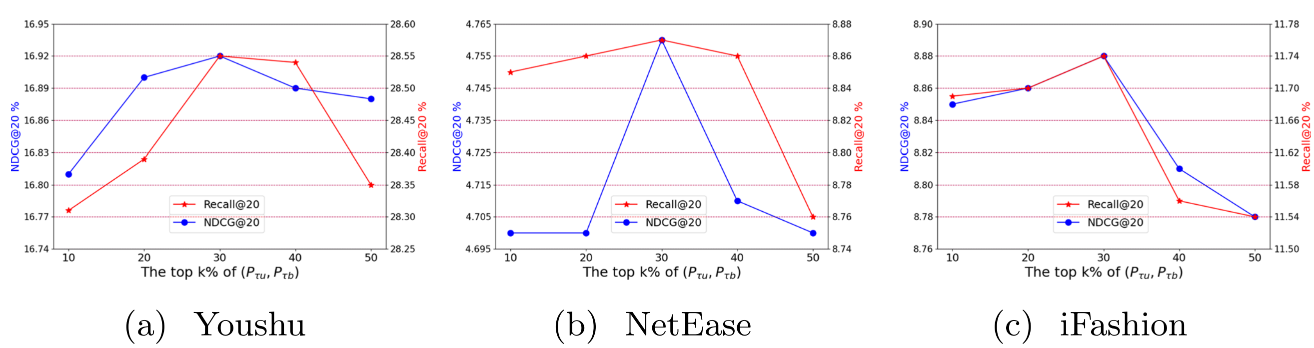
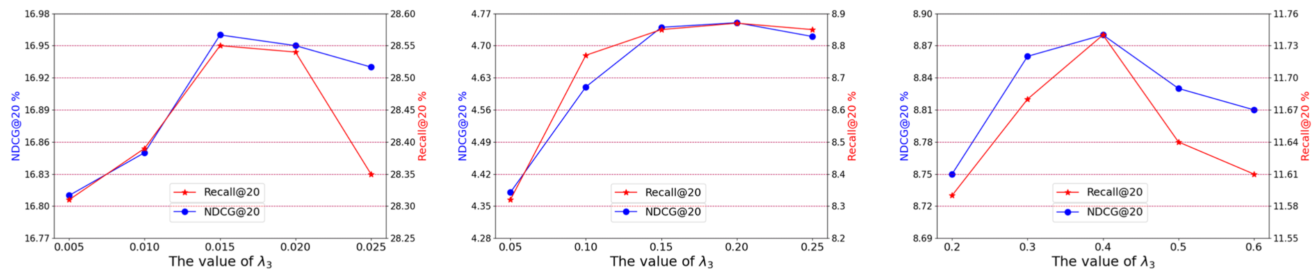


Fig. 7 The impact of hyper-parameter $(P_{\tau u}, P_{\tau b})$



(a) Youshu (b) NetEase (c) iFashion

Fig. 8 The impact of hyper-parameter λ_3

5.1 Bundle recommendation

Bundle recommendation addresses a specific recommendation scenario where recommendations are composed of items related to a particular theme, forming a bundle package. Early works often overlook the accompanying items within bundles, representing a bundle with a single ID. Subsequent works acknowledge the significance of associated items and develop diverse models by leveraging user-item interactions and bundle-item relationships, such as EFM [12] and DAM [11]. With the advent of GNN-based recommendation models, Deng proposes BundleNet [5] and Chang introduces BGCN [3]. While BundleNet integrates three types of relationships among users, bundles and items, BGCN effectively models user behavior at both the bundle and item levels, capturing diverse behaviors at each level and consequently improving overall performance. CrossCBR [10] leverages contrastive learning to capture collaborative relationships between two levels, enhancing the perceptual representations of each level and strengthening the complementary relationship between them.

Nevertheless, given the susceptibility of graph neural networks to noisy data, they often exhibit a tendency toward overfitting, with previous studies neglecting to consider the adverse effects of noisy data on these networks.

5.2 Graph-based Denoising recommendation

As denoising techniques in GNNs advance [18–20], there is a growing interest in addressing noise in graph-based collaborative filtering. SGCN [21] utilizes random binary masks, whereas HAR-GCF [22] employs attention mechanisms to mitigate interference caused by noise. DGNN [23] employs a combination of the ReLU function and the swin ReLU function to mitigate noise interference. SEDGN [24] extracts typical user behavioral patterns from both sequence and graph structures to reduce the impact of noise in session recommendations. KDGNN [25] integrates personalized and knowledge-aware signals to comprehensively capture user preferences, employing personalized knowledge-aware attention to clean noisy user-item interaction data.

Certainly, the effective integration of robust data augmentation strategies into the contrastive learning paradigm can effectively mitigate the potential impact of noise on the model.

5.3 Contrastive learning in recommendation

Recently, contrastive learning techniques achieves success in the fields of computer vision [26, 27] and natural language processing [28, 29], such as image generation and enhancement, as well as text generation. The recommendation community also embraces the application of contrastive learning in various scenarios, such as general collaborative filtering [17, 30–32], knowledge graphs [33–36] and multi-behavior recommendation [37–41], sequential [42–45] and social recommendation [46–49] , among others. The key to the success of contrastive learning in recommender systems lies in how to construct contrast pairs correctly. A branch of current methods involves creating more views through different data augmentation techniques. For instance, SGL [17] employs various graph augmentation methods (e.g., edge deletion or random walks). KGCL [33] introduces a knowledge graph to effectively guide data augmentation in user-item interactions. LightGCL [50] enhances user-item interactions through data augmentation using singular value decomposition. Another research approach involves mining multiple views present in the data and using multi-view contrastive learning to capture the supervised collaborative among these views. MCCLK [34] is based on contrastive learning and combines multiple graph views, including structural, collaborative and semantic views to obtain additional collaborative signals. CML [37] models multiple behavioral relationships between users and items and employs a multi-view contrastive learning approach to capture collaborative signals among these behaviors. In this study, we utilize the consistency score of bundles content and users behavior obtained from constructing two views from different data sources to guide data augmentation in user interactions with bundles. Subsequently, we capture collaborative signals between views through multi-view contrastive learning to achieve the goal of reducing the impact of noise.

6 Conclusion

In this work, we employ multi-view denoising contrastive learning to address the noise issues in bundle recommendation. By harnessing the collaborative relationships in the original view across different levels for data augmentation, creating augmented views and applying multi-view contrastive learning methods, we successfully capture collaboration across multi-views, consequently achieving effective denoising effects. Consequently, we present an effective and efficient multi-view denoising contrastive learning approach for bundle recommendation. Compared to the latest methods, our approach achieves better performance on all three datasets.

In future work, we have several key objectives. Firstly, our objective is to incorporate additional positive samples into the contrastive learning framework. For instance, we consider bundles with a higher number of shared items as positive samples. Secondly, we plan to extend the application of multi-view denoising contrastive learning to other similar application scenarios, provided that there are two or more perspectives available. Thirdly, considering the increased model complexity in this work due to generating two augmented views, we explore simpler and more efficient augmented views methods.

Acknowledgements This work is supported by the National Natural Science Foundation of China (No. 62272001 and No. 62206002), Anhui Provincial Natural Scielnce Foundation (2208085QF195), and Hefei Key Common Technology Project (GJ2022GX15).

Declarations

We would like to submit the manuscript titled "Multi-view Denoising Contrastive Learning for Bundle Recommendation" as a complete article for publication in Applied Intelligence.
(1)We would like to confirm that the paper is not currently under submission at any other journal or conference. None of the co-authors listed in this submission serve as Editors for Applied Intelligence. We have ensured that there is no conflict of interest in this regard. There is no conflict of interest that could compromise the fairness and integrity of the review process.
(2)The specific contributions of each author are outlined as follows: (a) Lei Sang: conceptualization, methodology, software development, review, writing-review. (b) Yang Hu: conceptualization, methodology, validation, software utilization, original draft preparation, writing-review. (c) Yi Zhang: methodology, formal analysis, review, writing-review. (d) Yiwen Zhang: conceptualization, methodology, formal analysis, supervision, writing, review, editing, resource management, project oversight, funding acquisition.
(3)The manuscript is approved by all authors for publication.
(4)This research does not involve any Human or Animals Participants.
(5)Authentic datasets, and ethical approval, participant consent, and publication consent are not applicable.
(6)The data for our manuscript is sourced from open.
Youshu
http://www.yousuu.com
https://github.com/mysbupt/CrossCBR/blob/master/dataset.tgz
NetEase
https://music.163.com
https://github.com/mysbupt/CrossCBR/blob/master/dataset.tgz
iFashion
https://tianchi.aliyun.com/dataset/137347
https://github.com/mysbupt/CrossCBR/blob/master/dataset.tgz

References

1. Bai J, Zhou C, Song J et al (2019) Personalized bundle list recommendation. In: The world wide web conference, pp 60–71. https://doi.org/10.1145/3308558.3313568
2. Zhu T, Harrington P, Li J et al (2014) Bundle recommendation in ecommerce. In: Proceedings of the 37th international ACM SIGIR conference on research & development in information retrieval, pp 657–666. https://doi.org/10.1145/2600428.2609603
3. Chang J, Gao C, He X et al (2020) Bundle recommendation with graph convolutional networks. In: Proceedings of the 43rd international ACM SIGIR conference on research and development in information retrieval, pp 1673–1676. https://doi.org/10.1145/3397271.3401198
4. Chang J, Gao C, He X et al (2021) Bundle recommendation and generation with graph neural networks. IEEE Trans Knowl Data Eng 35(3):2326–2340. https://doi.org/10.1109/TKDE.2021.3114586
5. Deng Q, Wang K, Zhao M et al (2020) Personalized bundle recommendation in online games. In: Proceedings of the 29th ACM international conference on information & knowledge management, pp 2381–2388. https://doi.org/10.1145/3340531.3412734
6. Liu G, Fu Y, Chen G et al (2017) Modeling buying motives for personalized product bundle recommendation. ACM Tran Knowl Discov Data (TKDD) 11(3):1–26. https://doi.org/10.1145/3022185
7. Liu Y, Xie M, Lakshmanan LV (2014) Recommending user generated item lists. In: Proceedings of the 8th ACM conference on recommender systems, pp 185–192. https://doi.org/10.1145/2645710.2645750
8. He X, Liao L, Zhang H et al (2017) Neural collaborative filtering. In: Proceedings of the 26th international conference on world wide web, pp 173–182. https://doi.org/10.1145/3038912.3052569
9. Wang X, He X, Wang M et al (2019) Neural graph collaborative filtering. In: Proceedings of the 42nd international ACM SIGIR conference on research and development in information retrieval, pp 165–174. https://doi.org/10.1145/3331184.3331267
10. Ma Y, He Y, Zhang A et al (2022) Crosscbr: cross-view contrastive learning for bundle recommendation. In: Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining, pp 1233–1241. https://doi.org/10.1145/3534678.3539229
11. Chen L, Liu Y, He X et al (2019) Matching user with item set: collaborative bundle recommendation with deep attention work. In: IJCAI, pp 2095–2101. https://doi.org/10.24963/IJCAI.2019/290
12. Cao D, Nie L, He X, et al (2017) Embedding factorization models for jointly recommending items and user generated lists. In: Proceedings of the 40th international ACM SIGIR conference on research and development in information retrieval, pp 585–594. https://doi.org/10.1145/3077136.3080779
13. Chen W, Huang P, Xu J et al (2019) Pog: personalized outfit generation for fashion recommendation at alibaba ifashion. In: Pro-

ceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining, pp 2662–2670. https://doi.org/10.1145/3292500.3330652
14. He X, Deng K, Wang X et al (2020) Lightgcn:simplifying and powering graph convolution network for recommendation. In: Proceedings of the 43rd international ACM SIGIR conference on research and development in information retrieval, pp 639–648. https://doi.org/10.1145/3397271.3401063
15. Gutmann M, Hyvärinen A (2010) Noise-contrastive estimation: a new estimation principle for unnormalized statistical models. In: Proceedings of the thirteenth international conference on artificial intelligence and statistics, JMLR workshop and conference proceedings, pp 297–304
16. Rendle S, Freudenthaler C, Gantner Z et al (2009) Bpr: Bayesian personalized ranking from implicit feedback. In: Proceedings of the twenty-fifth conference on uncertainty in artificial intelligence, pp 452–461
17. Wu J, Wang X, Feng F et al (2021) Self-supervised graph learning for recommendation. In: Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval, pp 726–735. https://doi.org/10.1145/3404835.3462862
18. Dai E, Jin W, Liu H et al (2022) Towards robust graph neural networks for noisy graphs with sparse labels. In: Proceedings of the fifteenth ACM international conference on web search and data mining, pp 181–191. https://doi.org/10.1145/3488560.3498408
19. Chen Y, Wu L, Zaki M (2020) Iterative deep graph learning for graph neural networks: better and robust node embeddings. Adv Neural Inf Process Syst 33:19314–19326
20. Ma Y, Liu X, Zhao T et al (2021) A unified view on graph neural networks as graph signal denoising. In: Proceedings of the 30th ACM international conference on information & knowledge management, pp 1202–1211. https://doi.org/10.1145/3459637.3482225
21. Chen H, Wang L, Lin Y et al (2021) Structured graph convolutional networks with stochastic masks for recommender systems. In: Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval, pp 614–623. https://doi.org/10.1145/3404835.3462868
22. Feng P, Qian Y, Liu X et al (2021) Robust graph collaborative filtering algorithm based on hierarchical attention. In: Web information systems and applications: 18th international conference, WISA 2021, Kaifeng, China, September 24–26, 2021, Proceedings 18, Springer, pp 625–632. https://doi.org/10.1007/978-3-030-87571-8_54
23. Dai J, Yuan W, Bao C et al (2022) Dgnn: denoising graph neural network for session-based recommendation. In: 2022 IEEE 9th international conference on data science and advanced analytics (DSAA), IEEE, pp 1–8. https://doi.org/10.1109/DSAA54385.2022.10032399
24. Zhang C, Zheng W, Liu Q et al (2022) Sedgn: sequence enhanced denoising graph neural network for session-based recommendation. Expert Syst Appl 203:117391. https://doi.org/10.1016/J.ESWA.2022.117391
25. Zhang H, Wang Le, Sun Z et al (2023) Knowledge-aware dual-channel graph neural networks for denoising recommendation. The Computer Journal p bxad085. https://doi.org/10.1093/comjnl/bxad085
26. Kang M, Park J (2020) Contragan: contrastive learning for conditional image generation. Adv Neural Inf Process Syst 33:21357–21369
27. Deng Y, Yang J, Chen D et al (2020) Disentangled and controllable face image generation via 3d imitative-contrastive learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 5154–5163. https://doi.org/10.1109/CVPR42600.2020.00520
28. Lee KS, Tran NT, Cheung NM (2021) Infomax-gan: improved adversarial image generation via information maximization and contrastive learning. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision, pp 3942–3952. https://doi.org/10.1109/WACV48630.2021.00399
29. Sarto S, Barraco M, Cornia M et al (2023) Positive-augmented contrastive learning for image and video captioning evaluation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 6914–6924. https://doi.org/10.1109/CVPR52729.2023.00668
30. Zhou C, Ma J, Zhang J et al (2021) Contrastive learning for debiased candidate generation in large-scale recommender systems. In: Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining, pp 3985–3995. https://doi.org/10.1145/3447548.3467102
31. Yu J, Yin H, Xia X et al (2022) Are graph augmentations necessary? simple graph contrastive learning for recommendation. In: Proceedings of the 45th international ACM SIGIR conference on research and development in information retrieval, pp 1294–1303. https://doi.org/10.1145/3477495.3531937
32. Zhou X, Sun A, Liu Y et al (2023) Selfcf: a simple framework for self-supervised collaborative filtering. ACM Transact Recommend Syst 1(2):1–25. https://doi.org/10.1145/3591469
33. Yang Y, Huang C, Xia L et al (2022) Knowledge graph contrastive learning for recommendation. In: Proceedings of the 45th international ACM SIGIR conference on research and development in information retrieval, pp 1434–1443 https://doi.org/10.1145/3477495.3532009
34. Zou D, Wei W, Mao XL et al (2022) Multi-level cross-view contrastive learning for knowledge-aware recommender system. In: Proceedings of the 45th international ACM SIGIR conference on research and development in information retrieval, pp 1358–1368. https://doi.org/10.1145/3477495.3532025
35. Wang L, Zhao W, Wei Z et al (2022) Simkgc: simple contrastive knowledge graph completion with pre-trained language models. In: Proceedings of the 60th annual meeting of the association for computational linguistics (vol 1: Long Papers), pp 4281–4294. https://doi.org/10.18653/V1/2022.ACL-LONG.295
36. Yang Y, Huang C, Xia L et al (2023) Knowledge graph self-supervised rationalization for recommendation. In: Proceedings of the 29th ACM SIGKDD conference on knowledge discovery and data mining, pp 3046–3056 https://doi.org/10.1145/3580305.3599400
37. Wei W, Huang C, Xia L et al (2022) Contrastive meta learning with behavior multiplicity for recommendation. In: Proceedings of the fifteenth ACM international conference on web search and data mining, pp 1120–1128. https://doi.org/10.1145/3488560.3498527
38. Xuan H, Liu Y, Li B et al (2023) Knowledge enhancement for contrastive multi-behavior recommendation. In: Proceedings of the sixteenth ACM international conference on web search and data mining, pp 195–203. https://doi.org/10.1145/3539597.3570386
39. Fu S, Ren Q (2023) Multi-behavior recommendation with svd graph neural networks. https://doi.org/10.48550/ARXIV.2309.06912
40. Wei Y, Ma H, Wang Y et al (2023) Dual graph attention networks for multi-behavior recommendation. International Journal of Machine Learning and Cybernetics, pp 1–16. https://doi.org/10.1007/S13042-023-01801-0
41. Qiao Z, Yan H, Han L (2023) Mixmbr: contrastive learning for multi-behavior recommendation. In: International conference on database systems for advanced applications. Springer, pp 434–445. https://doi.org/10.1007/978-3-031-30672-3_29
42. Xie X, Sun F, Liu Z et al (2022) Contrastive learning for sequential recommendation. In: 2022 IEEE 38th international conference on data engineering (ICDE), IEEE, pp 1259–1273. https://doi.org/10.1109/ICDE53745.2022.00099

43. Qiu R, Huang Z, Yin H et al (2022) Contrastive learning for representation degeneration problem in sequential recommendation. In: Proceedings of the fifteenth ACM international conference on web search and data mining, pp 813–823. https://doi.org/10.1145/3488560.3498433
44. Zhou P, Gao J, Xie Y et al (2023) Equivariant contrastive learning for sequential recommendation. In: Proceedings of the 17th ACM conference on recommender systems, pp 129–140. https://doi.org/10.1145/3604915.3608786
45. Chen Y, Liu Z, Li J et al (2022) Intent contrastive learning for sequential recommendation. In: Proceedings of the ACM web conference, vol 2022, pp 2172–2182. https://doi.org/10.1145/3485447.3512090
46. Wu J, Fan W, Chen J et al (2022) Disentangled contrastive learning for social recommendation. In: Proceedings of the 31st ACM international conference on information & knowledge management, pp 4570–4574. https://doi.org/10.1145/3511808.3557583
47. Wang T, Xia L, Huang C (2023) Denoised self-augmented learning for social recommendation. In: Proceedings of the thirty-second international joint conference on artificial intelligence, IJCAI 2023, 19th-25th August 2023, Macao, SAR, China. ijcai.org, pp 2324–2331. https://doi.org/10.24963/IJCAI.2023/258
48. Jiang F, Cao Y, Wu H et al (2022) Social recommendation based on multi-auxiliary information constrastive learning. Mathematics (2227-7390) 10(21). https://doi.org/10.3390/math10214130
49. Luo L, Liu B (2022) Dual-contrastive for federated social recommendation. In: 2022 International Joint Conference on Neural Networks (IJCNN), IEEE, pp 1–8. https://doi.org/10.1109/IJCNN55064.2022.9892278
50. Cai X, Huang C, Xia L et al (2022) Lightgcl: simple yet effective graph contrastive learning for recommendation. In: The eleventh international conference on learning representations