Enhancing Long-Tail Bundle Recommendations Utilizing Composition Pattern Modeling

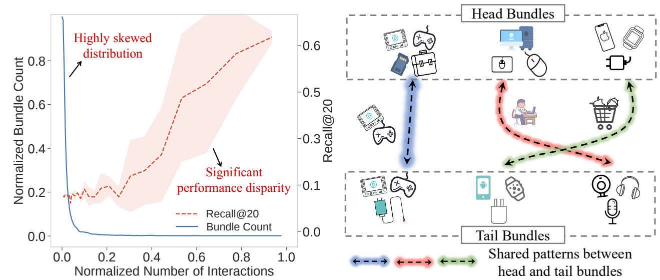
Tianhui Ma[1] , **Shuyao Wang**[1] , **Zhi Zheng**[2*] and **Hui Xiong**[3,4*]
[1]University of Science and Technology of China
[2]State Key Laboratory of Cognitive Intelligence, University of Science and Technology of China
[3]Thrust of Artificial Intelligence, The Hong Kong University of Science and Technology (Guangzhou)
[4]Department of Computer Science and Engineering, The Hong Kong University of Science and Technology, Hong Kong SAR, China
{matianhui, shuyaowang}@mail.ustc.edu.cn, zhengzhi97@ustc.edu.cn, xionghui@ust.hk

Abstract

Bundle recommendation aims to provide users with a one-stop service by offering a collection of related items. However, these systems face a significant challenge, where a small portion of bundles accumulate most interactions while the long-tail bundles receive few interactions. This imbalance leads to poor performance for long-tail bundles despite their potential to satisfy diverse user needs. Existing long-tail item recommendation methods fail to effectively address this problem, as long-tail bundle recommendation requires not only capturing the user-bundle interactions but also the item compositions in different bundles. Therefore, in this paper, we propose Composition-Aware Long-tail Bundle Recommendation (CALBRec), which leverages the inherent composition patterns shared across different bundles as valuable signals for further representation augmentation and recommendation enhancement. Specifically, to solve the complexity of modeling shared composition patterns due to the exponential explosion caused by the growing number of items and bundle sizes, we first introduce a composition-aware tail adapter to capture the shared composition patterns and then adaptively integrate them into individual bundle representations. Moreover, to mitigate the impact of noise in user-bundle interaction data, we propose to map the bundle representations into a set of learnable prototypes, and we further propose a prototype learning module to combine the composition patterns with interaction signals for tail bundles. Extensive experiments on three public datasets demonstrate that our method can improve the performance on bundle recommendation significantly, especially on the long-tail bundles.

1 Introduction

Bundle recommendation aims to provide a collection of items as a convenient option to meet specific user demands [Cao et al., 2017a; Chen et al., 2019a; Hu et al., 2020; Zheng et al., 2021; Zheng et al., 2023]. Recently, bundle recommendation has gained significance attention in many online applications such as e-commerce and music streaming platforms, as it offers one-stop convenience for users and aligns with widely adopted marketing strategies [Ma et al., 2022; Chang et al., 2023].



(a) (b)

Figure 1: (a) Long-tail distribution of bundles (Bundle Count, blue line) and recommendation performance (Recall@20, red) with varying interaction frequencies. (b) Illustration of composition patterns between head and tail bundles which provide stable signals for enhancing tail bundle representations.

In recent years, bundle recommendation has seen substantial progress mainly through matrix factorization-based approaches [Pathak et al., 2017; Cao et al., 2017a; Chen et al., 2019b; Brosh et al., 2022; He et al., 2019] and graph learning-based approaches [Deng et al., 2020; Zhang et al., 2022; Yu et al., 2022; Wei et al., 2023b; Ren et al., 2023]. Two-view graph-based modeling has emerged as an effective approach through a user-bundle view for modeling user-bundle interactions and a user-item view for modeling items compositions in different bundles. However, these approaches face critical limitations in handling long-tail bundle recommendation problem, which is very common in real-world scenarios. As shown in Figure. 1(a), we observe a highly skewed distribution of bundle interactions on Youshu [Chen et al., 2019b] dataset, where a small portion of bundles (head bundles) accumulates most interactions while the majority (tail bundles) receives few interactions. The performance curve of CrossCBR [Ma et al., 2022], which is a representative bundle recommendation method, plotted in the same figure reveals

*Zhi Zheng and Hui Xiong are both corresponding authors

significant performance disparity across bundles with varying interaction frequencies.

Indeed, some methods for long-tail item recommendation have already been conducted. For example, data augmentation methods adjust the training data distribution through resampling [Zhang and Hong, 2021; Park and Tuzhilin, 2008] or synthetic data generation [Tang *et al.*, 2020; Wu *et al.*, 2019] to enhance the interaction feature of tail items. Methods incorporating auxiliary information leverage item attributes [Zhu *et al.*, 2020; Ouyang *et al.*, 2021] or textual descriptions [Liang *et al.*, 2020] to enrich tail item features. However, these approaches fail to address the unique challenges in long-tail bundle recommendations where both user-bundle interactions and item compositions in different bundles need to be jointly considered.

To solve the above problems, we explore the potential of leveraging bundle composition patterns as a source of valuable signals for enhancing long-tail bundle recommendations. Specifically, we observe that different bundles naturally share common item composition patterns that reflect similar consumption behaviors across various scenarios (e.g., shared functional elements and usage contexts), as illustrated in Figure 1(b). Along this line, we propose to leverage these inherent patterns as stable signals for enhancing long-tail bundle representations. To achieve this goal, we need to address two notable challenges: (1)**Composition Pattern Complexity.** The complexity stems from the exponential explosion caused by the growing number of items and bundle sizes, complicating the modeling of composition patterns that are shared across different bundles. (2)**Long-tail Noisy Feature Integration.** The limited interaction records of tail bundles introduce noise from sparsity, leading to biased and under-represented behavior features [Wei *et al.*, 2023a]. When integrating these noisy features into the modeling process, the noise can propagate between different modeling views, which can further distort the bundle representations and degrade the model performance.

Therefore, in this paper, we propose Composition-Aware Long-tail Bundle Recommendation (CALBRec), a novel framework that jointly tackles the above challenges. Specifically, to address the **Composition Pattern Complexity** problem, we design a composition-aware tail adapter, which can model the shared composition patterns across bundles and adaptively integrate these patterns into bundle representations. Specifically, we employ an adaptive integration mechanism that minimizes pattern influence for head bundles while enhancing long-tail bundle representations, which ensures that tail bundles can benefit from shared composition patterns while preserving the effective representations of frequently interacted bundles. To tackle the **Long-tail Noisy Feature Integration**, we introduce a prototype learning module. Through mapping bundle representations from dual views into a set of learnable prototypes, this module achieves robust feature integration by comparing bundle assignments in a shared prototype space. This module enables effective representation learning for long-tail bundles despite their limited and biased interaction data. Our contributions are summarized as follows:

- **Problem.** We explicitly address the long-tail challenge in bundle recommendation, a challenging problem of significant impact in real-world scenarios.
- **Method.** We propose CALBRec, adaptively leveraging item composition patterns across different bundles to enhance long-tail bundle recommendation.
- **Experiments.** Extensive experiments on three public datasets validate that our method achieves significant improvements, particularly for long-tail bundles.

The code is available at https://anonymous.4open.science/r/CALBRec-A291.

2 Related Works

Bundle Recommendation. Prior work in bundle recommendation has mainly progressed along matrix factorization-based methods [Pathak *et al.*, 2017; Cao *et al.*, 2017a] and graph learning-based methods [Deng *et al.*, 2020; Chang *et al.*, 2020; Bai *et al.*, 2019]. Particularly, bundle recommendation research has explored various tasks, including accurate recommendation of existing bundles [Cao *et al.*, 2017b], personalized bundle creation [Deng *et al.*, 2021], leveraging conversational approaches [He *et al.*, 2022], increasing diversity [Jeon *et al.*, 2023], and improving recommendations for new bundles [Jeon *et al.*, 2024]. However, the long-tail distribution challenge in bundle recommendation, which is prevalent in real-world scenarios, remains under-investigated. Existing methods typically assume sufficient historical interactions for all bundles, making it difficult to learn effective representations for tail bundles with limited interactions, leading to biased and underrepresented bundle features. Our work addresses this limitation by developing a framework specifically designed for long-tail bundle recommendations.

Graph-based Long-tail Methods. The problem of long-tailed distributions in recommendation systems has motivated extensive research, with significant efforts made to address the issue of insufficient interactions for tail items [He *et al.*, 2020; Liu *et al.*, 2021; Deng *et al.*, 2020; Zheng *et al.*, 2022b; Chang *et al.*, 2023]. Existing solutions typically fall into two categories: data augmentation methods and auxiliary information-based methods. Data augmentation methods aim to adjust the training data distribution to enhance the interaction features of tail items. Approaches such as resampling [Zhang and Hong, 2021; Park and Tuzhilin, 2008] balance the interaction distribution, while synthetic data generation methods [Tang *et al.*, 2020; Wu *et al.*, 2019] create additional samples to augment the characteristics of tail items. In contrast, auxiliary information-based methods focus on enriching the features of tail items by incorporating supplementary data, such as item attributes [Zhu *et al.*, 2020; Wu *et al.*, 2024] or textual descriptions [Liang *et al.*, 2020; Zheng *et al.*, 2022a], to improve their representation in the recommendation process. However, existing long-tail methods are limited in bundle recommendation scenarios, as they primarily focus on single-item interactions without considering the complex item compositions within bundles. Our work addresses this gap by developing a framework that jointly models both user-bundle interactions and bundle composition patterns in the long-tail context.

3 Preliminary

In this section, we present the problem formulation for bundle recommendation and the foundational model structure with dual-view design and embedding propagation mechanisms.

3.1 Problem Formulation

Given a set of users $\mathcal{U} = \{u_1, u_2, \cdots, u_M\}$, a set of bundles $\mathcal{B} = \{b_1, b_2, \cdots, b_N\}$, and a set of items $\mathcal{I} = \{i_1, i_2, \cdots, i_O\}$, where M, N, and O are the number of users, bundles, and items, respectively. The user-bundle interactions, user-item interactions, and bundle-item affiliations are denoted as $\mathbf{X}_{M\times N} = \{x_{ub}|u \in \mathcal{U}, b \in \mathcal{B}\}$, $\mathbf{Y}_{M\times O} = \{y_{ui}|u \in \mathcal{U}, i \in \mathcal{I}\}$, and $\mathbf{Z}_{N\times O} = \{z_{bi}|b \in \mathcal{B}, i \in \mathcal{I}\}$, respectively. $x_{ub}, y_{ui}, z_{bi} \in \{0, 1\}$, where 1 represents an interaction between the user-bundle or user-item pair, or the item belongs to a certain bundle. Note that since we deduplicate the historical bundle and item interactions for each user, each element of $\mathbf{X}$ and $\mathbf{Y}$ is a binary value rather than an integer. In addition, $\mathbf{X}$ and $\mathbf{Y}$ are separately generated, where users are allowed to directly interact with both bundles and individual items. Therefore, $\mathbf{X}$ and $\mathbf{Y}$ contain different information, which heuristically enables the cooperative effect between the two different views. The goal of bundle recommendation task is to learn a model from the historical $\{\mathbf{X}, \mathbf{Y}, \mathbf{Z}\}$ and predict the unseen user-bundle interactions in $\mathbf{X}$.

3.2 Base Model

The base model employs a two-view graph structure to learn representations through user-bundle interactions and user-item interactions. Both views utilize LightGCN [He *et al.*, 2020] for embedding propagation, formalized as:

$$\text{LightGCN}(v, \mathcal{N}_v, k) = \sum_{u\in\mathcal{N}_v} \frac{1}{\sqrt{|\mathcal{N}_v|}\sqrt{|\mathcal{N}_u|}} \mathbf{e}_u^{(k-1)},$$

where $\mathbf{e}_v^{(k)}$ represents node embeddings at layer k, and $\mathcal{N}_v$ denotes the neighbor set of node v.

User-bundle View. For user-bundle interactions, at layer k, user and bundle embeddings $\mathbf{h}_u^{(k)}$ and $\mathbf{h}_b^{(k)}$ are obtained through LightGCN propagation with their respective neighbor sets $\mathcal{N}_u$ and $\mathcal{N}_b$. The final embeddings are computed by:

$$\mathbf{h}_u = \sum_{k=0}^{K} \frac{1}{k+1}\mathbf{h}_u^{(k)}, \quad \mathbf{h}_b = \sum_{k=0}^{K} \frac{1}{k+1}\mathbf{h}_b^{(k)},$$

where $\mathbf{h}_u, \mathbf{h}_b \in \mathbb{R}^d$ represent user and bundle embeddings.

User-item View. Similarly for user-item interactions, embeddings $\mathbf{a}_u^{(k)}$ and $\mathbf{a}_i^{(k)}$ are obtained through LightGCN propagation with neighbor sets $\mathcal{N}'_u$ and $\mathcal{N}_i$, respectively. The final embeddings are computed by:

$$\mathbf{a}_u = \sum_{k=0}^{K} \frac{1}{k+1}\mathbf{a}_u^{(k)}, \quad \mathbf{a}_i = \sum_{k=0}^{K} \frac{1}{k+1}\mathbf{a}_i^{(k)},$$

where $\mathbf{a}_u, \mathbf{a}_i \in \mathbb{R}^d$ represent user and item embeddings, respectively. For each bundle b, its representation $\mathbf{a}_b$ is obtained by averaging embeddings of its constituent items with $\mathbf{a}_b = \frac{1}{|\mathcal{N}'_b|}\sum_{i\in\mathcal{N}'_b} \mathbf{a}_i$, $\mathcal{N}'_b$ denotes the set of items in bundle b.

Prediction Scores. Based on the embeddings from both views, we compute:

$$h_{ub} = \mathbf{h}_u^\top \mathbf{h}_b, \quad a_{ub} = \mathbf{a}_u^\top \mathbf{a}_b, \tag{1}$$

where h_{ub} and a_{ub} denote the matching scores in user-bundle view and user-item view, respectively.

4 Methodology

In this section, we outline the overall architecture of the CALBRec framework (as shown in Figure 2) and provide a detailed description of its key components. The main idea of CALBRec is leveraging composition patterns between head and tail bundles to enhance bundle representations for long-tail bundle recommendation. Specifically, we develop a composition-aware long-tail adapter that enhances tail bundle representations by modeling shared composition patterns and integrating them into bundle representations through an adaptive fusion mechanism. Furthermore, we propose to map the bundle representations into a set of learnable prototypes and introduce a prototype learning module to combine the composition patterns with interaction signals for tail bundles. These modules are integrated into a unified learning framework.

4.1 Composition-aware Long-tail Adapter

We first model the shared compositional patterns across bundles and then personalize feature fusion, enabling tail bundles to benefit from patterns learned from head bundles while preserving the strong features of frequently interacted bundles.

Compositional Pattern Modeling and Adaptive Integration. We first construct a global feature vector $\mathbf{t}$, which is learnable and shared across all bundles to capture general composition patterns. Then, for each bundle b, we devise a personalization function ϕ that transforms $\mathbf{t}$ into a localized feature vector $\mathbf{t_b}$ by considering both the representation of bundle b and its constituent items:

$$\mathbf{t}_b^k = \phi(\mathbf{h}_b^k, \mathbf{a}_{i_1}^k, \mathbf{a}_{i_2}^k, \ldots, \mathbf{a}_{i_m}^k, \mathbf{t}^k), \tag{2}$$

where $\mathbf{h}_b^k$ is the representation of bundle b in the user-bundle view at layer k, and $\mathbf{a}_i^k$ represents the feature of item i in the user-item view that constitutes bundle b.

To personalize the global patterns to each bundle, we use a personalized function with complementary adaptation operations, while adjusting the degree of pattern integration:

$$\mathbf{t}_b^k = \phi\left(\mathbf{h}_b^k, \mathbf{a}_{i_1}^k, \mathbf{a}_{i_2}^k, \ldots, \mathbf{a}_{i_m}^k, \mathbf{t}^k\right) = (\boldsymbol{\gamma}_b^k + 1) \odot \mathbf{t}^k + \boldsymbol{\beta}_b^k,$$

where $\boldsymbol{\gamma}_b^k \in \mathbb{R}^{d_k}$ modulates the importance of different pattern dimensions through scaling, and $\boldsymbol{\beta}_b^k \in \mathbb{R}^{d_k}$ provides bundle-specific adjustments through shifting. The scaling factor is centered around one through $(\boldsymbol{\gamma}_b^k + 1)$ to maintain the original pattern information while allowing for adaptive adjustments. These adaptation vectors are computed as:

$$\boldsymbol{\gamma}_b^k = \text{LeakyReLU}\left(\mathbf{W}_\gamma^{k,1}\mathbf{h}_b^k + \mathbf{W}_\gamma^{k,2} \sum_{i\in\{i_1,i_2,\ldots,i_m\}} \mathbf{a}_i^k\right),$$

$$\boldsymbol{\beta}_b^k = \text{LeakyReLU}\left(\mathbf{W}_\beta^{k,1}\mathbf{h}_b^k + \mathbf{W}_\beta^{k,2} \sum_{i\in\{i_1,i_2,\ldots,i_m\}} \mathbf{a}_i^k\right),$$

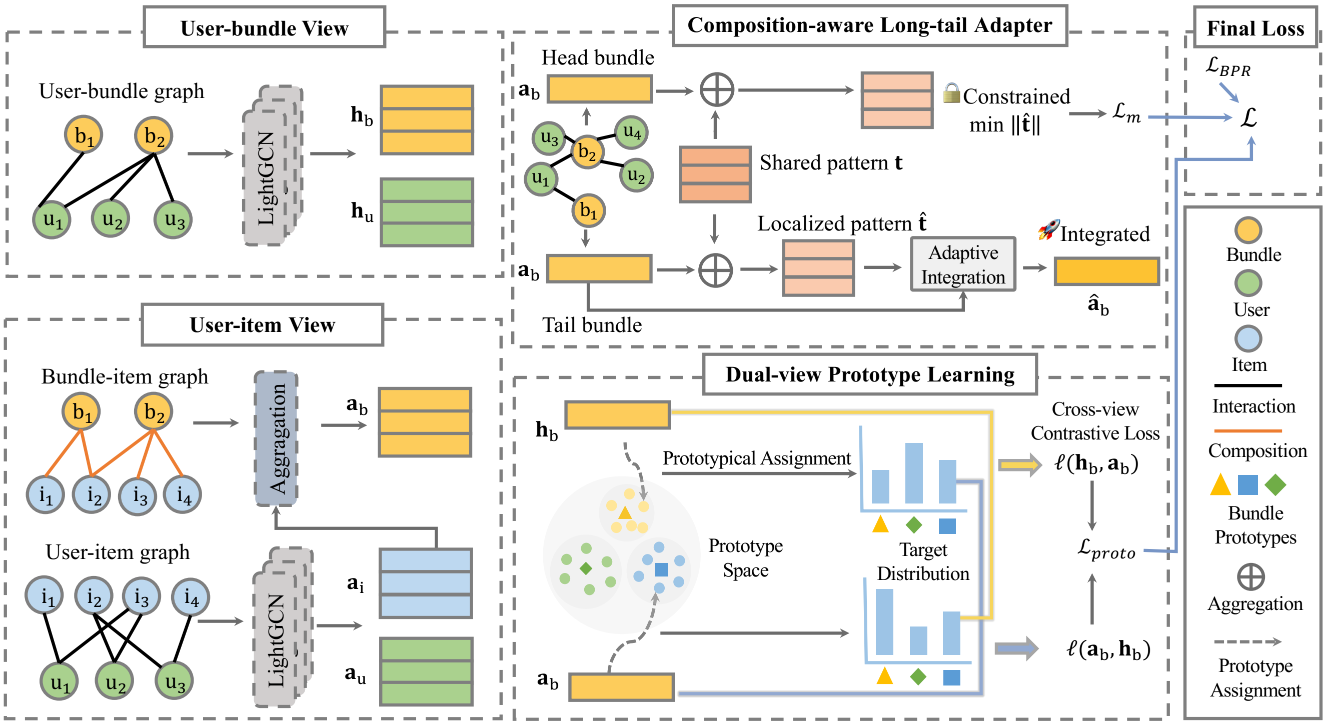


Figure 2: Overview of CALBRec. The framework consists of a dual-view structure with user-bundle and user-item views for representation learning, a composition-aware long-tail adapter for shared composition patterns modeling, and a dual-view prototype learning module that integrates composition patterns with interaction signals.

where $\mathbf{W}_{*}^{k,*} \in \mathbb{R}^{d_k \times d_k}$ are learnable parameters that transform the composition context into adaptation signals. This design enables each bundle to flexibly harness shared patterns, where the scaling operation governs the degree of pattern integration, and the shifting operation allows for targeted adjustments based on the compositional context.

However, not all bundles have the same need for complementary features. Head bundles with abundant interactions have already learned reliable representations from rich user feedback, while tail bundles with limited interactions can benefit more from the compositional patterns shared with similar head bundles. In the following, we will describe how CALBRec achieves differentiated control over feature fusion for head and tail bundles.

Frequency-aware Enhancement Control. To achieve this differentiated control, we design a frequency-aware enhancement mechanism that applies distinct strategies to head and tail bundles. For head bundles, we maintain their well-learned representations by minimizing additional pattern influence:

$$\forall b \in \mathcal{B}_{\text{head}} : \mathbf{t}_b \rightarrow \mathbf{0}.$$

This constraint ensures that the final representations of head bundles primarily rely on their rich interaction signals.

For tail bundles, we enhance their representations through controlled pattern integration: $\hat{\mathbf{a}}_b = \mathbf{a}_b + \lambda \mathbf{t}_b$. Here, a_b denotes the original representation and λ controls the degree of enhancement. This enhancement allows tail bundles to incorporate useful patterns while preserving their original characteristics, with the user-item view score computed using the enhanced representation for tail bundles as $a_{ub} = \mathbf{a}_u^T \hat{\mathbf{a}}_b$. Through this frequency-aware control mechanism, we can effectively improve the representation quality of tail bundles while maintaining the strong performance of head bundles.

4.2 Dual-view Prototype Learning

To mitigate the impact of noise in user-bundle interaction data, we introduce a prototype learning module to map the bundle representations into a set of learnable prototypes and further combine the composition patterns with interaction signals for tail bundles.

First, we construct K trainable prototype vectors $\mathbf{C} \in \mathbb{R}^{K \times D}$ to represent shared features at the bundle level, where D is the feature dimension. Instead of directly mapping bundle embeddings to the prototypes with maximum dot similarity, we use a soft assignment approach to assign bundles to multiple prototypes, which can mitigate the impact of mismatched mappings for biased representations of tail bundles. Specifically, we introduce assignment matrix $\mathbf{Q} \in \mathbb{R}^{B \times K}$, where each element $\mathbf{Q}_{b,m}$ represents the assignment strength between the bundle b and prototype m. Then, we introduce the following constraints:

$$\mathbf{Q}^{\top}\mathbf{1}_B = \frac{1}{K}\mathbf{1}_K, \qquad \mathbf{Q}\mathbf{1}_K = \frac{1}{B}\mathbf{1}_B,$$

where $\mathbf{1}_B$ and $\mathbf{1}_K$ represent vectors of ones in the number of bundles and prototypes, respectively. These dual constraints ensure that each prototype receives equal total assignment weight from bundles and each bundle distributes equal weights across prototypes. This balanced assignment mechanism mitigates representation bias for tail bundles by avoiding forced one-to-one prototype mappings and allowing them to leverage interaction patterns from other bundles assigned to the same prototypes.

The solution satisfying these constraints can be obtained through the Sinkhorn-Knopp algorithm [Cuturi, 2013], which achieves fast convergence through iterative row and column normalization. Specifically, we establish initial assignment relationships by computing dot-product similarity between bundle embeddings and prototype vectors. Then we obtain $\mathbf{Q}$ through the Algorithm 1.

For the obtained assignment matrix $\mathbf{Q}$, we further convert it into a probability distribution through a temperature-scaled softmax operation:

$$\mathbf{P}_{b,m} = \frac{\exp(\mathbf{Q}_{b,m}/\tau)}{\sum_{k=1}^{K} \exp(\mathbf{Q}_{b,k}/\tau)},$$

where $\mathbf{P}_{b,m}$ represents the probability of assigning bundle b to prototype m, and the temperature parameter τ controls the concentration of the distribution: a smaller τ produces more concentrated distributions that help reduce noise impact, while a larger τ encourages uniform distributions to preserve more association information.

The final prototype assignments are optimized through a contrastive objective detailed in Section 4.3.

Algorithm 1 Sinkhorn-Knopp Algorithm

Require: Initial probability matrix $\mathbf{Q} \in \mathbb{R}^{B\times K}$
Ensure: The optimal assignment matrix $\hat{\mathbf{Q}}$
1: **while** not converged **do**
2: Row normalization: $\mathbf{Q} \leftarrow \mathbf{Q}\text{diag}(\frac{1}{B\mathbf{Q}\mathbf{1}_K})$
3: Column normalization: $\mathbf{Q} \leftarrow \text{diag}(\frac{1}{K\mathbf{Q}^\top \mathbf{1}_B})\mathbf{Q}$
4: **end while**
5: $\hat{\mathbf{Q}} \leftarrow \mathbf{Q}$

4.3 Model Optimization

Recommendation Objective. The Bayesian Personalized Ranking (BPR) loss [Rendle *et al.*, 2012] is used as the main loss. The prediction score combines both views as $\hat{y}_{ub} = h_{ub} + a_{ub}$, where h_{ub} and a_{ub} are defined in Eq. (1). The BPR loss is formulated as:

$$\mathcal{L}_{BPR} = \mathop{\mathbb{E}}_{(u,b^+,b^-)\sim p_{data}} -\ln\sigma(\hat{y}_{ub^+} - \hat{y}_{ub^-}),$$

where p_{data} represents the user-bundle interaction distribution, b^+ and b^- denote the interacted and non-interacted bundles, respectively.

Composition-aware Adaptation Objective. In order to achieve the personalized composition pattern integration proposed in Section 4.1, we introduce a corresponding optimization objective. The loss function is defined as:

$$\mathcal{L}_m = \sum_{b\in\mathcal{B}_{\text{head}}} I_b \sum_{k=1}^{\ell} \|\mathbf{t}_b^{k-1}\|_2^2,$$

where I_b is the indicator function for head bundles (1 for head, 0 for tail), ℓ denotes the number of network layers, and $\mathbf{t}_b^{k-1}$ is the pattern representation obtained through the adaptive fusion mechanism in Section 4.1 at layer $k-1$.

Dataset	#U	#I	#B	#U-I	#U-B	#Avg.I/B
Youshu	8,039	32,770	4,771	138,515	51,377	37.03
NetEase	18,528	123,628	22,864	1,128,065	302,303	77.80
iFashion	53,897	42,563	27,694	2,290,645	1,679,708	3.86

Table 1: Dataset Statistics.

This objective implements the frequency-aware enhancement control proposed in Section 4.1 by minimizing the pattern representation norm of head bundles: for head bundles, the influence of additional patterns is limited through L_2 norm constraints; for tail bundles, no constraints are imposed, allowing them to fully utilize pattern enhancement obtained through adaptive fusion.

Prototype Contrastive Objective. Following Section 4.2, let $\mathbf{Q}^h$ and $\mathbf{Q}^a$ denote the assignment matrices for bundles from the user-bundle view and user-item view, respectively. The corresponding probability distributions after temperature-scaled softmax normalization are denoted as $\mathbf{P}^h$ and $\mathbf{P}^a$, respectively. We formulate the prototypical contrastive loss between views as:

$$\ell_b(\mathbf{h}_b, \mathbf{a}_b) = -\sum_{m=1}^{K} \mathbf{Q}_{b,m}^h \log \mathbf{P}_{b,m}^h,$$

The cross-entropy loss encourages bundle representations from both views to be consistently assigned to the shared prototype space, where each prototype serves as a learnable center for capturing common patterns across bundles. By minimizing the discrepancy between assignment distributions $\mathbf{Q}^h$ and $\mathbf{P}^h$, we enable bundles with similar characteristics to be mapped to similar prototypes, regardless of their interaction frequencies. The bidirectional consistency constraints ($\mathbf{h}_b \rightarrow \mathbf{a}_b$ and $\mathbf{a}_b \rightarrow \mathbf{h}_b$) further ensure that the learned prototypes effectively integrate information from both interaction patterns and compositional structures. The overall prototype contrastive loss is formulated as:

$$\mathcal{L}_{proto} = \frac{1}{2n}\sum_{i=1}^{n}[\ell_{b_i}(\mathbf{h}_{b_i}, \mathbf{a}_{b_i}) + \ell_{b_i}(\mathbf{a}_{b_i}, \mathbf{h}_{b_i})],$$

Overall Objective. The final objective function combines all three objectives:

$$\mathcal{L} = \mathcal{L}_{BPR} + \lambda\mathcal{L}_{proto} + \gamma\mathcal{L}_m,$$

where λ and γ are hyperparameters that balance different objectives. This multi-objective optimization framework both ensures overall recommendation performance and addresses the long-tail challenge through the two specially designed objectives. The hyperparameters are tuned on the validation set to balance the contribution of different objectives.

5 Experiments

In this section, we first introduce the experimental settings and then compare CALBRec with state-of-the-art methods. We further analyze model components through ablation studies and examine key parameter effects.

Model	Youshu						NetEase						iFashion					
	Recall@20			nDCG@20			Recall@20			nDCG@20			Recall@20			nDCG@20		
	Head	*Tail*	*All*	*Head*	*Tail*	*All*	*Head*	*Tail*	*All*	*Head*	*Tail*	*All*	*Head*	*Tail*	*All*	*Head*	*Tail*	*All*
LightGCN	.3001	.0606	.2425	.1666	.0276	.1382	.0664	.0339	.0496	.0391	.0186	.0254	.1261	.0830	.0837	.0789	.0611	.0612
Over-sampling	.2895	.0585	.2340	.1644	.0272	.1364	.0662	.0336	.0494	.0401	.0186	.0258	.1255	.0826	.0833	.0799	.0619	.0620
Down-sampling	.2683	.0542	.2169	.1321	.0219	.1096	.0610	.0298	.0461	.0352	.0170	.0221	.1184	.0779	.0786	.0757	.0586	.0587
MeLU	.2828	.0571	.2286	.1510	.0250	.1253	.0626	.0320	.0467	.0355	.0168	.0230	.1188	.0782	.0789	.0715	.0553	.0555
MIRec	.2883	.0582	.2330	.1535	.0254	.1274	.0638	.0326	.0477	.0361	.0171	.0234	.1211	.0797	.0804	.0727	.0563	.0564
MGL	.3109	.0628	.2513	.1756	.0291	.1457	.0743	.0379	.0555	.0439	.0208	.0285	.1410	.0928	.0937	.0885	.0685	.0686
BundleNet	.2344	.0473	.1895	.1356	.0225	.1125	.0484	.0098	.0391	.0242	.0040	.0201	.0943	.0621	.0626	.0576	.0446	.0447
BGCN	.3204	.0671	.2615	.1857	<u>.0342</u>	.1554	.0950	.0519	.0673	.0475	.0247	.0360	.1104	.0727	.0733	.0684	.0530	.0531
CrossCBR	<u>.3471</u>	<u>.0701</u>	<u>.2806</u>	<u>.2015</u>	.0334	<u>.1672</u>	.1146	<u>.0585</u>	<u>.0856</u>	.0588	<u>.0280</u>	.0433	.1706	.1123	.1133	.1128	.0873	.0875
CoHeat	.3305	.0585	.2676	.1898	.0262	.1576	**.1181**	.0561	.0771	**.0598**	.0270	<u>.0455</u>	<u>.1741</u>	<u>.1146</u>	<u>.1156</u>	<u>.1129</u>	<u>.0874</u>	<u>.0876</u>
CALBRec	**.3475**	**.0867**	**.2852**	**.2041**	**.0402**	**.1695**	<u>.1150</u>	**.0705**	**.0858**	<u>.0589</u>	**.0383**	**.0462**	**.1745**	**.1166**	**.1176**	**.1142**	**.0884**	**.0886**

Table 2: Performance comparison of CALBRec and baseline methods on three real-world datasets. Bold and underlined values indicate the best and the second best accuracies, respectively.

5.1 Experimental Setup

Datasets. We selected three real-world bundle recommendation datasets that represent diverse bundle recommendation scenarios: Youshu [Chen *et al.*, 2019b] for book bundles; NetEase [Cao *et al.*, 2017a] for music playlists, and iFashion [Chen *et al.*, 2019c] for fashion outfits. The detailed statistics of the datasets are listed in Table 1. These datasets were selected for their varying bundle sizes and interaction sparsity patterns, enabling comprehensive evaluation of long-tail bundle recommendation. Following [Ma *et al.*, 2022; Chang *et al.*, 2020; Deng *et al.*, 2020], we split each dataset into training/validation/testing sets at a 7:1:2 proportion.

Evaluation Metrics. We evaluated the recommendation performance using Recall@k and nDCG@k. These metrics were chosen because Recall@k directly measures hit rate in top recommendations while nDCG@k considers ranking quality [Deng *et al.*, 2020]. We set k to 20. Following the Pareto Principle [Box and Meyer, 1986; Reed, 2001], we separated bundles into head (top 20% most frequent) and tail (remaining 80%) groups to analyze model performance with different interaction frequencies. To ensure reliability, we conducted each experiment 5 times with different random initializations and reported average performance with standard deviations.

Baseline Methods. To fully demonstrate the effectiveness of CALBRec on long-tail bundle recommendation, we compared CALBRec with both bundle recommendation methods and long-tail methods.

Bundle Recommendation Methods:

- **BundleNet** [Deng *et al.*, 2020]: This method builds a user-bundle item tripartite graph, leverages GCN to learn the representations, and applies multi-task learning.
- **BGCN** [Chang *et al.*, 2020; Chang *et al.*, 2023]: This method that leverages graph convolutional networks to capture the complex interactions in bundle recommendation.
- **CrossCBR** [Ma *et al.*, 2022]: A bundle recommendation method that integrates cross-domain information to improve recommendation performance.
- **Coheat** [Jeon *et al.*, 2024]: A method designed for cold-start settings, which aims to handle the challenge of recommending new bundles with limited historical data.

Long-tail Recommendation Methods:

- **Over-sampling** [Amirruddin *et al.*, 2022]: This strategy samples from tail items and is more common in practice since user feedback data is highly valuable.
- **Down-sampling** [He *et al.*, 2008]: Tail items remain unchanged and the head items are down-sampled.
- **MeLU** [Lee *et al.*, 2019]: A meta-learning approach for cold-start prediction, extended to implicit feedback.
- **MIRec** [Zhang *et al.*, 2021]: A transfer learning framework for long-tail recommendation in two-tower mode.
- **MGL** [Liu *et al.*, 2020]: A meta-graph framework for long-tail recommendation via auxiliary item relations and popularity-aware contrastive meta-learning to address skewness and ensure consistency.

Hyper-parameter Settings. For all methods, the embedding size was set as 64, Xavier normal initialization was adopted, the models were optimized using Adam optimizer with the learning rate 0.001, and the batch size was set as 2048. We applied a grid search to optimize hyperparameters. The parameters of baselines were set following their official implementations and optimal configurations reported in the corresponding papers. All the models were trained using Pytorch and NVIDIA Titan-V GPUs.

5.2 Overall Performance

Comparison with Bundle Recommendation Methods. Table 2 presents the comparative analysis between CALBRec and baseline methods across three datasets. The results demonstrate that CALBRec achieves strong performance on tail bundles while maintaining competitive results across head bundles and overall metrics. Among baseline methods, contrastive learning improves performance as shown by CrossCBR over BGCN, while dual-view models (CrossCBR, CoHeat) outperform the single-view BundleNet. Building on these insights, CALBRec adopts a graph-based dual-view

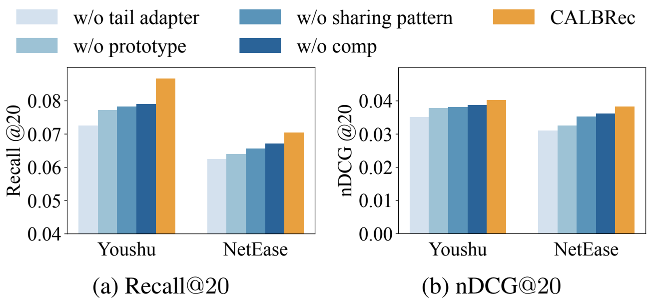


Figure 3: Ablation study in long-tail scenario.

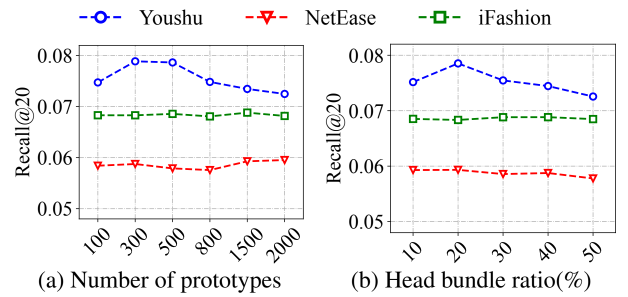


Figure 4: Effect of the number of prototypes and head bundle ratios on Recall@20.

architecture with contrastive learning. CALBRec significantly outperforms existing methods in handling the performance gap between head and tail bundles. Specifically on the Youshu dataset, CrossCBR shows a Recall@20 gap of 0.277 between head and tail bundles, indicating that learning from imbalanced interaction data leads to strong head bundle performance but poor tail bundle recommendations. CALBRec addresses this limitation, significantly narrowing the performance gap by introducing composition pattern modeling, which achieves a 23.7% improvement in Recall@20 for tail bundles compared to CrossCBR and substantially reduces the disparity between head and tail bundle recommendations.

Comparison with Long-tail Methods. The results show that CALBRec consistently surpassed the baselines across the tail-bundle scenario, verifying its superiority. Specifically, CALBRec achieves significant improvements of 28.46% on average on Youshu and NetEase datasets, while showing moderate improvements on iFashion. This performance differential can be explained by the dataset characteristics (Table 1): iFashion has substantially smaller bundles (3.86 items) while Youshu and NetEase contain larger bundles (37.03 and 77.80 items, respectively). The results suggest that our composition pattern modeling and prototype learning mechanisms are particularly effective when applied to bundles with richer compositional structures, as evidenced by the superior performance on Youshu and NetEase.

5.3 Ablation Study

To validate the effectiveness of each component of our model, we conducted ablation studies comparing the full model with four variants: (1) **no-adapter**: which eliminated the composition-aware long-tail adapter module; (2) **no-prototype**: which removed dual-view prototype learning module; (3) **no-sharing**: which replaced the global pattern in composition-aware adapter module with independently sampled random vectors from standard normal distribution; (4) **no-comp**: which removed bundle-specific pattern personalization process by eliminating item vectors in Eq. 2. Figure 3 presents comparative results on Youshu and NetEase datasets in the long-tail scenario. CALBRec demonstrates consistent superior performance across all variants. Notably, removing the long-tail adapter resulted in the most significant degradation, with Recall@20 decreasing by 16.4% and 11.3% on Youshu and NetEase respectively, highlighting the importance of composition-aware long-tail adapter in leveraging shared patterns between head and tail bundles.

5.4 Effect of the Parameter

Impact of the Number of Prototypes. As shown in Figure 4(a), the impact of prototype numbers varies across datasets. Specifically, the Youshu dataset shows notable performance variations, with accuracy declining beyond 500 prototypes. This performance degradation occurs because Youshu is relatively small (Table 1), where excess prototypes begin to memorize individual bundle characteristics rather than extracting generalizable composition patterns. Conversely, the larger NetEase and iFashion datasets maintain robust performance with increased prototypes. Based on these empirical findings, we configure K as 300, 2000, and 2000 for Youshu, NetEase and iFashion, respectively.

Impact of the Ratio of Head Bundles. To investigate the impact of the head-tail ratio, we evaluate CALBRec by varying the head ratio from 10% to 50% on three datasets, as shown in Figure 4(b). The experimental results reveal performance degradation when the ratio increases to 50%, with Youshu showing the largest decline of 7.61% from its peak at 20%, followed by NetEase with 2.62% and iFashion with 0.51%. The high ratio of 50% forces the model to treat a substantial portion of tail bundles as head features, undermining our strategy of distinctive enhancement for head and tail bundles described in Section 4.1. Our empirical results indicate that a moderate ratio of 20% achieves the optimal balance between preserving head representations and enhancing tail features across all evaluated datasets.

6 Conclusion

In this paper, we addressed the challenge of long-tail bundle recommendation and proposed the Composition-Aware Long-tail Bundle Recommendation (CALBRec) framework. CALBRec aims to leverage the inherent composition patterns across different bundles as valuable signals for recommendation enhancement. Specifically, we first designed a novel composition-aware long-tail adapter that effectively captures the shared composition patterns and enhances individual bundle representations adaptively. Additionally, a prototype learning module was introduced to facilitate robust feature integration against the noise in sparse interaction data. Extensive experiments on three public datasets demonstrated that CALBRec significantly improves recommendation performance, particularly for long-tail bundles. In the future, we will explore the potential of leveraging composition patterns in more tasks such as personalized bundle creation, diversity enhancement, and conversational bundle recommendations.

Acknowledgments

This work was supported in part by the National Key R&D Program of China(Grant No.2023YFF0725001),in part by the National Natural Science Foundation of China(Grant No.92370204),in part by the guangdong Basic and Applied Basic Research Foundation(Grant No.2023B1515120057).in part by Guangzhou-HKUST(GZ) Joint Funding Program(Grant No.2023A03J0008), Education Bureau of Guangzhou Municipality.

References

[Amirruddin *et al.*, 2022] Amiratul Diyana Amirruddin, Farrah Melissa Muharam, Mohd Hasmadi Ismail, Ngai Paing Tan, and Mohd Firdaus Ismail. Synthetic minority oversampling technique (SMOTE) and logistic model tree (lmt)-adaptive boosting algorithms for classifying imbalanced datasets of nutrient and chlorophyll sufficiency levels of oil palm (*Elaeis guineensis*) using spectroradiometers and unmanned aerial vehicles. *Comput. Electron. Agric.*, 193:106646, 2022.

[Bai *et al.*, 2019] Jinze Bai, Chang Zhou, Junshuai Song, Xiaoru Qu, Weiting An, Zhao Li, and Jun Gao. Personalized bundle list recommendation. In *WWW*, 2019.

[Box and Meyer, 1986] George EP Box and R Daniel Meyer. An analysis for unreplicated fractional factorials. *Technometrics*, 28(1), 1986.

[Brosh *et al.*, 2022] Tzoof Avny Brosh, Amit Livne, Oren Sar Shalom, Bracha Shapira, and Mark Last. BRUCE: bundle recommendation using contextualized item embeddings. In *RecSys*, 2022.

[Cao *et al.*, 2017a] Da Cao, Liqiang Nie, Xiangnan He, Xiaochi Wei, Shunzhi Zhu, and Tat-Seng Chua. Embedding factorization models for jointly recommending items and user generated lists. In *SIGIR*, 2017.

[Cao *et al.*, 2017b] Da Cao, Liqiang Nie, Xiangnan He, Xiaochi Wei, Shunzhi Zhu, and Tat-Seng Chua. 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*, pages 585–594, 2017.

[Chang *et al.*, 2020] Jianxin Chang, Chen Gao, Xiangnan He, Depeng Jin, and Yong Li. Bundle recommendation with graph convolutional networks. In *SIGIR*, 2020.

[Chang *et al.*, 2023] Jianxin Chang, Chen Gao, Xiangnan He, Depeng Jin, and Yong Li. Bundle recommendation and generation with graph neural networks. *IEEE Trans. Knowl. Data Eng.*, 2023.

[Chen *et al.*, 2019a] Liang Chen, Yang Liu, Xiangnan He, Lianli Gao, and Zibin Zheng. Matching user with item set: Collaborative bundle recommendation with deep attention network. In Sarit Kraus, editor, *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019*, pages 2095–2101. ijcai.org, 2019.

[Chen *et al.*, 2019b] Liang Chen, Yang Liu, Xiangnan He, Lianli Gao, and Zibin Zheng. Matching user with item set: Collaborative bundle recommendation with deep attention network. In *IJCAI*, 2019.

[Chen *et al.*, 2019c] Wen Chen, Pipei Huang, Jiaming Xu, Xin Guo, Cheng Guo, Fei Sun, Chao Li, Andreas Pfadler, Huan Zhao, and Binqiang Zhao. POG: personalized outfit generation for fashion recommendation at alibaba ifashion. In *KDD*, 2019.

[Cuturi, 2013] Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *NIPS*, pages 2292–2300, 2013.

[Deng *et al.*, 2020] Qilin Deng, Kai Wang, Minghao Zhao, Zhene Zou, Runze Wu, Jianrong Tao, Changjie Fan, and Liang Chen. Personalized bundle recommendation in online games. In *CIKM*, 2020.

[Deng *et al.*, 2021] Qilin Deng, Kai Wang, Minghao Zhao, Runze Wu, Yu Ding, Zhene Zou, Yue Shang, Jianrong Tao, and Changjie Fan. Build your own bundle - A neural combinatorial optimization method. In *ACM Multimedia*, pages 2625–2633. ACM, 2021.

[He *et al.*, 2008] Haibo He, Yang Bai, Edwardo A. Garcia, and Shutao Li. ADASYN: adaptive synthetic sampling approach for imbalanced learning. In *IJCNN*, pages 1322–1328. IEEE, 2008.

[He *et al.*, 2019] Yun He, Jianling Wang, Wei Niu, and James Caverlee. A hierarchical self-attentive model for recommending user-generated item lists. In *CIKM*, 2019.

[He *et al.*, 2020] Xiangnan He, Kuan Deng, Xiang Wang, Yan Li, Yong-Dong Zhang, and Meng Wang. Lightgcn: Simplifying and powering graph convolution network for recommendation. In *SIGIR*, 2020.

[He *et al.*, 2022] Zhankui He, Handong Zhao, Tong Yu, Sungchul Kim, Fan Du, and Julian J. McAuley. Bundle MCR: towards conversational bundle recommendation. In *RecSys*, pages 288–298. ACM, 2022.

[Hu *et al.*, 2020] Haoji Hu, Xiangnan He, Jinyang Gao, and Zhi-Li Zhang. Modeling personalized item frequency information for next-basket recommendation. In *SIGIR*, pages 1071–1080. ACM, 2020.

[Jeon *et al.*, 2023] Hyunsik Jeon, Jongjin Kim, Jaeri Lee, Jong-eun Lee, and U Kang. Aggregately diversified bundle recommendation via popularity debiasing and configuration-aware reranking. In *PAKDD (3)*, volume 13937 of *Lecture Notes in Computer Science*, pages 348–360. Springer, 2023.

[Jeon *et al.*, 2024] Hyunsik Jeon, Jong-eun Lee, Jeongin Yun, and U Kang. Cold-start bundle recommendation via popularity-based coalescence and curriculum heating. In *WWW*, pages 3277–3286. ACM, 2024.

[Lee *et al.*, 2019] Hoyeop Lee, Jinbae Im, Seongwon Jang, Hyunsouk Cho, and Sehee Chung. Melu: Meta-learned user preference estimator for cold-start recommendation. In *Proceedings of the 25th ACM SIGKDD International*

Conference on Knowledge Discovery & Data Mining, pages 1073–1082, 2019.

[Liang *et al.*, 2020] Tingting Liang, Congying Xia, Yuyu Yin, and Philip S. Yu. Joint training capsule network for cold start recommendation. In *SIGIR*, pages 1769–1772. ACM, 2020.

[Liu *et al.*, 2020] Zemin Liu, Wentao Zhang, Yuan Fang, Xinming Zhang, and Steven C. H. Hoi. Towards locality-aware meta-learning of tail node embeddings on networks. In *CIKM*, pages 975–984. ACM, 2020.

[Liu *et al.*, 2021] Zemin Liu, Trung-Kien Nguyen, and Yuan Fang. Tail-gnn: Tail-node graph neural networks. In *KDD*, pages 1109–1119. ACM, 2021.

[Ma *et al.*, 2022] Yunshan Ma, Yingzhi He, An Zhang, Xiang Wang, and Tat-Seng Chua. Crosscbr: Cross-view contrastive learning for bundle recommendation. In *KDD*, 2022.

[Ouyang *et al.*, 2021] Wentao Ouyang, Xiuwu Zhang, Shukui Ren, Li Li, Kun Zhang, Jinmei Luo, Zhaojie Liu, and Yanlong Du. Learning graph meta embeddings for cold-start ads in click-through rate prediction. In *SIGIR*, pages 1157–1166. ACM, 2021.

[Park and Tuzhilin, 2008] Yoon-Joo Park and Alexander Tuzhilin. The long tail of recommender systems and how to leverage it. In *RecSys*, pages 11–18. ACM, 2008.

[Pathak *et al.*, 2017] Apurva Pathak, Kshitiz Gupta, and Julian J. McAuley. Generating and personalizing bundle recommendations on *Steam*. In *SIGIR*, 2017.

[Reed, 2001] William J Reed. The pareto, zipf and other power laws. *Economics letters*, 74(1):15–19, 2001.

[Ren *et al.*, 2023] Yuyang Ren, Haonan Zhang, Luoyi Fu, Xinbing Wang, and Chenghu Zhou. Distillation-enhanced graph masked autoencoders for bundle recommendation. In *SIGIR*, 2023.

[Rendle *et al.*, 2012] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. BPR: bayesian personalized ranking from implicit feedback. *CoRR*, abs/1205.2618, 2012.

[Tang *et al.*, 2020] Xianfeng Tang, Huaxiu Yao, Yiwei Sun, Yiqi Wang, Jiliang Tang, Charu C. Aggarwal, Prasenjit Mitra, and Suhang Wang. Graph convolutional networks against degree-related biases. *CoRR*, abs/2006.15643, 2020.

[Wei *et al.*, 2023a] Chunyu Wei, Jian Liang, Di Liu, Zehui Dai, Mang Li, and Fei Wang. Meta graph learning for long-tail recommendation. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 2512–2522, 2023.

[Wei *et al.*, 2023b] Yinwei Wei, Xiaohao Liu, Yunshan Ma, Xiang Wang, Liqiang Nie, and Tat-Seng Chua. Strategy-aware bundle recommender system. In *SIGIR*, 2023.

[Wu *et al.*, 2019] Jun Wu, Jingrui He, and Jiejun Xu. Demonet: Degree-specific graph neural networks for node and graph classification. In Ankur Teredesai, Vipin Kumar, Ying Li, Rómer Rosales, Evimaria Terzi, and George Karypis, editors, *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2019, Anchorage, AK, USA, August 4-8, 2019*, pages 406–415. ACM, 2019.

[Wu *et al.*, 2024] Likang Wu, Zhi Li, Hongke Zhao, Zhenya Huang, Yongqiang Han, Junji Jiang, and Enhong Chen. Supporting your idea reasonably: A knowledge-aware topic reasoning strategy for citation recommendation. *IEEE Transactions on Knowledge and Data Engineering*, 2024.

[Yu *et al.*, 2022] Zhouxin Yu, Jintang Li, Liang Chen, and Zibin Zheng. Unifying multi-associations through hypergraph for bundle recommendation. *Knowl. Based Syst.*, 2022.

[Zhang and Hong, 2021] Chuanyan Zhang and Xiaoguang Hong. Challenging the long tail recommendation on heterogeneous information network. In *ICDM (Workshops)*, pages 94–101. IEEE, 2021.

[Zhang *et al.*, 2021] Yin Zhang, Derek Zhiyuan Cheng, Tiansheng Yao, Xinyang Yi, Lichan Hong, and Ed H. Chi. A model of two tales: Dual transfer learning framework for improved long-tail item recommendation. In *WWW*, pages 2220–2231. ACM / IW3C2, 2021.

[Zhang *et al.*, 2022] Zhenning Zhang, Boxin Du, and Hanghang Tong. Suger: A subgraph-based graph convolutional network method for bundle recommendation. In *CIKM*, 2022.

[Zheng *et al.*, 2021] Zhi Zheng, Chao Wang, Tong Xu, Dazhong Shen, Penggang Qin, Baoxing Huai, Tongzhu Liu, and Enhong Chen. Drug package recommendation via interaction-aware graph induction. In *Proceedings of the Web Conference 2021*, pages 1284–1295, 2021.

[Zheng *et al.*, 2022a] Zhi Zheng, Zhaopeng Qiu, Hui Xiong, Xian Wu, Tong Xu, Enhong Chen, and Xiangyu Zhao. Ddr: Dialogue based doctor recommendation for online medical service. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 4592–4600, 2022.

[Zheng *et al.*, 2022b] Zhi Zheng, Zhaopeng Qiu, Tong Xu, Xian Wu, Xiangyu Zhao, Enhong Chen, and Hui Xiong. Cbr: context bias aware recommendation for debiasing user modeling and click prediction. In *Proceedings of the ACM Web Conference 2022*, pages 2268–2276, 2022.

[Zheng *et al.*, 2023] Zhi Zheng, Chao Wang, Tong Xu, Dazhong Shen, Penggang Qin, Xiangyu Zhao, Baoxing Huai, Xian Wu, and Enhong Chen. Interaction-aware drug package recommendation via policy gradient. *ACM Transactions on Information Systems*, 41(1):1–32, 2023.

[Zhu *et al.*, 2020] Yu Zhu, Jinghao Lin, Shibi He, Beidou Wang, Ziyu Guan, Haifeng Liu, and Deng Cai. Addressing the item cold-start problem by attribute-driven active learning. *IEEE Trans. Knowl. Data Eng.*, 32(4):631–644, 2020.