# Counterfactual Multi-player Bandits for Explainable Recommendation Diversification

Yansen Zhang[1], Bowei He[1], Xiaokun Zhang[1], Haolun Wu[2], Zexu Sun[3], and Chen Ma[1] ✉

[1] Department of Computer Science, City University of Hong Kong, Hong Kong SAR, China yanszhang7-c@my.cityu.edu.hk, boweihe2-c@my.cityu.edu.hk, dawnkun1993@gmail.com, chenma@cityu.edu.hk

[2] School of Computer Science, McGill University, Montreal, Quebec, Canada haolun.wu@mail.mcgill.ca

[3] Gaoling School of Artificial Intelligence, Renmin University of China, Beijing, China sunzexu21@ruc.edu.cn

**Abstract.** Existing recommender systems tend to prioritize items closely aligned with users' historical interactions, inevitably trapping users in the dilemma of "filter bubble". Recent efforts are dedicated to improving the diversity of recommendations. However, they mainly suffer from two major issues: 1) a lack of explainability, making it difficult for the system designers to understand how diverse recommendations are generated, and 2) limitations to specific metrics, with difficulty in enhancing non-differentiable diversity metrics. To this end, we propose a **C**ounterfactual **M**ulti-player **B**andits (CMB) method to deliver explainable recommendation diversification across a wide range of diversity metrics. Leveraging a counterfactual framework, our method identifies the factors influencing diversity outcomes. Meanwhile, we adopt the multi-player bandits to optimize the counterfactual optimization objective, making it adaptable to both differentiable and non-differentiable diversity metrics. Extensive experiments conducted on three real-world datasets demonstrate the applicability, effectiveness, and explainability of the proposed CMB.

**Keywords:** Diversified recommendation · Counterfactual framework · Multi-armed bandits.

## 1 Introduction

Recommendation systems (RS) are widely deployed on various online platforms, such as Google, Facebook, and Yahoo!, to mitigate information overload. However, existing recommendation methods [13,21,35] only prioritize recommending the most relevant items to users, which can have negative consequences for both users and service providers. Users may experience the "filter bubble" [20] problem, leading to limited content diversity, while content providers may face the "Matthew Effect" [18] where new content lacks exposure. Therefore, improving recommendation diversity is essential to enhance the overall user experience and maintain a healthy ecosystem for content providers.

Various approaches have been proposed to diversify the recommended items. Existing methods for diversification can be generally classified into three categories [31]: pre-processing, in-processing, and post-processing methods. Pre-processing methods involve modifying or selecting interaction data before model training [7,40]. In-processing methods, such as treating the need for diversity as a kind of regularization [5,29] or a ranking score [17], integrate diversification strategies into the training process directly. Post-processing methods, like MMR [22,27] and DPP [4,14,30], re-rank the recommended items based on relevance and diversity metrics after the model training.

Unfortunately, current methods still suffer from two main limitations. Firstly, current methods, such as [4,5,7,29,30,40], do not provide adequate explainability regarding how factors affect the diversity of recommendations at the (latent) feature level. This limitation makes it difficult for system designers to understand the underlying drivers of diversity, hindering efforts to enhance model diversity and potentially reducing user satisfaction. Secondly, most diversification methods, like [4,5,17,30,40], rely on diversity metrics to evaluate recommendation results, but they often fail to optimize these metrics directly because these metrics are mostly non-differentiable, as highlighted in a recent survey [31]. While several methods, such as those described in [33], strive to optimize some diversity metrics directly, they are only suitable for very few specific non-differentiable diversity metrics, like $\alpha$-nDCG [8], and cannot handle more commonly used metrics like Prediction Coverage or Subtopic Coverage [10].

To address the aforementioned challenges, we propose a counterfactual framework for explainable recommendation diversification. In response to the first limitation, we propose to identify the factors influencing diversity outcomes under the counterfactual framework. In this framework, perturbations are applied to the representation of items to adjust the diversity level of the ranking lists. Our goal is to identify the "minimal" changes to a specific factor in the factor space that can effectively switch the recommendation results to a desired level of diversity. Then in response to the second limitation, we design a gradient-free Counterfactual Multi-player Bandits (CMB) method to learn these perturbations by optimizing the diversity of recommended items, which is no longer constrained by diversity metrics and recommendation models. The bandit-based approach searches for the best perturbations applied to different factors, which also provides insights for explaining the recommendation diversification: *the factors with more perturbations have more potential to influence both the accuracy and diversity.* Finally, as there is a growing need to achieve a better trade-off between accuracy and diversity, we redesign the optimization objective considering accuracy and diversity metrics simultaneously. Overall, our proposed approach offers a more flexible and adaptable framework that can optimize various diversity metrics directly, and provides a promising solution to the explanation of recommendation diversification.

To summarize, the contributions of this work are as follows:

- To explain recommendation diversification, we employ the counterfactual framework to discover the meaningful factors that affect recommendation accuracy and diversity trade-off.
- To optimize a range of differentiable and non-differentiable diversity metrics, we propose a bandit-based diversity optimization approach that is agnostic to diversity metrics and recommendation models.
- To validate the applicability, effectiveness, and explainability of our method, we conducted extensive experiments on multiple real-world datasets and diversity metrics.

## 2 Related Work

### 2.1 Recommendation Diversification

To address the "filter bubble" and reduced provider engagement issues, it is crucial to recommend accurate and diverse items for a healthier online marketplace. Existing diversification methods are typically offline and categorized as pre-processing, in-processing, and post-processing methods [31]. Pre-processing methods [7,40] involve preparing interaction data before the model training. In-processing integrates diversity into the training process, using it as regularization [5,29] or a ranking score [17,33]. Post-processing, the most scalable, includes greedy-based methods like MMR [3,22,27] and DPP [4,14,30], which adjust item selection and rankings to balance relevance and diversity, and refinement-based methods [26], which modify positions or replace items based on diversity metrics. Other online methods, such as bandit strategies [9], treat diversity as part of the score on each arm (item or topic) in the bandit recommendation algorithms and reinforcement learning [23,39], continuously update based on user feedback for long-term optimization.

Although these methods enhance recommendation diversity, they do not provide explanations of the monopoly phenomenon of recommended items or the mechanisms behind their diversity improvements. Our work seeks to optimize recommendation diversity and offer explanations for these issues.

### 2.2 Explainable Recommendation

Explainable recommendations have attracted significant attention in academia and industry, aiming to enhance transparency, user satisfaction, and trust [28,36,37,38]. Early methods focused on generating individualized explanations, often customizing models and using auxiliary information [32,38]. For example, the Explicit Factor Model (EFM) [38] recommends products based on features extracted from user reviews. Other approaches decouple explanations from the recommendation model, making them post-hoc and model-agnostic [6,24]. Recently, counterfactual reasoning has been widely used to improve explainability. For instance, CEF [11] uses counterfactual reasoning to explain fairness in feature-aware recommendation systems.

This work focuses on explaining recommendation diversification. While existing approaches help interpret recommendation models, they overlook diversity, which is the main focus of our work.

## 3 Methodology

### 3.1 Preliminaries

**Problem Formulation** Given a user set $\mathcal{U}$, an item set $\mathcal{V}$, and the corresponding user-item interactions set $\mathcal{T}$, the purpose of explainable diversification is to recommend accurate and also diverse items that meet user interests, while offering explainability to the diversification. Formally, we need to provide the diverse top-$K$ recommendation list $R^u \subset \mathcal{V}(|R^u| = K)$ to each user $u$, and analyze what leads to the diversified results.

**Base Recommendation Models** Given the user latent feature matrix $\mathbf{P} \in \mathbb{R}^{d\times|\mathcal{U}|}$ and item latent feature matrix $\mathbf{Q} \in \mathbb{R}^{d\times|\mathcal{V}|}$, where $d$ is the dimension of the latent feature matrices. We define a base recommendation model $g$ that predicts the user-item ranking score $\hat{y}_{u,v}$ for user $u$ and item $v$ by:

$$\hat{y}_{u,v} = g(\mathbf{p}_u, \mathbf{q}_v \mid \mathbf{Z}, \mathbf{\Theta}), \tag{1}$$

where $\mathbf{p}_u \in \mathbb{R}^d$ and $\mathbf{q}_v \in \mathbb{R}^d$ are the latent feature vector of user $u$ and item $v$, respectively. The symbol $\mathbf{\Theta}$ denotes the model parameters, and $\mathbf{Z}$ represents all other auxiliary information. Since collaborative filtering (CF) methods are still mainstream in current recommendation systems, we mostly work on the factors with latent features of CF methods. Without loss of generality, we can also target the raw features (e.g., age, gender, etc.), which will be discussed in Sec. 4.4. We explore two popular and effective instances of $g$: **BPRMF** [21] and **LightGCN** [13]. The loss function for the base model adopts the Bayesian Personalized Ranking loss function.

**Diversity Metrics** Among all diversity metrics, we discuss the following four most popular metrics [7,31]. **Novelty-biased Normalized Discounted Cumulative Gain ($\alpha$-nDCG)** [8], which is a subtopic-level metric derived from NDCG, accounting for subtopics and item redundancy, where $\alpha$ applies geometric penalization for redundancy. **Subtopic Coverage (SC)** [10], which is a subtopic-level coverage of a recommended item list $R^u$ in the whole item set. **Prediction Coverage (PC)** [10], which is an item-level coverage of all recommendation lists $R^u$ in the whole item set. **Intra-List Average Distance (ILAD)** [34], which is an item-level metric that measures diversity by averaging the dissimilarity between item pairs in the recommendation list $R^u$. We use cosine similarity for dissimilarity calculation.

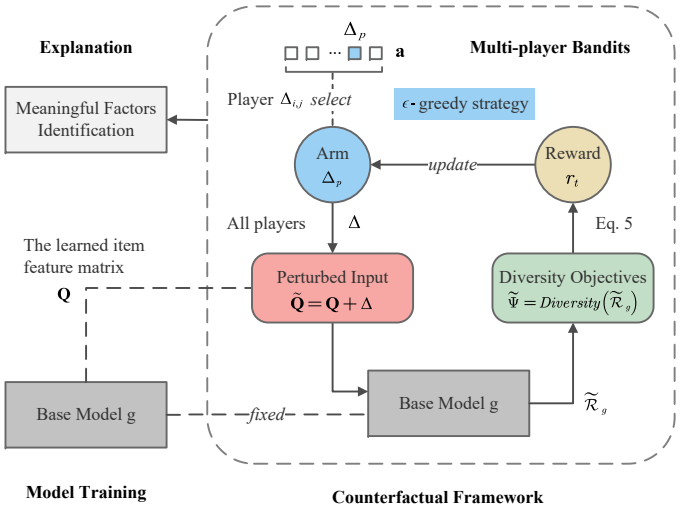


**Fig. 1.** The architecture of CMB. CMB consists of three major stages: the first stage of base model training, the second stage of counterfactual framework within multi-player bandits optimization, and the third stage of explanation. The $\mathbf{Q}$ is the learned item latent feature matrix from the base model $g$. The $\widetilde{\Psi}$ in the green part is alterable, which can be diversity or the trade-off between diversity and accuracy.

### 3.2 Counterfactual Framework for Explainable Diversification

Current diversification approaches generate diverse lists that are hard to explain and control. However, understanding the underlying diversity mechanism is crucial for making intelligent decisions in real-world applications. Inspired by counterfactual reasoning [11], we develop a perturbation-based framework for explaining the diversification of the recommendation lists.

The essential idea behind the proposed explanation model is to discover a perturbation matrix $\Delta$ on items' factors by solving a counterfactual optimization problem that maximizes diversity, as well as identify which factors are the underlying drive of diversified recommendations. After identifying these factors, it is easy to generate feature-based explanations for the given recommendation model $g$ and guide the system to make appropriate decisions that increase the recommendation diversity. Generally, given a recommendation model $g$, we have a certain recommendation result $\mathcal{R}_g = \{R^{u_1}, R^{u_2}, \cdots, R^{u_i}, \cdots, R^{u_{|\mathcal{U}|}}\}(|R^{u_i}| = K, i = 1, 2, \cdots, |\mathcal{U}|)$ containing all users' top-$K$ recommendation lists, where $R^{u_i}$ represents the top-$K$ items list recommended to user $u_i$ by the base model $g$. We denote the recommendation diversity of $g$ as,

$$\Psi = Diversity(\mathcal{R}_g), \tag{2}$$

where $Diversity(\cdot)$ can be any of the previously introduced diversity measurements in Sec. 3.1.

Specifically, for the learned item latent feature matrix $\mathbf{Q} \in \mathbb{R}^{d \times |\mathcal{V}|}$ from $g$, we slightly intervene with an equal-size matrix $\Delta \in \mathbb{R}^{d \times |\mathcal{V}|}$. In detail, a small perturbation $\Delta_{i,j}$ will be added to feature $i$ of item $j$ ($\mathbf{Q}_{i,j}$) to obtained the perturbed input $\widetilde{\mathbf{Q}}$. That is,

$$\widetilde{\mathbf{Q}} = \mathbf{Q} + \Delta. \tag{3}$$

With this perturbed item latent feature matrix $\widetilde{\mathbf{Q}}$, the base model $g$ will change the recommendation from $\mathcal{R}_g$ to a new counterfactual result $\widetilde{\mathcal{R}}_g$ with a new diversity measure $\widetilde{\Psi}$,

$$\widetilde{\Psi} = Diversity(\widetilde{\mathcal{R}}_g). \tag{4}$$

Here, our goal is to find the minimum intervention on item factors that will result in the maximum improvement in terms of diversity. Thus, the objective function would be:

$$\max_{\Delta} \|\widetilde{\Psi}\|_2^2 - \lambda_1 \|\Delta\|_1, \tag{5}$$

where $\lambda_1$ is a hyperparameter that controls the balance between two terms: the first maximizes the predefined diversity, and the second constrains the perturbation by reflecting the distance between the original input and the counterfactuals. To minimize changes in item factors, we apply $L_1$ norm constraint on $\Delta$ and scale its absolute values between $[0, 1]$, encouraging more $\Delta$ as 0 and highlighting the factors that most influence diversity.

### 3.3 Multi-player Bandits for Diversity Optimization

According to certain needs of the diversity, various metrics can be used to optimize Eq. 5 for learning the perturbation $\Delta$. For example, $\alpha$-nDCG or SC metric can be employed to ensure broader coverage of subtopics in the recommendation list, while ILAD or PC metric can be used to enhance item-level diversity.

However, a significant challenge is that most diversity metrics are non-differentiable, making it difficult to define a proxy for their optimization [31,33]. For instance, among the four metrics discussed in Sec. 3.1, only ILAD is differentiable, complicating the integration of non-differentiable metrics into gradient-based counterfactual frameworks. To overcome this, we propose a bandit-based method to learn the perturbation matrix $\Delta$, enabling the optimization of diverse objectives without relying on gradient computation.

The multi-armed bandit problem [2] models decision-making under uncertain rewards, where a player chooses among various options ("arms") to maximize cumulative payoff by balancing exploration and exploitation. This approach is well-suited for optimizing non-differentiable objectives within a counterfactual framework. Therefore, in our specific problem, each item feature is treated as a player, selecting an appropriate arm to construct the final perturbation matrix. Moreover, our problem can be further conceptualized as a multi-player bandit scenario [1], where players collaborate to optimize a shared reward, which serves as the objective in the counterfactual framework.

Specifically, we treat each variable $\Delta_{i,j}$ in $\Delta$ as a player; for simplicity, we denote it as $p(|p| = d \times |\mathcal{V}|)$, and then select a suitable arm $\Delta_p$ from arms $\mathbf{a}$ for

every $p$ iteratively to maximize the objective in the counterfactual framework. Assume the total number of iterations for one player $p$ to select an arm is $T$; our problem can be formulated by maximizing the following $T$-step cumulative reward for each player:

$$V^T = \max \sum_{t=1}^{T} r_t, \tag{6}$$

where $r_t$ is the reward at the iteration step $t$ obtained from all players by selecting arms, as calculated by Eq. 5. In each iteration, every player selects a particular arm and gets $\Delta$ from all players to calculate the reward. This process continues as players select arms in subsequent iterations based on the obtained rewards. The iterative process repeats until the final iteration, where each player selects the optimal arm to achieve diversification. At this point, the value of the perturbation $\Delta$ can be justified.

More specifically, before commencing the algorithm, it is necessary to define the arm values from which the players can make their selections. To achieve this, we utilize the following method to initialize the arms $\mathbf{a}$ for each player $p$:

$$\mathbf{a} = \texttt{INIT}(A, n_A), \tag{7}$$

where $A$ and $n_A$ represent the perturbation threshold and the number of arms, respectively. The $\texttt{INIT}(\cdot)$ method samples $n_A$ values evenly from $[-A, A]$. For simplicity, this initialization is applied to each player, and the item latent feature matrix $\mathbf{Q}$ is scaled to $[-1, 1]$ using maximum absolute scaling ($\mathbf{Q} = \mathbf{Q}/|max(\mathbf{Q})|$). At each iteration $t$, players independently choose an arm $\Delta_p$ from $\mathbf{a}$ based on arm selection strategies like $\epsilon$-greedy or UCB [16,25], with experiments showing that $\epsilon$-greedy is more efficient and effective. The selection strategy of $\epsilon$-greedy strategy is as follows:

$$\Delta_p^t = \begin{cases} \arg\max_{\Delta_p \in \mathbf{a}}(V_{\mathbf{a}}^t) & \text{with probability } 1-\epsilon, \\ \text{a random arm} & \text{with probability } \epsilon, \end{cases} \tag{8}$$

where $\Delta_p^t$ and $V_{\mathbf{a}}^t$ are the arm value selected by the player and the cumulative reward vector containing all arms for the player in the $t$ iteration, respectively.

When all players have selected the arm $\Delta_p^t$ based on the arm selection strategy, we can get the $\Delta$ and further the reward $r_t$ (Eq. 5) based on the counterfactual result $\widetilde{\mathcal{R}}_g$. To reduce the computational complexity, we update the cumulative reward value $V_{\Delta_p}^{t+1}$ of the corresponding arm $\Delta_p$ selected by each

player $p$ in iteration $t+1$ by an incremental average method:

$$\begin{aligned} V_{\Delta_p}^{t+1} &= \frac{1}{n}\sum_{i=1}^{n} r_i, \\ &= \frac{1}{n}\left(r_n^t + \sum_{i=1}^{n-1} r_i\right), \\ &= \frac{1}{n}\left(r_t + (n-1)\,\frac{1}{n-1}\sum_{i=1}^{n-1} r_i\right), \\ &= \frac{1}{n}\left(r_t + (n-1)\,V_{\Delta_p}^t\right), \\ &= \frac{1}{n}\left(r_t + nV_{\Delta_p}^t - V_{\Delta_p}^t\right), \\ &= V_{\Delta_p}^t + \frac{1}{n}\left(r_t - V_{\Delta_p}^t\right), \end{aligned} \tag{9}$$

where $n$ is the times that arm $\Delta_p$ has been selected by player $p$ till iteration $t$. Thus, the model iteratively learns and adjusts the $\Delta$ until convergence.

The challenge in diversified recommendation is to enhance diversity while preserving accuracy, i.e., maximizing diversity without significantly compromising accuracy. To better balance these two aspects, we propose a redesign of the optimization objective (Eq. 5), particularly the counterfactual diversity measurement $\widetilde{\Psi}$. As previously discussed, while a suitable diversity metric for $\widetilde{\Psi}$ can be chosen, it often leads to some loss in accuracy. To achieve an optimal balance, inspired by [3,4], we redesign $\widetilde{\Psi}$ to balance both accuracy and diversity simultaneously. Specifically,

$$\widetilde{\Psi} = \lambda_2 \times Accuracy(\widetilde{\mathcal{R}}_g) + (1-\lambda_2) \times Diversity(\widetilde{\mathcal{R}}_g), \tag{10}$$

where $\lambda_2$ is a hyperparameter to control the trade-off between accuracy and diversity, and $Accuracy(\cdot)$ and $Diversity(\cdot)$ represent accuracy metrics (e.g., Recall@K, NDCG@K, etc.) and diversity metrics (Sec. 3.1), respectively.

**Time Complexity Analysis** In the $\epsilon$-greedy strategy for multi-armed bandits, arm selection and incremental reward updates both have $O(1)$ time complexity in each step. However, if the reward update uses a complex equation (e.g., Eq. 5), the overall time complexity in each step will depend on that equation.

**Discussion** 1) We mainly work on the latent features of items to explain the recommendation diversification. The key consideration is that current mainstream recommendation models still originate from collaborative filtering methods, which are based on latent features. These motivate us to work on the latent features for controlling the diversity level of recommendation results. It is worth noting that our model can both work raw and latent features. 2) The main reason why we decided to apply the perturbation $\Delta$ to the features of items is that

perturbing at the items' feature level enables us to identify specific features that impact the diversity of the model at a more fine-grained feature level. 3) The multi-armed bandits method, unlike previous approaches, is primarily utilized in online recommendation scenarios. In our study, we employ this method to optimize the learning objective in the counterfactual framework, especially targeting and optimizing the non-differentiable diversity metrics directly.

### 3.4 Meaningful Factors Identification as Explanation

Once finishing optimization, we get the "minimal" changes $\Delta$ and the corresponding recommendation results under such changes. The values of $\Delta$ indicate the influence of item factors on the accuracy-diversity trade-off of the recommendation lists generated by the base model $g$. Specifically, compared with the initial item latent feature matrix $\mathbf{Q}$, after adding the values of $\Delta$, the model $g$ is supposed to generate more diverse lists. Therefore, the perturbation $\Delta$ provides insights for our explanation. In particular, larger absolute values of $\Delta$ correspond to a greater need for the corresponding factors to promote greater diversity.

Based on the above analysis, after we identify each factor's "ability" to incur the diversity of the recommendation list, we further select the most meaningful factors of the items affecting diversity and give insights into recommendation systems. We provide two perspectives on detecting the most meaningful factors here, namely CMB-Individual and CMB-Shared.

$$\begin{cases} \text{CMB-Individual}, & \textit{feature-level} \\ \text{CMB-Shared}, & \textit{item-level} \end{cases}$$

Specifically, the strategy of CMB-Individual is to directly select the factors corresponding to the higher absolute values of $\Delta$ on the factors as an explanation for each user. The CMB-Shared strategy is that we take the absolute values of $\Delta$, compute the mean value by rows, and compress the $\Delta$ into a vector $\Delta_v \in \mathbb{R}^d$,

$$\Delta_v = MEAN(|\Delta|, dim = 0), \tag{11}$$

and then choose the factors corresponding to the higher values of $\Delta_v$ as an explanation. After discovering the most meaningful factors, we can adjust the values of these factors to meet the corresponding needs of diversity.

### 3.5 Overall Procedure

The entire procedure contains three stages, as shown in Fig. 1. In the first stage, the base model $g$ introduced above will be trained. In the second stage, the counterfactual framework is first constructed. Based on this, the bandit algorithm is used to learn the perturbations by optimizing the diversity of the recommended top-$K$ lists. Our framework is model-agnostic and applicable to any recommendation model $g$. Meanwhile, our model is metric-agnostic since the optimization objective can be any diversity metric. In the final stage, two strategies are utilized to discover the most meaningful factors for recommendation diversification.

**Table 1.** The statistics of datasets.

| Dataset | #User | #Item | #Subtopic | #Interaction | Density |
|---|---|---|---|---|---|
| *ML1M* | 5,950 | 3,125 | 18 | 573,726 | 0.0309 |
| *ML10M* | 51,692 | 7,135 | 19 | 4,752,578 | 0.0129 |
| *CDs* | 13,364 | 29,294 | 30 | 371,204 | 0.0009 |

## 4 Experiments

In this section, we mainly focus on the following questions:

- **RQ1:** Does our method get better recommendation diversification effects than the state-of-art methods? Especially the trade-off between recommendation accuracy and diversity.
- **RQ2:** Do the selected top features play a significant role in diversification performance?
- **RQ3:** Are our generated feature-level diversification explanations reasonable and intuitive in real cases?

Given the limited space, for a more detailed experiment setup and the results, we invite the reader to check out the Appendix for the supplementary material.

### 4.1 Experiment Setup

**Datasets** We performed experiments on three widely used real-world datasets, *MovieLens 1M* [12] (*ML1M*), *MovieLens 10M* [12] (*ML10M*), and *Amazon CDs and Vinyl* [19] (*CDs*) to evaluate the models under different data scales and application scenarios. The statistics of the datasets are shown in Table 1.

For all datasets, we convert ratings to implicit feedback, treating ratings no less than four (out of five) as positive and all other ratings as missing entries. To optimize base models, we randomly sample 3 negative instances for each user's positive interaction. For the *CDs* dataset, we use the top 30 categories with the highest frequency as subtopic information according to the metadata. Each dataset is split 8:1:1 for training, test, and validation. We independently run all models five times and report the average results.

**Baselines** To verify the effectiveness of our proposed CMB method, we compare it with the following representative baselines. Two vanilla recommendation models **BPRMF** [21] and **LightGCN** [13], which are introduced in Sec. 3.1. Two recommendation diversification methods **MMR** [3] and **DPP** [4]. In addition, we also explore the method $\mathbf{CMB}^{\mathbf{Gradient}}$, which represents our method is directly optimized for differentiable metrics (e.g., ILAD) by using the gradient method instead of the proposed bandit method.

**Evaluation Metrics** In all experiments, we evaluate the recommendation performance using accuracy and diversity metrics on Top-$K$ ($K$ = 10, 20) lists, where $K$ represents the list length as discussed earlier during training. For all metrics, the higher the value is, the better the performance is. **Accuracy:** we evaluate the accuracy of the ranking list using Recall@K and NDCG@K. **Diversity:** we evaluate the recommendation diversity by the $\alpha$-nDCG@K, SC@K, PC@K, and ILAD@K introduced previously.

**Implementation Details** In our experiments, we employ BPRMF and LightGCN as the base model $g$, with LightGCN set to three layers of the graph neural network. For baselines, BPRMF calculates the relevance scores for MMR and constructs the kernel matrix for DPP. The trade-off parameter for MMR and DPP is empirically set to 0.9 after testing values from {0.1, 0.3, 0.5, 0.7, 0.9}, prioritizing guaranteeing the optimal accuracy performance as much as possible. All models have a latent feature dimension of 50, with $K$ set to 20 for counterfactual learning. The $\alpha$ in $\alpha$-nDCG is set to 0.5 same as [17,22,27,33], and model parameters are optimized by Adam [15] with a learning rate of 0.005. In the bandit algorithm, we set the threshold of arm values $A$ to 0.3, the number of arms $n_A$ to 61, and $\epsilon$ to 0.1. The hyperparameters $\lambda_1$, $\lambda_2$, and the total number of iterations $T$ are set to 5, 0.9, and 200, respectively. The source code is available at `https://github.com/Forrest-Stone/CMB`.

### 4.2 Performance Comparison (RQ1)

Tables 2, 3, and 4 compare the performance of different methods under two base models. Our model is denoted as $\text{CMB}_{\text{BPRMF}}$ when using BPRMF, and $\text{CMB}_{\text{LightGCN}}$ when using LightGCN. The * in CMB-* represents the specific objective $\widetilde{\Psi}$ optimized in Eq. 5. For instance, CMB-$\alpha$-nDCG means that we adopt the $\alpha$-nDCG metric for $\widetilde{\Psi}$, while CMB-$\alpha$-nDCG-NDCG means that we use the trade-off optimization objective (Eq. 10) of $\alpha$-nDCG and NDCG. When optimizing a single diversity metric, $\lambda_1$ is set to 0 for optimal results. BPRMF is the default base model $g$ unless otherwise specified. The observations from Tables 2, 3, and 4 are as follows.

**Trade-off observations**. First, as shown in Table 2, CMB achieves a better balance between accuracy and diversity than other methods. While diversification methods like DPP improve diversity performance, they often significantly compromise accuracy performance. For example, DPP increases ILAD@10 on *ML1M* by 83.73% but causes a 75.43% drop in Recall@10, which is counterproductive to the primary goal of recommendation systems. In contrast, our method achieves an acceptable balance between accuracy and diversity. For instance, $\text{CMB}_{\text{BPRMF}}$-SC-Recall increases ILAD@10 by 11.74% on *ML1M* and 8.48% on *ML10M*, with only a 5.05% and 3.81% reduction in Recall@10, respectively. Similar trends are observed in other CMB variants, as evidenced in Tables 2 and 3, demonstrating the effectiveness of our trade-off objective.

**Table 2.** Comparisons of the accuracy and diversity performance. The base model $g$ here adopts BPRMF. The bold scores are the best in each column, and the underlined scores are the second best. The symbols ↑ and ↓, along with their preceding values, represent the percentage (% is omitted) improvement and decrease of a given method in the corresponding metric, in comparison to the base model $g$.

| Metric | Recall@10 | NDCG@10 | $\alpha$-nDCG@10 | SC@10 | PC@10 | ILAD@10 |
|---|---|---|---|---|---|---|
| | | | *ML1M* | | | |
| BPRMF | **0.1465** | **0.2742** | 0.7035 | 0.4993 | 0.3206 | 0.2010 |
| MMR | 0.0441 (69.90↓) | 0.0741 (72.98↓) | 0.6980 (0.78↓) | 0.4692 (6.03↓) | 0.0970 (69.74↓) | 0.1709 (14.98↓) |
| DPP | 0.0360 (75.43↓) | 0.0689 (74.87↓) | **0.7186** (2.15↑) | **0.5558** (11.32↑) | **0.4554** (42.05↑) | **0.3693** (83.73↑) |
| $\text{CMB}_{\text{BPRMF}}$-$\alpha$-nDCG-Recall | 0.1388 (5.26↓) | 0.2588 (5.62↓) | 0.7094 (0.84↑) | 0.5097 (2.08↑) | 0.3291 (2.65↑) | <u>0.2253</u> (12.09↑) |
| $\text{CMB}_{\text{BPRMF}}$-SC-Recall | <u>0.1391</u> (5.05↓) | <u>0.2594</u> (5.40↓) | 0.7078 (0.61↑) | 0.5102 (2.18↑) | 0.3307 (3.15↑) | 0.2246 (11.74↑) |
| $\text{CMB}_{\text{BPRMF}}$-PC-NDCG | 0.1388 (5.26↓) | 0.2584 (5.76↓) | 0.7093 (0.82↑) | 0.5107 (2.28↑) | 0.3286 (2.50↑) | 0.2247 (11.79↑) |
| $\text{CMB}_{\text{BPRMF}}$-ILAD-NDCG | 0.1387 (5.32↓) | 0.2587 (5.65↓) | <u>0.7109</u> (1.05↑) | <u>0.5127</u> (2.68↑) | <u>0.3313</u> (3.34↑) | 0.2246 (11.74↑) |
| | | | *ML10M* | | | |
| BPRMF | **0.1549** | **0.2648** | 0.7043 | 0.5483 | 0.2453 | 0.1886 |
| MMR | 0.0402 (74.05↓) | 0.0602 (77.27↓) | 0.7095 (0.74↑) | 0.5155 (5.98↓) | 0.0416 (83.04↓) | 0.1623 (13.94↓) |
| DPP | 0.0253 (83.67↓) | 0.0541 (79.57↓) | 0.6977 (0.94↓) | **0.6072** (10.74↑) | **0.3735** (52.26↑) | **0.3764** (99.58↑) |
| $\text{CMB}_{\text{BPRMF}}$-$\alpha$-nDCG-Recall | 0.1488 (3.94↓) | 0.2531 (4.42↓) | **0.7126** (1.18↑) | 0.5544 (1.11↑) | 0.2475 (0.90↑) | 0.2045 (8.43↑) |
| $\text{CMB}_{\text{BPRMF}}$-SC-Recall | <u>0.1490</u> (3.81↓) | <u>0.2535</u> (4.27↓) | <u>0.7123</u> (1.14↑) | 0.5544 (1.11↑) | <u>0.2477</u> (0.98↑) | 0.2046 (8.48↑) |
| $\text{CMB}_{\text{BPRMF}}$-PC-NDCG | 0.1487 (4.00↓) | 0.2532 (4.38↓) | <u>0.7123</u> (1.14↑) | <u>0.5546</u> (1.15↑) | 0.2471 (0.73↑) | <u>0.2051</u> (8.75↑) |
| $\text{CMB}_{\text{BPRMF}}$-ILAD-NDCG | 0.1486 (4.07↓) | 0.2527 (4.57↓) | 0.7110 (0.95↑) | 0.5541 (1.06↑) | 0.2460 (0.29↑) | 0.2050 (8.70↑) |
| | | | *CDs* | | | |
| BPRMF | **0.0515** | **0.0457** | <u>0.7206</u> | 0.1700 | 0.1665 | 0.2332 |
| MMR | 0.0033 (93.59↓) | 0.0032 (93.00↓) | **0.7240** (0.47↑) | 0.1705 (0.29↑) | 0.0247 (85.17↓) | 0.2372 (1.72↑) |
| DPP | 0.0115 (77.67↓) | 0.0128 (71.99↓) | 0.7116 (1.25↓) | **0.2409** (41.71↑) | **0.3261** (95.86↑) | **0.4013** (72.08↑) |
| $\text{CMB}_{\text{BPRMF}}$-$\alpha$-nDCG-NDCG | <u>0.0477</u> (7.38↓) | <u>0.0422</u> (7.66↓) | 0.7183 (0.32↓) | <u>0.1739</u> (2.29↑) | <u>0.1825</u> (9.61↑) | 0.2511 (7.68↑) |
| $\text{CMB}_{\text{BPRMF}}$-SC-NDCG | 0.0475 (7.77↓) | 0.0421 (7.88↓) | 0.7192 (0.19↓) | 0.1736 (2.12↑) | 0.1824 (9.55↑) | 0.2510 (7.63↑) |
| $\text{CMB}_{\text{BPRMF}}$-PC-Recall | 0.0476 (7.57↓) | 0.0421 (7.88↓) | 0.7180 (0.36↓) | 0.1737 (2.18↑) | 0.1816 (9.07↑) | 0.2509 (7.59↑) |
| $\text{CMB}_{\text{BPRMF}}$-ILAD-Recall | <u>0.0477</u> (7.38↓) | <u>0.0422</u> (7.66↓) | 0.7189 (0.24↓) | 0.1736 (2.12↑) | 0.1823 (9.49↑) | <u>0.2513</u> (7.76↑) |

Second, not only does the combined optimization objective help achieve a reasonable balance between accuracy and diversity, but also the single diversity objective does. For instance, as shown in Table 4, when compared to BPRMF, $\text{CMB}_{\text{BPRMF}}$-ILAD shows a decrease of 18.79% in Recall@10 and an increase of 40.62% in ILAD@10 on *ML10M*, while showing a decrease of 31.29% in NDCG@10 and an increase of 59.58% in PC@10 on *CDs*. These results demonstrate that the proposed single diversity objective can improve diversity performance while maintaining accuracy performance, and outperforming other diversity methods like MMR and DPP.

Finally, compared with the single diversity objective optimized by CMB, the combined optimization objective also achieves a better balance between accuracy and diversity. For example, according to Table 2 and 4, on *ML10M* and *CDs*, the Recall@10 of $\text{CMB}_{\text{BPRMF}}$-SC-NDCG is 19.04% and 35.71% higher than $\text{CMB}_{\text{BPRMF}}$-SC, while the SC@10 only decreases by 5.06% and 10.33%, respectively. Similar findings can also be found in other cases. These results demonstrate the superiority of our proposed trade-off target in achieving a balance between accuracy and diversity.

**Diversification observations**. First, the CMB model can effectively optimize various diversity metrics while yielding satisfactory results. For instance, as shown in Table 2 and 4, $\text{CMB}_{\text{BPRMF}}$-$\alpha$-nDCG outperforms the best baseline

**Table 3.** Comparisons of the accuracy and diversity performance. The base model $g$ here adopts LightGCN. The bold scores are the best in each column, and the underlined scores are the second best. The symbols ↑ and ↓, along with their preceding values, represent the percentage (% is omitted) improvement and decrease of a given method in the corresponding metric, in comparison to the base model $g$.

| Metric | Recall@10 | NDCG@10 | $\alpha$-nDCG@10 | SC@10 | PC@10 | ILAD@10 |
|---|---|---|---|---|---|---|
| | | | *ML10M* | | | |
| LightGCN | **0.1724** | **0.2912** | 0.7056 | **0.5680** | 0.1979 | 0.1480 |
| $\text{CMB}_{\text{LightGCN}}$-$\alpha$-nDCG-NDCG | <u>0.1706</u> (1.04↓) | <u>0.2866</u> (1.58↓) | 0.7061 (0.07↑) | 0.5655 (0.44↓) | <u>0.2131</u> (7.68↑) | <u>0.1562</u> (5.54↑) |
| $\text{CMB}_{\text{LightGCN}}$-SC-NDCG | <u>0.1706</u> (1.04↓) | 0.2865 (1.61↓) | 0.7060 (0.06↑) | 0.5650 (0.53↓) | **0.2142** (8.24↑) | **0.1565** (5.74↑) |
| $\text{CMB}_{\text{LightGCN}}$-PC-Recall | 0.1704 (1.16↓) | <u>0.2866</u> (1.58↓) | <u>0.7062</u> (0.09↑) | <u>0.5664</u> (0.28↓) | 0.2122 (7.23↑) | <u>0.1562</u> (5.54↑) |
| $\text{CMB}_{\text{LightGCN}}$-ILAD-Recall | 0.1705 (1.10↓) | <u>0.2866</u> (1.58↓) | **0.7072** (0.23↑) | 0.5661 (0.33↓) | 0.2130 (7.63↑) | 0.1561 (5.47↑) |
| | | | *CDs* | | | |
| LightGCN | **0.0567** | **0.0500** | **0.7260** | 0.1616 | 0.0931 | 0.1659 |
| $\text{CMB}_{\text{LightGCN}}$-$\alpha$-nDCG-Recall | <u>0.0554</u> (2.29↓) | <u>0.0490</u> (2.00↓) | <u>0.7240</u> (0.28↓) | <u>0.1643</u> (1.67↑) | **0.0938** (0.75↑) | **0.1751** (5.55↑) |
| $\text{CMB}_{\text{LightGCN}}$-SC-Recall | <u>0.0554</u> (2.29↓) | 0.0489 (2.20↓) | 0.7238 (0.30↓) | 0.1642 (1.61↑) | 0.0935 (0.43↑) | **0.1751** (5.55↑) |
| $\text{CMB}_{\text{LightGCN}}$-PC-NDCG | 0.0553 (2.47↓) | 0.0489 (2.20↓) | 0.7238 (0.30↓) | **0.1644** (1.73↑) | 0.0934 (0.32↑) | 0.1748 (5.36↑) |
| $\text{CMB}_{\text{LightGCN}}$-ILAD-NDCG | 0.0552 (2.65↓) | 0.0488 (2.40↓) | 0.7237 (0.32↓) | 0.1642 (1.61↑) | <u>0.0936</u> (0.54↑) | <u>0.1750</u> (5.49↑) |

by 2.26% in $\alpha$-nDCG@10 on *ML10M*. Additionally, from these two tables, we also observe that $\text{CMB}_{\text{BPRMF}}$ significantly improves all diversity metrics compared to the base model BPRMF when optimized individually.

Second, as illustrated in Tables 2 and 4, $\text{CMB}_{\text{BPRMF}}$-SC achieves the second-best SC@10 on *ML10M*, only 0.12% decrease compared to the best baseline. Similarly, $\text{CMB}_{\text{BPRMF}}$-PC also achieves the second-best performance in the PC metric on *ML10M* and *CDs*. These results demonstrate that CMB can achieve a satisfactory diversification. Moreover, while CMB sometimes trails DPP in diversity metrics, this is because DPP prioritizes diversity over accuracy. In contrast, our method balances both, leading to improved accuracy even if diversity scores are slightly lower when compared to the DPP.

**Application observations**. First, from Table 4, by comparing $\text{CMB}^{\text{Gradient}}_{\text{BPRMF}}$-ILAD and $\text{CMB}_{\text{BPRMF}}$-ILAD, we observe that the gradient descent optimization method outperforms the bandit optimization method in terms of accuracy on *CDs*, while exhibiting better diversity results on *ML10M* for PC and ILAD metrics. Thus, the choice of which optimization method should be based on the specific application scenario (e.g., dataset, diversity level, etc.).

Second, as shown in Table 2 and 4, compared to $\text{CMB}_{\text{Random}}$, CMB that only optimizes a single diversity metric also achieves good results on that metric, which shows that our bandit method is effective for optimizing different diversity metrics. Moreover, optimizing the combination of trade-off objectives simultaneously also achieves a good balance in accuracy and diversity, further highlighting the effectiveness of our combined optimization objectives. For example, NDCG@10 of $\text{CMB}_{\text{BPRMF}}$-SC-NDCG on *ML10M* and *CDs* is 18.88% and 34.08% higher than $\text{CMB}_{\text{BPRMF}}$-Random, respectively, while the SC@10 only decreases by 1.35% and 6.16%.

**Other observations.** First, the accuracy and diversity trade-off exists widely. No method can achieve the best results in both accuracy and diversity since

**Table 4.** Comparisons among the accuracy and diversity performance of CMB optimizes the single diversity metrics. The base model $g$ here adopts BPRMF. $\text{CMB}_{\text{ILAD}}^{\text{Gradient}}$ represents CMB that directly optimizes the differentiable metric ILAD by using the gradient method. $\text{CMB}_{\text{BPRMF}}$-Random represents CMB that chooses the arm randomly for each player. The bold scores are the best in each column.

| Metric | Recall@K | | NDCG@K | | $\alpha$-nDCG@K | | SC@K | | PC@K | | ILAD@K | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | K=10 | K=20 | K=10 | K=20 | K=10 | K=20 | K=10 | K=20 | K=10 | K=20 | K=10 | K=20 |
| *ML10M* | | | | | | | | | | | | |
| $\text{CMB}_{\text{BPRMF}}$-Random | 0.1266 | 0.2032 | 0.2129 | 0.2217 | 0.7108 | 0.8045 | 0.5630 | 0.6930 | 0.2625 | 0.3514 | 0.2584 | 0.2787 |
| $\text{CMB}_{\text{BPRMF}}$-$\alpha$-nDCG | **0.1267** | **0.2033** | **0.2138** | **0.2223** | **0.7467** | **0.8336** | 0.5794 | 0.7002 | 0.2620 | 0.3500 | 0.2554 | 0.2757 |
| $\text{CMB}_{\text{BPRMF}}$-SC | 0.1250 | 0.2024 | 0.2106 | 0.2201 | 0.7436 | 0.8295 | **0.6065** | **0.7253** | 0.2607 | 0.3496 | 0.2542 | 0.2751 |
| $\text{CMB}_{\text{BPRMF}}$-PC | 0.1259 | 0.2018 | 0.2135 | 0.2216 | 0.7104 | 0.8036 | 0.5726 | 0.7012 | 0.2714 | 0.3598 | 0.2582 | 0.2784 |
| $\text{CMB}_{\text{BPRMF}}$-ILAD | 0.1258 | 0.2021 | 0.2121 | 0.2209 | 0.7147 | 0.8073 | 0.5650 | 0.6928 | 0.2651 | 0.3556 | 0.2652 | 0.2843 |
| $\text{CMB}_{\text{BPRMF}}^{\text{Gradient}}$-ILAD | 0.1174 | 0.1914 | 0.1982 | 0.2086 | 0.6992 | 0.7948 | 0.5003 | 0.6279 | **0.3089** | **0.3780** | **0.2900** | **0.3034** |
| *CDs* | | | | | | | | | | | | |
| $\text{CMB}_{\text{BPRMF}}$-Random | 0.0353 | 0.0582 | 0.0314 | 0.0397 | 0.7053 | 0.8087 | 0.1850 | 0.2593 | 0.2659 | 0.3909 | 0.3081 | 0.3220 |
| $\text{CMB}_{\text{BPRMF}}$-$\alpha$-nDCG | 0.0358 | 0.0583 | 0.0319 | 0.0400 | 0.7150 | **0.8155** | 0.1866 | 0.2593 | 0.2621 | 0.3859 | 0.3071 | 0.3209 |
| $\text{CMB}_{\text{BPRMF}}$-SC | 0.0350 | 0.0579 | 0.0310 | 0.0393 | 0.7043 | 0.8074 | **0.1936** | **0.2684** | 0.2649 | 0.3884 | 0.3064 | 0.3205 |
| $\text{CMB}_{\text{BPRMF}}$-PC | 0.0347 | 0.0572 | 0.0310 | 0.0391 | 0.7050 | 0.8086 | 0.1862 | 0.2610 | **0.2755** | **0.4050** | **0.3117** | **0.3255** |
| $\text{CMB}_{\text{BPRMF}}$-ILAD | 0.0354 | 0.0585 | 0.0314 | 0.0397 | 0.7046 | 0.8083 | 0.1855 | 0.2598 | 0.2657 | 0.3885 | 0.3102 | 0.3233 |
| $\text{CMB}_{\text{BPRMF}}^{\text{Gradient}}$-ILAD | **0.0524** | **0.0846** | **0.0458** | **0.0574** | **0.7164** | 0.8154 | 0.1717 | 0.2353 | 0.1768 | 0.2438 | 0.2788 | 0.2866 |

an increase in accuracy generally corresponds to a decrease in diversity. From the results in Table 2, DPP achieves the best results in SC@10, PC@10, and ILAD@10 on *ML1M* and *ML10M*, but it achieves the worst performance in Recall@10 and NDCG@10 on these datasets.

Second, generally, no single method demonstrates superior performance across all diversity metrics. For example, as shown in Table 2, DPP has the highest SC@10, PC@10, and ILAD@10 performance on *ML10M* and *CDs*, but has the lowest $\alpha$-nDCG@K performance. This indicates the inherent gap between different diversity evaluation metrics, proving the necessity of optimizing different metrics in a general framework, which is just the focus of our work.

### 4.3 Validity Analysis of Explanations (RQ2)

As discussed in Sec.3.4, the values of $\Delta$ affect the diversity of recommendation lists generated by the base model $g$. To evaluate whether $\Delta$ can discover the meaningful factors that improve diversity or balance accuracy and diversity, we follow the widely deployed erasure-based evaluation criterion [11] from Explainable AI. Specifically, we erase the "most meaningful factors" from $\Delta$ (setting them to 0) and input this modified $\Delta$ into the pre-trained model $g$ to generate new recommendations. We then assess our model's effectiveness regarding the diversity and accuracy of these new results. We explore two erasure strategies, CMB-Individual (CMB-I) and CMB-Shared (CMB-S) – by erasing the top, least, or random $F$ factors, where $F$ is the number of erasing factors. For CMB-I, we erase the top/least/random-$F$ factors of each column of the absolute values of $\Delta$. For CMB-S, we average each row of the absolute values of $\Delta$, then erase the top/least/random-$F$ factors by row. Compared with the least/random manners

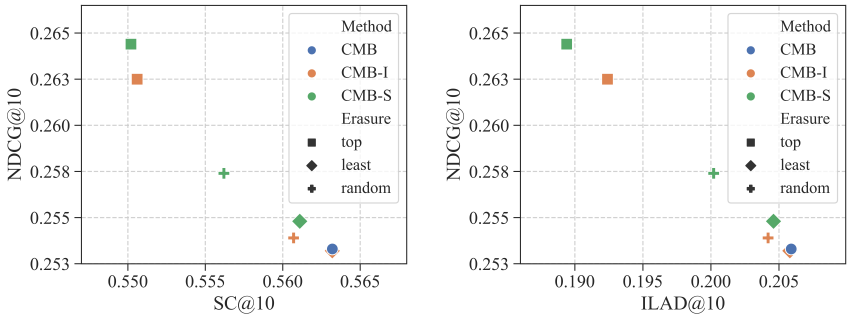


**Fig. 2.** Factor validity analysis on *ML10M* dataset when utilizing different erasure methods (CMB-I and CMB-S) with top/least/random manners.

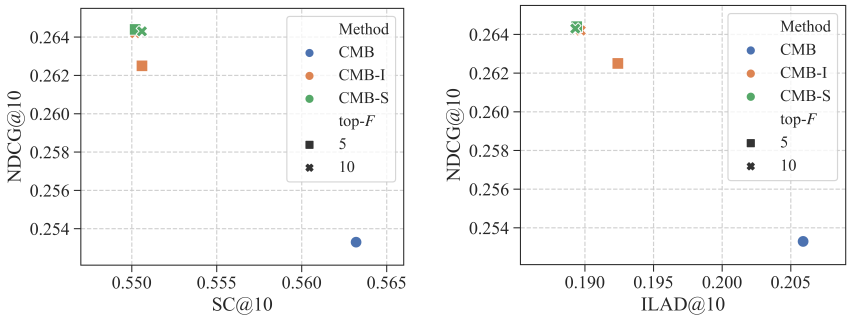


**Fig. 3.** Factor validity analysis on *ML10M* dataset when utilizing different erasure methods (CMB-I and CMB-S) with different $F$.

in Fig. 2, we observe that omitting these meaningful factors by the top manner reduces the diversity scores much while increasing the accuracy measures a lot. And the least manner does little to alter the performance of diversity or accuracy. Therefore, it verifies that the meaningful factors we discover can benefit the trade-off between diversity and accuracy of recommendation results. Furthermore, as shown in Fig. 3, the results of the CMB-Shared approach with top-$F$ ($F = 5/10$) are highly equivalent, indicating that the CMB-Shared approach can identify only a few factors that significantly impact the model's diversity or accuracy. Observations from other approaches are similar.

### 4.4 Case Study of Explanations (RQ3)

The purpose of the case study is to demonstrate the applicability of our method to both latent and raw features. As described in Sec. 3.4, we illustrate how to generate explanations using raw features in this section. Following [11,38], we adopt the same method to extract the features and obtain the raw user and item

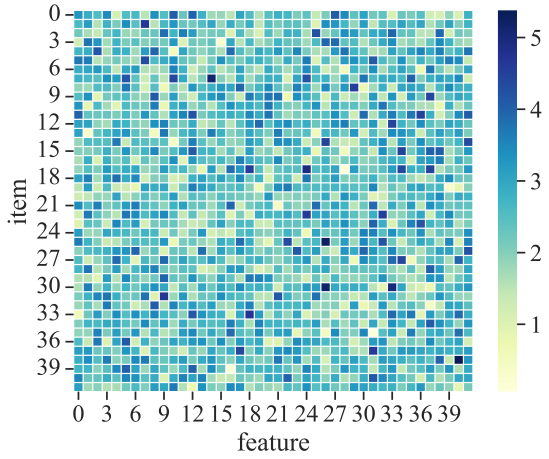


**Fig. 4.** The feature explanations of CMB-Individual-$\alpha$-nDCG-Recall. Only the results of partial items are shown.

**Table 5.** Top-5 feature-based explanations on *Phones* dataset.

| **Method** | **Feature-based Explanations** |
|---|---|
| CMB-Shared-SC-Recall | sound, volume, connector, headphone, protection |
| CMB-Shared-SC-NDCG | charger, button, flashlight, cable, protection |
| CMB-Shared-ILAD-Recall | connector, volume, pocket, charger, sound |
| CMB-Shared-ILAD-NDCG | port, headset, plug, volume, package |

feature matrices. Then, we apply two different feature-based explanations introduced in Sec. 3.4 on *Phones* dataset from Amazon. The explanation results are presented in Fig. 4 and Table 5. These findings support our idea that it is challenging to manually discover feature explanations for diversity in recommender systems. For example, as shown in Table 5, it is difficult to know how input features (such as sound, charger, and connector) would determine the diversity of phone recommendations. As a result, explainable diversity approaches like ours are necessary to discover such features in the recommendation.

## 5 Conclusion and Future Work

In this work, we propose CMB, a general bandit-based method, which optimizes the recommendation diversity while providing corresponding explanations. The method exhibits wide applicability and is agnostic to both recommendation models and diversity metrics. The proposed combination optimization target helps reach a more reasonable trade-off between recommendation accuracy and diversity performance. Besides, the explanations regarding diversification can

be provided with the meaningfulness of the factors obtained from counterfactual optimization. Extensive experiments on real-world datasets demonstrate our method's applicability, effectiveness, and explainability. In the future, we plan to design more efficient methods for generating explanations.

**Acknowledgments.** This work was supported by the Early Career Scheme (No. CityU 21219323) and the General Research Fund (No. CityU 11220324) of the University Grants Committee (UGC), and the NSFC Young Scientists Fund (No. 9240127).

## References

1. Bistritz, I., Bambos, N.: Cooperative multi-player bandit optimization. In: NeurIPS. pp. 2016–2027 (2020)
2. Bubeck, S., Munos, R., Stoltz, G.: Pure exploration in multi-armed bandits problems. In: International conference on Algorithmic learning theory. pp. 23–37 (2009)
3. Carbonell, J., Goldstein, J.: The use of mmr, diversity-based reranking for reordering documents and producing summaries. In: SIGIR. pp. 335–336 (1998)
4. Chen, L., Zhang, G., Zhou, E.: Fast greedy MAP inference for determinantal point process to improve recommendation diversity. In: NeurIPS. pp. 5627–5638 (2018)
5. Chen, W., Ren, P., Cai, F., Sun, F., de Rijke, M.: Improving end-to-end sequential recommendations with intent-aware diversification. In: CIKM. pp. 175–184 (2020)
6. Chen, Z., Silvestri, F., Wang, J., Zhu, H., Ahn, H., Tolomei, G.: Relax: Reinforcement learning agent explainer for arbitrary predictive models. In: CIKM. pp. 252–261 (2022)
7. Cheng, P., Wang, S., Ma, J., Sun, J., Xiong, H.: Learning to recommend accurate and diverse items. In: WWW. pp. 183–192 (2017)
8. Clarke, C.L., Kolla, M., Cormack, G.V., Vechtomova, O., Ashkan, A., Büttcher, S., MacKinnon, I.: Novelty and diversity in information retrieval evaluation. In: SIGIR. pp. 659–666 (2008)
9. Ding, Q., Liu, Y., Miao, C., Cheng, F., Tang, H.: A hybrid bandit framework for diversified recommendation. In: AAAI. pp. 4036–4044 (2021)
10. Ge, M., Delgado-Battenfeld, C., Jannach, D.: Beyond accuracy: evaluating recommender systems by coverage and serendipity. In: RecSys. pp. 257–260 (2010)
11. Ge, Y., Tan, J., Zhu, Y., Xia, Y., Luo, J., Liu, S., Fu, Z., Geng, S., Li, Z., Zhang, Y.: Explainable fairness in recommendation. In: SIGIR. pp. 681–691 (2022)
12. Harper, F.M., Konstan, J.A.: The movielens datasets: History and context. TIIS **5**(4), 1–19 (2015)
13. He, X., Deng, K., Wang, X., Li, Y., Zhang, Y., Wang, M.: Lightgcn: Simplifying and powering graph convolution network for recommendation. In: SIGIR. pp. 639–648 (2020)
14. Huang, Y., Wang, W., Zhang, L., Xu, R.: Sliding spectrum decomposition for diversified recommendation. In: KDD. pp. 3041–3049 (2021)
15. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
16. Li, L., Chu, W., Langford, J., Schapire, R.E.: A contextual-bandit approach to personalized news article recommendation. In: WWW. pp. 661–670 (2010)
17. Li, S., Zhou, Y., Zhang, D., Zhang, Y., Lan, X.: Learning to diversify recommendations based on matrix factorization. In: DASC/PiCom/DataCom/CyberSciTech. pp. 68–74 (2017)

18. Merton, R.K.: The matthew effect in science: The reward and communication systems of science are considered. Science **159**(3810), 56–63 (1968)
19. Ni, J., Li, J., McAuley, J.J.: Justifying recommendations using distantly-labeled reviews and fine-grained aspects. In: EMNLP-IJCNLP. pp. 188–197 (2019)
20. Pariser, E.: The filter bubble: How the new personalized web is changing what we read and how we think (2011)
21. Rendle, S., Freudenthaler, C., Gantner, Z., Schmidt-Thieme, L.: BPR: bayesian personalized ranking from implicit feedback. In: UAI. pp. 452–461 (2009)
22. Santos, R.L., Macdonald, C., Ounis, I.: Exploiting query reformulations for web search result diversification. In: WWW. pp. 881–890 (2010)
23. Shi, X., Liu, Q., Xie, H., Wu, D., Peng, B., Shang, M., Lian, D.: Relieving popularity bias in interactive recommendation: A diversity-novelty-aware reinforcement learning approach. TOIS **42**(2), 1–30 (2023)
24. Singh, J., Anand, A.: Exs: Explainable search using local model agnostic interpretability. In: WSDM. pp. 770–773 (2019)
25. Slivkins, A., et al.: Introduction to multi-armed bandits. Foundations and Trends® in Machine Learning **12**(1-2), 1–286 (2019)
26. Tsukuda, K., Goto, M.: Dualdiv: diversifying items and explanation styles in explainable hybrid recommendation. In: RecSys. pp. 398–402 (2019)
27. Vargas, S., Castells, P., Vallet, D.: Intent-oriented diversity in recommender systems. In: SIGIR. pp. 1211–1212 (2011)
28. Wang, X., Chen, Y., Yang, J., Wu, L., Wu, Z., Xie, X.: A reinforcement learning framework for explainable recommendation. In: ICDM. pp. 587–596 (2018)
29. Wasilewski, J., Hurley, N.: Incorporating diversity in a learning to rank recommender system. In: FLAIRS. pp. 572–578 (2016)
30. Wilhelm, M., Ramanathan, A., Bonomo, A., Jain, S., Chi, E.H., Gillenwater, J.: Practical diversified recommendations on youtube with determinantal point processes. In: CIKM. pp. 2165–2173 (2018)
31. Wu, H., Zhang, Y., Ma, C., Lyu, F., He, B., Mitra, B., Liu, X.: Result diversification in search and recommendation: A survey. TKDE (2024)
32. Wu, L., Quan, C., Li, C., Wang, Q., Zheng, B., Luo, X.: A context-aware user-item representation learning for item recommendation. TOIS **37**(2), 1–29 (2019)
33. Yu, H.: Optimize what you evaluate with: Search result diversification based on metric optimization. In: AAAI. pp. 10399–10407 (2022)
34. Zhang, M., Hurley, N.: Avoiding monotony: improving the diversity of recommendation lists. In: RecSys. pp. 123–130 (2008)
35. Zhang, Y., Hu, C., Dai, G., Kong, W., Liu, Y.: Self-adaptive graph neural networks for personalized sequential recommendation. In: ICONIP. pp. 608–619 (2021)
36. Zhang, Y., Zhang, X., Cui, Z., Ma, C.: Shapley value-driven data pruning for recommender systems. In: KDD (2025)
37. Zhang, Y., Chen, X., et al.: Explainable recommendation: A survey and new perspectives. Foundations and Trends® in Information Retrieval **14**(1), 1–101 (2020)
38. Zhang, Y., Lai, G., Zhang, M., Zhang, Y., Liu, Y., Ma, S.: Explicit factor models for explainable recommendation based on phrase-level sentiment analysis. In: SIGIR. pp. 83–92 (2014)
39. Zheng, G., Zhang, F., Zheng, Z., Xiang, Y., Yuan, N.J., Xie, X., Li, Z.: Drn: A deep reinforcement learning framework for news recommendation. In: WWW. pp. 167–176 (2018)
40. Zheng, Y., Gao, C., Chen, L., Jin, D., Li, Y.: Dgcn: Diversified recommendation with graph convolutional networks. In: WWW. pp. 401–412 (2021)