Diversified Interactive Recommendation with Implicit Feedback

Yong Liu,[1,2] **Yingtai Xiao,**[2,4*] **Qiong Wu,**[2] **Chunyan Miao,**[3†] **Juyong Zhang,**[4]
Binqiang Zhao,[5] **Haihong Tang**[5]

[1]Alibaba-NTU Singapore Joint Research Institute
[2]Joint NTU-UBC Research Centre of Excellence in Active Living for the Elderly (LILY)
[3]School of Computer Science and Engineering, Nanyang Technological University
[4]University of Science and Technology of China, [5]Alibaba Group
{stephenliu, wu.qiong, ascymiao}@ntu.edu.sg, {gzxyt@mail., juyong@}ustc.edu.cn
binqiang.zhao@alibaba-inc.com, piaoxue@taobao.com

Abstract

Interactive recommender systems that enable the interactions between users and the recommender system have attracted increasing research attention. Previous methods mainly focus on optimizing recommendation accuracy. However, they usually ignore the diversity of the recommendation results, thus usually results in unsatisfying user experiences. In this paper, we propose a novel diversified recommendation model, named $\underline{D}$iversified $\underline{C}$ontextual $\underline{C}$ombinatorial $\underline{B}$andit (DC^2B), for interactive recommendation with users' implicit feedback. Specifically, DC^2B employs determinantal point process in the recommendation procedure to promote diversity of the recommendation results. To learn the model parameters, a Thompson sampling-type algorithm based on variational Bayesian inference is proposed. In addition, theoretical regret analysis is also provided to guarantee the performance of DC^2B. Extensive experiments on real datasets are performed to demonstrate the effectiveness of the proposed method in balancing the recommendation accuracy and diversity.

Introduction

Conventional recommender systems are usually developed in non-interactive manner and learn the user preferences from logged user behavior data (Liu et al. 2017; Yang et al. 2018; Liu et al. 2018; Wang et al. 2018). One main drawback of these systems is that they cannot capture the changes of users' preferences in time. This requires the development of interactive recommender system that enables interactions (Steck, van Zwol, and Johnson 2015). In the literature, contextual bandit learning has been demonstrated to be a promising solution to interactive recommendation problems (Li et al. 2010; Zhao, Zhang, and Wang 2013; Tang et al. 2015; Wang, Wu, and Wang 2017; Qi et al. 2018). In these methods, the recommender system sequentially recommends a set of items to a user and adopts the user's immediate feedback to improve its recommendation policy.

In practice, users' implicit feedback (e.g., clicking history) are usually utilized to build recommender systems, be-

cause implicit feedback is user centric, and can be easily collected (Shi, Larson, and Hanjalic 2014; Wang et al. 2019). However, the implicit feedback usually brings bias signals which make the recommendation problems much more challenging. This bias comes from the fact that the implicit feedback can only capture the positive user preferences (i.e., observed user-item interactions), and all negative user preferences are missing. Although the non-interaction between the user and an item is usually treated as negative user preference in previous research work (Shi, Larson, and Hanjalic 2014), it does not explicitly indicate that the user dislikes the item, as non-interaction may also be caused by that the item has not been exposed to the user (Liang et al. 2016).

Moreover, previous interactive recommendation methods mainly focus on optimizing recommendation accuracy. They usually ignore other important properties of the recommendation results, for example the diversity of the recommended item set (Kunaver and Požrl 2017; Wu et al. 2019b). Therefore, the items in the recommendation lists generated by these approaches may usually be very similar with each other, and the recommendation results may only cover a small fraction of items. This usually leads to inferior user experiences, and thus reduces the commercial values of recommender systems. Intuitively, it is very challenging to achieve both high accuracy and diversity. The methods focusing too strongly on diversity usually put accuracy at risk. Because there is a lack of data for less popular items, considering such items for recommendation may lead to a decrease in recommendation accuracy (Adomavicius and Kwon 2011). Therefore, the main objective of diversified recommendation methods is to optimize the trade-off between accuracy and diversity, which is usually referred as the "accuracy–diversity dilemma" (Zhou et al. 2010).

In this paper, we propose a novel bandit learning framework for interactive recommender systems based on users' implicit feedback, which strives to achieve a good balance between accuracy and diversity in the recommendation results. To solve the bias problems caused by implicit feedback, we model the interactions between users and the recommender system from two perspectives: i) Diversified Item Exposure: the recommender system selects a set of relevant yet diverse items to expose to the user; ii) User Engage-

---

ments: the user eventually engages with some of the exposed items (e.g., clicks on the items). Specifically, the determinantal point process (DPP) (Kulesza, Taskar, and others 2012) is employed to select a set of diverse items to expose to users, considering both the qualities of items and the diversity of the selected item set. The advantage of DPP is that it explicitly models the probability that an item set would be selected to be shown to the user, thus can help solve the bias problem caused by implicit feedback (Liang et al. 2016). In addition, the contextual features of items are also utilized to model the observed user engagements on the recommended items.

To summarize, the major contributions made in this paper are as follows: (1) we propose a novel bandit learning method, i.e., Diversified Contextual Combinatorial Bandit (DC^{2}B), to improve the recommendation diversity for interactive recommender systems; (2) we propose a variational Bayesian inference algorithm under the Thompson sampling framework to learn the model parameters; (3) we also provide theoretical regret analysis for the proposed DC^{2}B method; (4) we perform extensive experiments on real datasets to demonstrate the effectiveness of DC^{2}B in balancing the recommendation accuracy and diversity.

Related Work

Diversified Recommendation

One major group of diversified recommendation methods are based on greedy heuristics. The pioneering work is maximal marginal relevance (MMR) (Carbonell and Goldstein 1998), which defines a marginal relevance to combine the relevance and diversity metrics, and creates a diversified ranking of items by choosing an item in each interaction such that it maximizes the marginal relevance. Other greedy heuristics methods vary in the definition of the marginal relevance, often in the form of a sub-modular objective function (Qin and Zhu 2013; Sha, Wu, and Niu 2016), which can be solved greedily with an approximation to the optimal solution. Another group of methods are based on refinement heuristics, which usually re-rank a pre-ranked item list through post-processing actions (Zhang and Hurley 2008; Antikacioglu and Ravi 2017). From another perspective, (Cheng et al. 2017) formulates the diversified recommendation problem as a supervised learning task, and proposes a diversified collaborative filtering model to solve the optimization problems. Recently, DPP has been demonstrated to be effective in modeling diversity in various machine learning problems (Kulesza, Taskar, and others 2012), and some recent work (Chen, Zhang, and Zhou 2018; Wilhelm et al. 2018; Wu et al. 2019a) employs DPP to improve recommendation diversity.

Interactive Recommendation

Contextual bandit has been often used for building interactive recommender systems. These methods mainly focus on optimizing the recommendation accuracy. For instance, (Li et al. 2010) proposes a contextual bandit algorithm, named LinUCB, which sequentially recommended articles to users based on the contextual information of users and

articles. (Zhao, Zhang, and Wang 2013) combines probabilistic matrix factorization with Thompson sampling and upper confidence bound based bandit algorithms to interactively select items. (Tang et al. 2015) proposes a parameter-free bandit approach that uses online bootstrap to learn the online recommendation model. Recently, (Wang et al. 2017) extends the LinUCB to incorporate users' social relationships into interactive recommender system. (Wang, Wu, and Wang 2017) proposes a factorization-based bandit approach to solve the online interactive recommendation problem. Moreover, in (Qi et al. 2018), the Thompson sampling framework is employed to solve the bandit problems with implicit feedback, where the implicit feedback is modeled as a composition of user result examination and relevance judgement. There also exist some interactive recommender systems focusing on promoting the recommendation diversity. For example, (Qin, Chen, and Zhu 2014) proposes a contextual combinatorial bandit framework, incorporating the entropy regularizer (Qin and Zhu 2013) to diversify the recommendation results. Differing from (Qin, Chen, and Zhu 2014), DC^{2}B is a full Bayesian framework which is more effective in balancing the recommendation accuracy and diversity.

Problem Formulation

We employ contextual bandit to build the diversified interactive recommender system. The recommender system is treated as an agent, and each item is treated as an arm. Let $\mathcal{A} = \{a_i\}_{i=1}^N$ denote the set of N arms (i.e., items). We assume each arm a_i has a contextual feature vector $\boldsymbol{x}_i \in \mathbb{R}^{1 \times d}$ summarizing its side information, and denote the features of all arms by $\boldsymbol{X} \in \mathbb{R}^{N \times d}$. At each trial, the recommender agent would firstly choose a subset of arms $\mathcal{S}$ from $\mathcal{A}$, considering the qualities of the arms and the diversity of selected arms. $\mathcal{S}$ is usually called as a *super arm*. Here, we empirically define the quality of an arm a_i as follows:

$$r_i = \exp(\boldsymbol{\theta}\boldsymbol{x}_i^\top), \tag{1}$$

where $\boldsymbol{\theta}$ is the bandit parameter that describes the user preferences. The diversity of the selected super arm $\mathcal{S}$ can be measured by the intra-list distance metric (Zhang and Hurley 2008). Once a diversified super arm $\mathcal{S}$ has been selected according to a policy π and displayed to the user, the user's engagements on displayed items (e.g., clicks on items) are used as the rewards for recommender agent to optimize its recommendation policy. Through interactions with the user, the recommender agent aims to adjust its super arm selection strategy to maximize its cumulative reward over time.

Diversified Item Exposure

The DPP is an elegant probabilistic model with the ability to model diversity in various machine learning problems (Kulesza, Taskar, and others 2012). In this work, we utilize DPP to model the selection probability of a relevant yet diverse super arm $\mathcal{S}$. Formally, a DPP $\mathcal{P}$ on the set of candidate arms $\mathcal{A}$ is a probability measure on $2^{\mathcal{A}}$, describing the probability for the set of all subsets of $\mathcal{A}$. If $\mathcal{P}$ assigns nonzero probability on the empty set $\emptyset$, there exists a real,

positive semi-definite kernel matrix $\boldsymbol{L} \in \mathbb{R}^{N \times N}$, such that the probability of the super arm $\mathcal{S}$ can be defined as follows:

$$p(\mathcal{S}) = \frac{\det(\boldsymbol{L}_{[\mathcal{S}]})}{\det(\boldsymbol{L} + \boldsymbol{I})}, \tag{2}$$

where $\boldsymbol{I}$ is the identity matrix, $\boldsymbol{L}_{[\mathcal{S}]} \equiv [\boldsymbol{L}_{ij}]_{a_i, a_j \in \mathcal{S}}$ is the sub-matrix of $\boldsymbol{L}$. As revealed in (Kulesza, Taskar, and others 2012), $\boldsymbol{L}$ can be written as a Gram matrix, $\boldsymbol{L} = \boldsymbol{V}\boldsymbol{V}^{\top}$, where the rows of $\boldsymbol{V}$ are vectors representing the arms. Following (Chen, Zhang, and Zhou 2018; Wilhelm et al. 2018), we empirically set $\boldsymbol{V}_i = (r_i)^{\alpha} \boldsymbol{x}_i$, where $\alpha > 0$ is a parameter controlling the impacts of item qualities. Then, the elements of $\boldsymbol{L}$ are defined as $\boldsymbol{L}_{ij} = (r_i r_j)^{\alpha} \boldsymbol{x}_i \boldsymbol{x}_j^{\top}$. If $\boldsymbol{x}_i$ is normalized, i.e., $\|\boldsymbol{x}_i\|_2 = 1$, the Cosine similarity between a_i and a_j can be calculated as $\boldsymbol{C}_{ij} = \boldsymbol{x}_i \boldsymbol{x}_j^{\top}$. We can re-write $\boldsymbol{L}$ as follows:

$$\boldsymbol{L} = \mathrm{Diag}\{\exp(\alpha \widetilde{\boldsymbol{r}})\} \cdot \boldsymbol{C} \cdot \mathrm{Diag}\{\exp(\alpha \widetilde{\boldsymbol{r}})\}, \tag{3}$$

where $\mathrm{Diag}\{\widetilde{\boldsymbol{r}}\}$ is a diagonal matrix with the ith diagonal element being $\widetilde{r}_i = \boldsymbol{\theta}\boldsymbol{x}_i^{\top}$, and $\boldsymbol{C}$ is the similarity matrix. Then, the log-probability of the super arm $\mathcal{S}$ is:

$$\log p(\mathcal{S}) \propto 2\alpha \sum_{a_i \in \mathcal{S}} \widetilde{r}_i + \log \det(\boldsymbol{C}_{[\mathcal{S}]}), \tag{4}$$

where the last term is maximized when the features of arms in $\mathcal{S}$ are orthogonal, thus it helps promote recommendation diversity (Chen, Zhang, and Zhou 2018). In addition, Eq. (4) also indicates the parameter α can help balance the relevance and diversity of items for recommendation.

User Engagements

The user's engagements on displayed items are expressed by her implicit feedback (e.g., clicks on the items), which is usually described by a set of binary variables. If the user engages in the arm a_i, we set y_i to 1; otherwise, we set y_i to 0. Once an arm $a_i \in \mathcal{S}$ has been displayed to the user, we assume the user's engagements on a_i is only determined by its quality. Thus, the probability of the observed user engagement on a_i, i.e., $y_i = 1$, can be defined as follows:

$$p_i \triangleq \rho(\boldsymbol{\theta}\boldsymbol{x}_i^{\top}) = \frac{\exp(\boldsymbol{\theta}\boldsymbol{x}_i^{\top})}{1 + \exp(\boldsymbol{\theta}\boldsymbol{x}_i^{\top})} = \frac{r_i}{1 + r_i}. \tag{5}$$

This can be explained as that when an arm a_i is offered to the user, the user engages in this arm or a *virtual arm* a_0 with a relevance score 1. Based on these assumptions, we can define the joint probability of observed user engagements $\mathcal{Y} = \{y_i | a_i \in \mathcal{S}\}$ as follows:

$$\begin{aligned} p(\mathcal{Y}, \mathcal{S}, \boldsymbol{\theta}) &= p(\boldsymbol{\theta})p(\mathcal{S}|\boldsymbol{\theta})p(\mathcal{Y}|\mathcal{S}, \boldsymbol{\theta}) \\ &= p(\boldsymbol{\theta})\frac{\det(\boldsymbol{L}_{[\mathcal{S}]})}{\det(\boldsymbol{L} + \boldsymbol{I})} \prod_{a_i \in \mathcal{S}} p_i^{y_i}(1 - p_i)^{1-y_i}, \end{aligned} \tag{6}$$

where $p(\boldsymbol{\theta})$ is the prior assigned to bandit parameters. In addition, we assume $p(\boldsymbol{\theta})$ follows a Gaussian distribution $\mathcal{N}(\boldsymbol{m}, \boldsymbol{\Sigma})$, and $\boldsymbol{m}$, $\boldsymbol{\Sigma}$ are bounded. This assumption is typically used in practice.

Algorithm 1 Thompson Sampling for DC^{2}B

Initialize $\boldsymbol{m} = \boldsymbol{0}$, $\boldsymbol{\Sigma} = \lambda \boldsymbol{I}$, and $\mathcal{R} = \emptyset$.
for $t = 0$ **to** T **do**
  $\mathcal{A}_t \leftarrow \mathcal{A} \setminus \mathcal{R}$, $\boldsymbol{X}_t = \{\boldsymbol{x}_i | a_i \in \mathcal{A}_t\}$
  Randomly sample $\widehat{\boldsymbol{\theta}} \sim \mathcal{N}(\boldsymbol{m}, \boldsymbol{\Sigma})$
  $\mathcal{S} \leftarrow \mathcal{O}(\widehat{\boldsymbol{\theta}}, \boldsymbol{X}_t)$
  Play super arm $\mathcal{S}$ and observe the reward $\mathcal{Y}$
  Update $\boldsymbol{\Sigma}$ and $\boldsymbol{m}$ according to Eq. (9), (10), and (11).
  $\mathcal{R} \leftarrow \mathcal{R} \cup \mathcal{S}$
end for

Parameter Inference

Once a newly observation $(\mathcal{S}, \mathcal{Y})$ is available, we employ variational Bayesian inference (Blei, Kucukelbir, and McAuliffe 2017) to develop a closed form approximation to the posterior of $\boldsymbol{\theta}$. According to (Blei, Kucukelbir, and McAuliffe 2017), the approximated posterior $q(\boldsymbol{\theta})$ of $\boldsymbol{\theta}$ can be expressed as $\log q^*(\boldsymbol{\theta}) = \mathbb{E}_{param \neq \boldsymbol{\theta}}[\log p(\mathcal{Y}, \mathcal{S}, \boldsymbol{\theta})] + \mathrm{const}$. Moreover, based on the knowledge in Linear Algebra, we have $\det(\boldsymbol{L}_{[\mathcal{S}]}) = \prod_{a_i \in \mathcal{S}} r_i^{2\alpha} \det(\boldsymbol{X}_{[\mathcal{S}]} \boldsymbol{X}_{[\mathcal{S}]}^{\top})$ and $\det(\boldsymbol{L} + \boldsymbol{I}) = \exp(\mathrm{tr}(\log(\boldsymbol{L} + \boldsymbol{I})))$. Then, we can have the following log-likelihood function:

$$\begin{aligned} &\log p(\mathcal{Y}, \mathcal{S}|\boldsymbol{\theta}) \\ &= \sum_{a_i \in \mathcal{S}} (\varphi_i + 2\alpha \log r_i) + \log \det(\boldsymbol{X}_{[\mathcal{S}]} \boldsymbol{X}_{[\mathcal{S}]}^{\top}) \\ &\quad - \sum_{j=1}^{N} \log(1 + r_j^{2\alpha} \boldsymbol{x}_j \boldsymbol{x}_j^{\top}), \end{aligned} \tag{7}$$

where $\varphi_i = y_i \log p_i + (1 - y_i) \log(1 - p_i)$. In Eq. (7), the likelihood function is a logistic function, which is not conjugate with the Gaussian priors on $\boldsymbol{\theta}$. To address this issue, the following Gaussian lower bound on the logistic function is employed to approximate the likelihood (Jaakkola and Jordan 1997), $\rho(x) \geq \rho(\xi)e^{\frac{x-\xi}{2} - \lambda(\xi)(x^2 - \xi^2)}$, where $\lambda(\xi) = \frac{1}{2\xi}(\rho(\xi) - \frac{1}{2})$, and ξ is an auxiliary variable needs to be adjusted to make the bound tight at $x = \pm\xi$. Moreover, by assuming $\|\boldsymbol{\theta}\|_2 \leq A$ and $\|\boldsymbol{x}_j\|_2 \leq B$, we have $-\log[1 + \exp(2\alpha\boldsymbol{\theta}\boldsymbol{x}_j^{\top})\boldsymbol{x}_j\boldsymbol{x}_j^{\top}] \geq -\exp(2\alpha\boldsymbol{\theta}\boldsymbol{x}_j^{\top})\boldsymbol{x}_j\boldsymbol{x}_j^{\top} \geq -\exp(2\alpha AB)B^2$. As we assume $\boldsymbol{m}$ and $\boldsymbol{\Sigma}$ are bounded, it is reasonable to infer that $\boldsymbol{\theta}$ is bounded. By normalizing $\boldsymbol{x}_j$, we can make $\boldsymbol{x}_j$ bounded. Then, we have the following lower bound of the log-likelihood function in Eq. (7):

$$\log p(\mathcal{Y}, \mathcal{S}|\boldsymbol{\theta}) \geq \mathrm{const.} +$$
$$\underbrace{\sum_{a_i \in \mathcal{S}} \left[\frac{(2y_i - 1 + 4\alpha)\boldsymbol{\theta}\boldsymbol{x}_i^{\top}}{2} - \lambda(\xi_i)(\boldsymbol{\theta}(\boldsymbol{x}_i^{\top}\boldsymbol{x}_i)\boldsymbol{\theta}^{\top}) + \phi(\xi_i) \right]}_{\log h(\boldsymbol{\theta}, \boldsymbol{\xi})},$$
$$\tag{8}$$

where $\phi(\xi_i) = \log \rho(\xi_i) - \frac{\xi_i}{2} + \lambda(\xi_i)\xi_i^2$. The optimal variational distribution of $\boldsymbol{\theta}$ is as follows: $\log q^*(\boldsymbol{\theta}) \approx \mathbb{E}[\log h(\boldsymbol{\theta}, \boldsymbol{\xi})] + \mathbb{E}[\log p(\boldsymbol{\theta})] + \mathrm{const}$. Due to model conjugacy, we can know that $q(\boldsymbol{\theta})$ shall follow a Gaussian distri-

Algorithm 2 DPP Greedy Search $\mathcal{S} \leftarrow \mathcal{O}(\widehat{\boldsymbol{\theta}}, \boldsymbol{X}_t)$

Startup: Construct $\boldsymbol{L}, \boldsymbol{p}$ according to $\widehat{\boldsymbol{\theta}}, \boldsymbol{X}_t$
Initialize $\mathbf{c}_i = [\,]$, $d_i^2 = \boldsymbol{L}_{ii}$, $j = \arg\max_{i \in Z} \log\left(d_i^2\right) + \log(p_i)$, $\mathcal{S} = \{j\}$.
for $k = 0$ **to** K **do**
  for $i \in Z \backslash \mathcal{S}$ **do**
    $e_i = (\boldsymbol{L}_{ji} - \langle \mathbf{c}_j, \mathbf{c}_i \rangle)/d_j$
    $\mathbf{c}_i = \begin{bmatrix} \mathbf{c}_i & e_i \end{bmatrix}, d_i^2 = d_i^2 - e_i^2$
  end for
  $j = \arg\max_{i \in Z \backslash Y_{\mathrm{g}}} \log\left(d_i^2\right) + \log(p_i), \mathcal{S} = \mathcal{S} \cup \{j\}$
end for
Return $\mathcal{S}$

---

bution $\mathcal{N}(\boldsymbol{m}_{post}, \boldsymbol{\Sigma}_{post})$, where the mean and variance are as follows:

$$\boldsymbol{\Sigma}_{post}^{-1} = \boldsymbol{\Sigma}^{-1} + 2 \sum_{a_i \in \mathcal{S}} \lambda(\xi_i) \boldsymbol{x}_i^\top \boldsymbol{x}_i, \tag{9}$$

$$\boldsymbol{m}_{post} = \boldsymbol{\Sigma}_{post}\Big[\boldsymbol{\Sigma}^{-1}\boldsymbol{m} + \sum_{a_i \in \mathcal{S}}(y_i + 2\alpha - \frac{1}{2})\boldsymbol{x}_i\Big]. \tag{10}$$

Since no prior has been assigned to ξ_i, the optimal value of ξ_i can be derived by maximizing the expected log-likelihood function: $\ell(\xi_i) = \mathbb{E}[\log p(\mathcal{Y}, \mathcal{S}|\boldsymbol{\theta}, \xi_i)]$. Taking the derivative of $\ell(\xi_i)$ with respect to ξ_i and setting it to zero, the optimal value of ξ_i can be obtained as follows:

$$\xi_i = \sqrt{\boldsymbol{x}_i(\boldsymbol{\Sigma}_{post} + \boldsymbol{m}_{post}^\top \boldsymbol{m}_{post})\boldsymbol{x}_i^\top}. \tag{11}$$

We employ Thompson sampling (TS) to update the model parameters by balancing exploration and exploitation. The details of the TS algorithm are summarized in Algorithm 1. In standard TS method, it is required to sample from the true posterior of model parameter $\boldsymbol{\theta}$. As the logistic likelihood function is not conjugate with the Gaussian prior, we propose to sample from the approximated poster distribution $q(\boldsymbol{\theta})$. Once completing the sampling of $\boldsymbol{\theta}$, the DPP kernel matrix $\boldsymbol{L}$ is fixed, and we can select the optimal super arm $\mathcal{S}$ by maximizing $f_{\boldsymbol{\theta}}(\mathcal{S}) = \prod_{a_i \in \mathcal{S}} p_i \det(\boldsymbol{L}_{[\mathcal{S}]})$. We employ the fast greedy MAP inference algorithm (Chen, Zhang, and Zhou 2018) to obtain the optimal super arm. The details of the greedy algorithm are summarized in Algorithm 2.

Regret Analysis

We consider a model involving a set of actions $\boldsymbol{S}$ and a set of functions $\mathcal{F} = \{f_{\boldsymbol{\theta}} : \boldsymbol{S} \mapsto \mathbb{R}|\boldsymbol{\theta} \in \Theta\}$ indexed by a random variable $\boldsymbol{\theta}$ which belongs to an index set Θ. At each time t, a random subset $\boldsymbol{S}_t \subseteq \boldsymbol{S}$ is presented and an action $\mathcal{S}_t \in \boldsymbol{S}_t$ is selected after which the reward R_t is gained. We define the reward function as: $\mathbb{E}[R_t] \triangleq f_{\boldsymbol{\theta}}(\mathcal{S}_t) = \prod_{a_i \in \mathcal{S}_t} p_i \det(\boldsymbol{L}_{[\mathcal{S}_t]}) = \prod_{a_i \in \mathcal{S}_t} p_i r_i^{2\alpha} \det(\boldsymbol{X}_{[\mathcal{S}_t]}\boldsymbol{X}_{[\mathcal{S}_t]}^T)$, and define the reward at trial t as: $R_t = f_{\boldsymbol{\theta}}(\mathcal{S}_t) + \epsilon_t$. Therefore, we have $\mathbb{E}[\epsilon_t] = 0$. In addition we assume $\forall f_{\boldsymbol{\theta}} \in \mathcal{F}, \forall \mathcal{S}_t \in \boldsymbol{S}, f_{\boldsymbol{\theta}}(\mathcal{S}_t) \in [0, C]$. For a recommendation policy π, we can define the Bayesian risk bound as follows:

$$Regret(T, \pi) = \sum_{t=1}^{T} \mathbb{E}\left[\max_{s \in \boldsymbol{S}_t} f_{\boldsymbol{\theta}}(s) - f_{\boldsymbol{\theta}}(\mathcal{S}_t)\right]. \tag{12}$$

To perform the regret analysis, we first introduce the following two Lemmas, which can be proofed following the Proposition 9 and 10 in (Russo and Van Roy 2014). The difference is that the variable in our approach is a set of arms $\mathcal{S}_t$ instead of a single arm a, the proofs are similar so we omit them due to space limitation. We set $\sigma = 1$ according to Lemma 7 in (Qi et al. 2018), which also shows that $f_{\boldsymbol{\theta}}$ satisfies Assumption 2 in (Russo and Van Roy 2014).

Lemma 1. *For all $T \in \mathbb{N}$, $\alpha_0 > 0$ and $\delta \leq 1/2T$,*

$$Regret\left(T, \pi^{\mathrm{TS}}\right) \leq 4\sqrt{\dim_M\left(\mathcal{F}, T^{-1}\right)\beta_T^*(\mathcal{F}, \alpha_0, \delta)T} + 1 + \left[\dim_M\left(\mathcal{F}, T^{-1}\right) + 1\right]C,$$

where $\dim_M\left(\mathcal{F}, T^{-1}\right)$ is the ϵ-dimension, $\beta_T^*(\mathcal{F}, \alpha_0, \delta) := 8\ln\left(N\left(\mathcal{F}, \alpha_0, \|\cdot\|_\infty\right)/\delta\right) + 2\alpha t\left(\frac{15}{2}C + \ln(2t^2/\delta)\right)$, and $N\left(\mathcal{F}, \alpha_0, \|\cdot\|_\infty\right)$ denotes the α_0-covering number of $\mathcal{F}$.

Lemma 2. *Suppose $\Theta \subset \mathbb{R}^d$, and $|f_{\boldsymbol{\theta}}(\mathcal{S}_t) - f_{\boldsymbol{\theta}^*}(\mathcal{S}_t)| \leq |h(\boldsymbol{\theta} - \boldsymbol{\theta}^*)^\top \phi(\mathcal{S}_t)|$, where $\phi(\mathcal{S}_t) = \sum_{i \in \mathcal{S}_t} \boldsymbol{x}_i$ and h is a constant. Assume there exist constants γ, S_0 such that $\forall \mathcal{S}_t \in \boldsymbol{S}$ and $\boldsymbol{\theta} \in \Theta$, $\|\boldsymbol{\theta}\|_2 \leq S_0$, and $\|\phi(\mathcal{S}_t)\|_2 \leq \gamma$. Then we have*

$$\dim_M(\mathcal{F}, \epsilon) \leq 3d\frac{e}{e-1}\ln\left\{3\left(\frac{2S_0 h\gamma}{\epsilon}\right)^2\right\} + 1. \tag{13}$$

According to Lemma 1, in our problem, $C = 1$, and we can choose $\alpha_0 = 1/T^2, \delta = 1/T^2$. Then, the Bayesian risk bound of DC^{2}B is given by the following Theorem.

Theorem 1. *When T is sufficient large, the Bayesian risk bound of DC^{2}B is*

$$Regret\left(T, \pi^{TS}\right) = O\left(d\ln\left(\alpha s e^{2\alpha s}T\sqrt{d}\right)\sqrt{T}\right), \tag{14}$$

where s is the number of items in a selected set, d is the dimension of $\boldsymbol{\theta}$.

Proof. We first assume $\|\boldsymbol{\theta}\|_2 \leq 1$, $\|\boldsymbol{x}\|_2 \leq 1$, and introduce the following inequalities: (1) Mean Value Theorem: we have $|\boldsymbol{\theta}\boldsymbol{x}| \leq \|\boldsymbol{\theta}\|_2\|\boldsymbol{x}\|_2 \leq 1$, then $|p - p^*| = |\rho(\boldsymbol{\theta}^\top \boldsymbol{x}) - \rho(\boldsymbol{\theta}^{*\top}\boldsymbol{x})| = |\rho'(\xi)(\boldsymbol{\theta} - \boldsymbol{\theta}^*)^\top \boldsymbol{x}| \leq \frac{1}{4}\|\boldsymbol{\theta} - \boldsymbol{\theta}^*\|_2\|\boldsymbol{x}\|_2$, and $|r^2 - r^{*2}| = |\exp(2\alpha\boldsymbol{\theta}^\top \boldsymbol{x}) - \exp(2\alpha\boldsymbol{\theta}^{*\top}\boldsymbol{x})| = |\exp(2\alpha\zeta)2(\boldsymbol{\theta} - \boldsymbol{\theta}^*)^\top \boldsymbol{x}| \leq 2\alpha e^{2\alpha}\|\boldsymbol{\theta} - \boldsymbol{\theta}^*\|_2\|\boldsymbol{x}\|_2$, where $\rho'(x) = \rho(x)(1 - \rho(x)) \leq \frac{1}{4}, 0 \leq \xi \leq 1, 0 \leq \zeta \leq 1$; (2) Gram Inequality: $|\det(\boldsymbol{X}_{[\mathcal{S}]}^\top \boldsymbol{X}_{[\mathcal{S}]})| = |\det(G(\boldsymbol{x}_1, \cdots, \boldsymbol{x}_s))| \leq \|\boldsymbol{x}_1\|_2^2 \cdots \|\boldsymbol{x}_s\|_2^2 \leq 1$, where $[G(\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n)]_{i,j} = \boldsymbol{x}_i^\top \boldsymbol{x}_j$ defines a gram matrix; (3) Triangle inequality: $|x_1 x_2 - y_1 y_2| = |x_1 x_2 - y_1 x_2 + y_1 x_2 - y_1 y_2| \leq |x_1 - y_1||x_2| + |y_1||x_2 - y_2|$. Based on these inequalities, we have,

$$|f_{\boldsymbol{\theta}}(\mathcal{S}_t) - f_{\boldsymbol{\theta}^*}(\mathcal{S}_t)|$$

$$= |\det(\boldsymbol{X}_{[\mathcal{S}_t]}^\top \boldsymbol{X}_{[\mathcal{S}_t]})|\Big|\prod_{i=1}^{s} p_i r_i^{2\alpha} - \prod_{i=1}^{s} p_i^* r_i^{*2\alpha}\Big|$$

$$\leq (\frac{8\alpha + 1}{4})e^{2\alpha s}|(\boldsymbol{\theta} - \boldsymbol{\theta}^*)^\top \sum_{i=1}^{s} \boldsymbol{x}_i|$$

$$\leq \frac{8\alpha + 1}{4}se^{2\alpha s}\sqrt{d}\|\boldsymbol{\theta} - \boldsymbol{\theta}^*\|_\infty, \tag{15}$$

where we use inequality $\|\boldsymbol{\theta}\|_2 \leq \sqrt{d}\|\boldsymbol{\theta}\|_\infty$. According to Eq. (15), an α_0-covering of $\mathcal{F}$ can therefore be attained through an (α_0/γ)-covering of $\Theta \subset \mathbb{R}^d$, where $\gamma = \frac{8\alpha+1}{4}se^{2\alpha s}\sqrt{d}$. Evenly divide $\mathbb{R}^d$ in each dimension, we can obtain $N\left(\mathbb{R}^d, \alpha_0, \|\cdot\|_\infty\right) = (1/\alpha_0)^d$. Then, we have

$$N\left(\mathcal{F}, \alpha_0, \|\cdot\|_\infty\right) = (\gamma/\alpha_0)^d = \left(\frac{(8\alpha+1)se^{2\alpha s}\sqrt{d}}{4\alpha_0}\right)^d. \tag{16}$$

In our problem, $S_0 = 1$, $h = \frac{8\alpha+1}{4}e^{2\alpha s}$ and $\gamma = s$. According to Lemma 2 and Eq. (15), we have the following bound on $\dim_M\left(\mathcal{F}, T^{-1}\right)$:

$$\dim_M(\mathcal{F}, T^{-1})$$
$$\leq 3d\frac{e}{e-1}\ln\left\{3\left(\frac{(8\alpha+1)e^{2\alpha s}Ts}{2}\right)^2\right\} + 1. \tag{17}$$

Let $\alpha_0 = 1/T^2$, $\delta = 1/T^2$, $C = 1$. When T is sufficient large, the second part of $\beta_T^*(\mathcal{F}, \alpha_0, \delta)$ will decrease to zero. After some calculation together with above two bounds, we can finish the proof. $\square$

The upper bound in Theorem 1 mainly depends on the dimensionality of model parameter d, the size of recommended item set s, and the quality controlling parameter α. Here, d describes the model complexity. As d increases, $\boldsymbol{\theta}$ is able to model more complex scenarios. However, a sophisticated model would cause over-fitting, resulting in poor performances. Therefore, the regret bound would be high, when d is large. The Proposition 9 in (Russo and Van Roy 2014) gives the Bayesian risk bound for non-combinatorial bandit methods as $O(rd\sqrt{T}\log(rT))$, where r is a parameter determined by the reward function. By simply repeating the recommendation s times to get a set of items, the bound would be $O(srd\sqrt{T}\log(rT))$. In DC^{2}B, if we set $\alpha = 1$, the Bayesian regret bound would be $O(d\sqrt{T}\log(se^{2s}\sqrt{d}T))$, which is slightly different from multiplying s to the bound of non-combinatorial methods. This is because our reward function also takes the recommendation diversity into account. As α controls the impacts of item qualities, the increase of α would increase the risks caused by the estimation of item qualities. Thus, the regret will grow as α increases.

Experiments

Experimental Settings

Datasets The experiments are performed on the following datasets: Movielens-100K, Movielens-1M[1], and Anime[2]. Movilens-100K contains 100,000 ratings given by 943 users to 1,682 movies, and Movielens-1M contains 1,000,209 ratings given by 6,040 users to 3,706 movies. There are 18 movie categories in both Movielens datasets. We denote these two datasets by ML-100K and ML-1M, respectively.

---

[1] https://grouplens.org/datasets/movielens/

[2] https://www.kaggle.com/CooperUnion/anime-recommendations-database

Table 1: The statistics of experimental datasets.

Datasets	# Users	# Items	# Inter.	# Cate.
ML-100K	942	1,447	55,375	18
ML-1M	6,038	3,533	575,281	18
Anime	69,400	8,825	5,231,117	44

For Anime dataset, there are 7,813,737 ratings given by 73,515 users to 11,200 animes, and there are 44 anime categories. Following (Liu et al. 2018), we keep the ratings larger than 3 as positive feedback on ML-100K and ML-1M datasets, and keep the ratings larger than 6 as positive feedback on the Anime dataset. Table 1 summarizes the statistics of the experimental datasets, where movies and animes are "items". In these datasets, each item may belong to multiple categories. The density of ML-100K, ML-1M, and Anime datasets are 4.06%, 2.70%, and 0.85%, respectively.

Setup and Metrics For interactive recommendation methods, it is most appropriate to use an online experimental setting with real time user interactions for evaluation. However, it is typically impossible to have such an environment in academic research. Hence, following (Zhao, Zhang, and Wang 2013; Qin, Chen, and Zhu 2014; Wang et al. 2017), we assume that users' ratings on items recorded in our experimental datasets are not biased by the recommender system, and these records can be regarded as unbiased user feedback in our experimental settings. The unbiased offline evaluation strategy (Li et al. 2011) is used to evaluate the recommendation methods. In the experiments, we randomly partition each dataset into two non-overlapping sets, by randomly sampling 80% of the users for training and using the remaining 20% users for testing. Moreover, we employ BPRMF (Rendle et al. 2009) to learn the embeddings of items based on training data, which are used as the contextual features of arms. Empirically, we set the dimensionality of the item embeddings to 10. As users are usually interested in a few top-ranked recommendation items, we adopt Precision@N to evaluate the recommendation accuracy (Shi, Larson, and Hanjalic 2014), by aggregating the recommended items in $\lfloor N/|\mathcal{S}|\rfloor$ trials and computing the precision. Specifically, N is set to 10, 30, and 50. We also evaluate the average recommendation diversity of each method over all recommendation trials, by the intra-list distance (ILD) (Zhang and Hurley 2008) metric as follows: $\frac{1}{T}\sum_{t=1}^{T}\left[\frac{2}{|\mathcal{S}_t|(|\mathcal{S}_t|-1)}\sum_{a_i\in\mathcal{S}_t}\sum_{a_j\in\mathcal{S}_t,i\neq j}(1-sim_{ij})\right]$, where $\mathcal{S}_t$ is recommended item set at trial t, $|\mathcal{S}_t|$ denotes the size of $\mathcal{S}_t$, T is the total number of recommendation trials, sim_{ij} denotes the similarity between a_i and a_j. As an item may belong to multiple item categories, we define the item similarity sim_{ij} by using the Jaccard similarity of the categories of two items. For these accuracy and diversity metrics, we first compute the value for each user, and then report the averaged value over all users. Following (Cheng et al. 2017), we also employ F-measure to evaluate the performances of different methods on trading-off between accuracy and diversity, where *F-measure=2*accuracy*diversity / (accuracy+diversity)*.

Table 2: Recommendation performances of different algorithms. The best results are in **bold faces**, and the second best results are underlined.

Datasets	Metrics	LogRank	MMR	ϵ-Greedy	DPP^{map}	C^{2}UCB	EC-Bandit	DC^{2}B
ML-100K	Precision@10	0.3548	**0.3665**	0.3421	**0.3665**	0.3633	0.2128	<u>0.3649</u>
	Precision@30	0.2872	0.2872	0.2792	0.2846	**0.3415**	0.1633	<u>0.3211</u>
	Precision@50	0.2507	0.2499	0.2433	0.2554	**0.3146**	0.1453	<u>0.2882</u>
	Diversity	0.8024	<u>0.8151</u>	0.8145	0.7985	0.7827	**0.8356**	0.8118
	F-measure	0.3820	0.3825	0.3747	0.3870	**0.4488**	0.2476	<u>0.4254</u>
ML-1M	Precision@10	**0.3785**	0.3754	0.3631	<u>0.3764</u>	0.3418	0.2160	**0.3785**
	Precision@30	<u>0.3204</u>	0.3173	0.3084	0.3173	0.3192	0.1750	**0.3401**
	Precision@50	0.2841	0.2824	0.2745	0.2807	<u>0.2998</u>	0.1611	**0.3117**
	Diversity	<u>0.8516</u>	**0.8531**	0.8462	0.8174	0.8319	0.8326	0.8367
	F-measure	0.4261	0.4221	0.4145	0.4179	<u>0.4408</u>	0.2700	**0.4542**
Anime	Precision@10	<u>0.3141</u>	**0.3157**	0.2867	**0.3157**	0.0095	0.1733	0.3003
	Precision@30	0.2527	0.2534	0.2366	<u>0.2541</u>	0.1116	0.1326	**0.2666**
	Precision@50	0.2165	<u>0.2178</u>	0.2025	0.2164	0.1518	0.1168	**0.2419**
	Diversity	0.8323	<u>0.8495</u>	**0.8521**	0.8414	0.5031	0.8460	0.8355
	F-measure	0.3436	<u>0.3467</u>	0.3272	0.3443	0.2332	0.2053	**0.3752**

Evaluated Recommendation Methods

As the training users are non-overlapping with the testing users, the recommendation algorithms (Shi, Larson, and Hanjalic 2014) designed for warm-start settings are not suitable as baselines. In this paper, we compare DC^{2}B with the following recommendation methods: (1) **LogRank**: In this method, we define the quality score of each arm a_i as $r_i = 1/(1 + \exp(-\bar{\boldsymbol{u}}\boldsymbol{x}_i^\top))$, where $\bar{\boldsymbol{u}}$ is the mean of the user embeddings learnt from the training data. Then, the $|\mathcal{S}_t|$ available arms with the highest quality scores are selected as a super arm $\mathcal{S}_t$ for recommendation at trial t; (2) **MMR**: This method employs MMR strategy (Carbonell and Goldstein 1998) to promote the recommendation diversity. At trial t, this method sequentially selects an available arm with the largest maximal marginal relevance score into $\mathcal{S}_t$. The maximal marginal relevance score is defined as $\tilde{r}_i = \alpha r_i - \frac{(1-\alpha)}{|\mathcal{S}_t|} \sum_{j \in \mathcal{S}_t} sim(\boldsymbol{x}_i, \boldsymbol{x}_j)$, where r_i is the arm quality defined in the LogRank method, and $sim(\boldsymbol{x}_i, \boldsymbol{x}_j)$ is the Cosine similarity between $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$; (3) ϵ**-Greedy**: This method randomly adds an available arm into $\mathcal{S}_t$ with probability ϵ, and adds the arm with highest quality into $\mathcal{S}_t$ with probability $1 - \epsilon$. The item quality is defined the same as in LogRank method; (4) **DPP^{map}** (Chen, Zhang, and Zhou 2018): This non-interactive method uses determinantal point process to promote recommendation diversity. The item quality is defined the same as in LogRank; (5) **C^{2}UCB** (Qin, Chen, and Zhu 2014): This methods integrates the LinUCB framework with an entropy regularizer to promote diversity for interactive recommendation. (6) **EC-Bandit** (Qi et al. 2018): This bandit method is based on Thompson sampling framework and developed for interactive recommendation with users' implicit feedback. In this method, the user needs to interact with the recommender $|\mathcal{S}_t|$ times to generate the recommended item set at trial t.

For all methods, we empirically set the size of $\mathcal{S}_t$ to 10 in each trial. A validation set is sampled from training data to choose hyper-parameters. The best parameter settings for
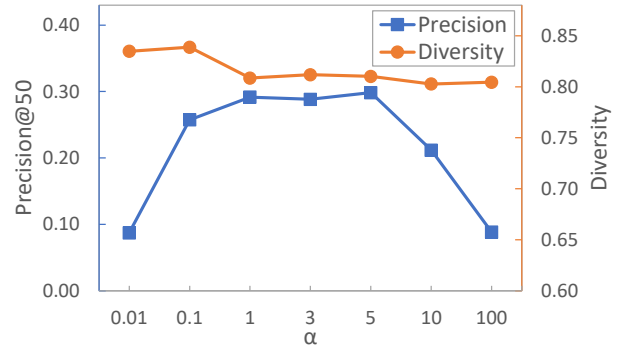

Figure 1: Performance trend of DC^{2}B with respect to different settings of α on ML-100K dataset.

each method are as follows. α is set to 0.9 for MMR. ϵ is set to 0.1 for ϵ-Greedy, and θ is set to 0.6 for DPP^{map}. In C^{2}UCB, we set $\lambda_0 = 100$, $\lambda = 0.1$, and $\sigma = 1$. In EC-Bandit, we set the parameter $\lambda = 1$. For DC^{2}B, we empirically set $\alpha = 3$, and $\lambda = 1$, on all datasets.

Performance Comparison

The recommendation accuracies and diversity of different algorithms are summarized in Table 2. As shown in Table 2, the proposed DC^{2}B method usually achieves the best recommendation accuracy (i.e., Precision@N) on ML-1M and Anime datsets, and achieves the second best accuracy on ML-100K dataset. For example, on Anime dataset, DC^{2}B significantly outperforms C^{2}UCB and EC-Bandit by 59.35% and 107.11%, and achieves 11.73%, 11.07%, 19.46%, and 11.78% better performances than LogRank, MMR, ϵ-Greedy, and DPP^{map}, in terms of Precision@50. These results indicate that DC^{2}B is more effective than baseline methods on large and sparse dataset. Moreover, we also note the combinatorial bandit methods C^{2}UCB and DC^{2}B significantly outperform EC-Bandit. One potential reason is

Table 3: Relative improvements of DC^{2}B over baselines. The positive improvements are highlighted in **bold**.

Methods	ML-100K			ML-1M			Anime		
	Prec.@50	Div.	F-m.	Prec.@50	Div.	F-m.	Prec.@50	Div.	F-m.
LogRank	**+14.96%**	**+1.17%**	**+11.36%**	**+9.71%**	-1.75%	**+6.59%**	**+11.73%**	**+0.38%**	**+9.20%**
MMR	**+15.33%**	-0.40%	**+11.22%**	**+10.38%**	-1.92%	**+7.60%**	**+11.07%**	-1.65%	**+ 8.22%**
ϵ-Greedy	**+18.45%**	-0.33%	**+13.53%**	**+13.55%**	-1.12%	**+9.58%**	**+19.46%**	-1.95%	**+14.67%**
DPP^{map}	**+12.84%**	**+1.67%**	**+9.92%**	**+11.04%**	**+2.36%**	**+8.69%**	**+11.78%**	-0.70%	**+8.97%**
C^{2}UCB	-8.39%	**+3.72%**	-5.21%	**+3.97%**	**+0.58%**	**+3.04%**	**+59.35%**	**+66.07%**	**+60.89%**
EC-Bandit	**+98.35%**	-2.85%	**+71.81%**	**+93.48%**	**+0.49%**	**+68.22%**	**+107.11%**	-1.21%	**+82.76%**
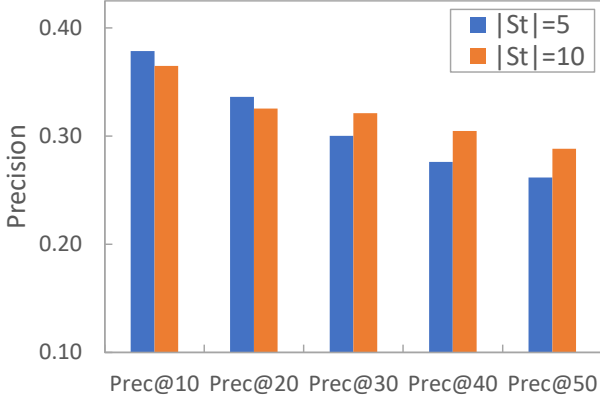
Figure 2: Performance trend of DC^{2}B with respect to different settings of $|\mathcal{S}_t|$ on ML-100K dataset.

that the combinatorial methods employ the user's feedback on a set of items to update model parameters. However, EC-Bandit uses the user's feedback on a single item to update model parameters. The parameter learning of C^{2}UCB and DC^{2}B is more stable than that of EC-Bandit, thus C^{2}UCB and DC^{2}B can achieve better recommendation accuracy. In addition, we can note that the non-interactive methods MMR, and ϵ-Greedy usually achieves slightly higher recommendation diversity than DC^{2}B, and DC^{2}B attains better recommendation diversity than DPP^{map} on ML-100K and ML-1M datasets. Comparing with interactive methods, Table 2 indicates that the recommendation diversity of DC^{2}B is higher than that of C^{2}UCB on all datasets, and EC-Bandit achieves higher recommendation diversity than DC^{2}B on ML-100K and Anime datasets.

For better understanding the results, F-measure is used to study the effectiveness of each recommendation algorithm in balancing the recommendation accuracy and diversity. Here, we use Precision@50 and Diversity to compute the F-measure. As shown in Table 2, we can note that DC^{2}B achieves the best F-measure value on ML-1M and Anime datasets, and the second best F-measure value on ML-100K dataset. In addition, we summarize the relative improvements of DC^{2}B over baseline methods on Precision@50, Diversity, and F-measure in Table 3. These results demonstrate that the proposed DC^{2}B method is more effective in balancing the recommendation accuracy and diversity than the baseline methods, especially on larger and sparser datasets.

Parameter Sensitivity Analysis

Moreover, we also evaluate the impacts of α and the size of super arm $|\mathcal{S}_t|$ on the performances of DC^{2}B, on ML-100K dataset. The parameter α is varied in $\{0.01, 0.1, 1, 3, 5, 10, 100\}$. Figure 1 shows the performances of DC^{2}B with respect to different settings of α. As shown in Figure 1, the recommendation accuracy in terms of Precision@50 firstly increases with the increase of α. When α is larger than 5, the recommendation accuracy of DC^{2}B drops drastically by changing α to 10 and 100. We can also note that the diversity value decreases with the increase of α, because larger α makes DC^{2}B focus more on the item qualities in generating recommendations. Overall, the results in Figure 1 indicate that α can effectively control the item qualities and item diversities when generating recommendations. Additionally, we vary the size of recommendation list $|\mathcal{S}_t|$ at each trial in $\{5, 10\}$. Figure 2 summarizes the accuracy of DC^{2}B with respect to different sizes of super arm. As shown in Figure 2, larger super arm size tends to results better recommendation accuracy, when enough number of interactions (e.g., more than 3 interactions) between the user and the recommender system have been performed. This is because the model updating based on the user's feedback on a larger set of items is expected to be more stable and accurate. Moreover, the average recommendation diversity with respect to $|\mathcal{S}_t| = 5$ and $|\mathcal{S}_t| = 10$ are 0.8110 and 0.8118, respectively. This indicates that $|\mathcal{S}_t|$ does not have significant impacts on diversity.

Conclusion and Future Work

This work proposes a novel bandit learning method, namely Diversified Contextual Combinatorial Bandit (DC^{2}B), for interactive recommendation based on users' implicit feedback. Specifically, DC^{2}B is a full Bayesian recommendation framework, which enables the interactions between recommender system and the user, and employs determinantal point process (DPP) to promote the recommendation diversity. We develop a Thompson sampling-type optimization algorithm to iteratively learn the model parameters, and conduct regret analysis to provide theoretical guarantee for DC^{2}B. Moreover, empirical experiments on real datasets also demonstrate the effectiveness of the proposed DC^{2}B in balancing the recommendation accuracy and diversity. As for the future work, we intend to develop more complex DPP kernels and efficient DPP inference algorithms for interactive recommender systems. In addition, we are also inter-

ested in developing more sophisticated models to describe the user engagements on the recommendation results.

Acknowledgments

References

Adomavicius, G., and Kwon, Y. 2011. Improving aggregate recommendation diversity using ranking-based techniques. *IEEE Transactions on Knowledge and Data Engineering* 24(5):896–911.

Antikacioglu, A., and Ravi, R. 2017. Post processing recommender systems for diversity. In *KDD*, 707–716. ACM.

Blei, D. M.; Kucukelbir, A.; and McAuliffe, J. D. 2017. Variational inference: A review for statisticians. *Journal of the American Statistical Association* 112(518).

Carbonell, J., and Goldstein, J. 1998. The use of mmr, diversity-based reranking for reordering documents and producing summaries. In *SIGIR*, 335–336. ACM.

Chen, L.; Zhang, G.; and Zhou, E. 2018. Fast greedy map inference for determinantal point process to improve recommendation diversity. In *NIPS*, 5622–5633.

Cheng, P.; Wang, S.; Ma, J.; Sun, J.; and Xiong, H. 2017. Learning to recommend accurate and diverse items. In *WWW*, 183–192. International World Wide Web Conferences Steering Committee.

Jaakkola, T., and Jordan, M. 1997. A variational approach to bayesian logistic regression models and their extensions. In *Sixth International Workshop on Artificial Intelligence and Statistics*, volume 82, 4.

Kulesza, A.; Taskar, B.; et al. 2012. Determinantal point processes for machine learning. *Foundations and Trends® in Machine Learning* 5(2–3).

Kunaver, M., and Požrl, T. 2017. Diversity in recommender systems–a survey. *Knowledge-Based Systems* 123:154–162.

Li, L.; Chu, W.; Langford, J.; and Schapire, R. E. 2010. A contextual-bandit approach to personalized news article recommendation. In *WWW*, 661–670. ACM.

Li, L.; Chu, W.; Langford, J.; and Wang, X. 2011. Unbiased offline evaluation of contextual-bandit-based news article recommendation algorithms. In *WSDM*, 297–306. ACM.

Liang, D.; Charlin, L.; McInerney, J.; and Blei, D. M. 2016. Modeling user exposure in recommendation. In *WWW*, 951–961. International World Wide Web Conferences Steering Committee.

Liu, Y.; Zhao, P.; Liu, X.; Wu, M.; Duan, L.; and Li, X.-L. 2017. Learning user dependencies for recommendation. In *IJCAI*, 2379–2385. AAAI Press.

Liu, Y.; Zhao, L.; Liu, G.; Lu, X.; Gao, P.; Li, X.-L.; and Jin, Z. 2018. Dynamic bayesian logistic matrix factorization for recommendation with implicit feedback. In *IJCAI*, 3463–3469. AAAI Press.

Qi, Y.; Wu, Q.; Wang, H.; Tang, J.; and Sun, M. 2018. Bandit learning with implicit feedback. In *NIPS*, 7276–7286.

Qin, L., and Zhu, X. 2013. Promoting diversity in recommendation by entropy regularizer. In *IJCAI*, 2698–2704. AAAI Press.

Qin, L.; Chen, S.; and Zhu, X. 2014. Contextual combinatorial bandit and its application on diversified online recommendation. In *SDM*, 461–469. SIAM.

Rendle, S.; Freudenthaler, C.; Gantner, Z.; and Schmidt-Thieme, L. 2009. Bpr: Bayesian personalized ranking from implicit feedback. In *UAI*, 452–461. AUAI Press.

Russo, D., and Van Roy, B. 2014. Learning to optimize via posterior sampling. *Mathematics of Operations Research* 39(4).

Sha, C.; Wu, X.; and Niu, J. 2016. A framework for recommending relevant and diverse items. In *IJCAI*, 3868–3874. AAAI Press.

Shi, Y.; Larson, M.; and Hanjalic, A. 2014. Collaborative filtering beyond the user-item matrix: A survey of the state of the art and future challenges. *ACM Computing Surveys* 47(1):3.

Steck, H.; van Zwol, R.; and Johnson, C. 2015. Interactive recommender systems: Tutorial. In *RecSys*, 359–360. ACM.

Tang, L.; Jiang, Y.; Li, L.; Zeng, C.; and Li, T. 2015. Personalized recommendation via parameter-free contextual bandits. In *SIGIR*, 323–332. ACM.

Wang, X.; Hoi, S. C.; Liu, C.; and Ester, M. 2017. Interactive social recommendation. In *CIKM*, 357–366. ACM.

Wang, S.; Hu, L.; Cao, L.; Huang, X.; Lian, D.; and Liu, W. 2018. Attention-based transactional context embedding for next-item recommendation. In *AAAI*, 2532–2539. AAAI Press.

Wang, S.; Hu, L.; Wang, Y.; Cao, Longbingand Sheng, Q. Z.; and Orgun, M. 2019. Sequential recommender systems: Challenges, progress and prospects. In *IJCAI*, 6332–6338. AAAI Press.

Wang, H.; Wu, Q.; and Wang, H. 2017. Factorization bandits for interactive recommendation. In *AAAI*, 2695–2702. AAAI Press.

Wilhelm, M.; Ramanathan, A.; Bonomo, A.; Jain, S.; Chi, E. H.; and Gillenwater, J. 2018. Practical diversified recommendations on youtube with determinantal point processes. In *CIKM*, 2165–2173. ACM.

Wu, Q.; Liu, Y.; Miao, C.; Zhao, B.; Zhao, Y.; and Guan, L. 2019a. Pd-gan: adversarial learning for personalized diversity-promoting recommendation. In *IJCAI*, 3870–3876. AAAI Press.

Wu, Q.; Liu, Y.; Miao, C.; Zhao, Y.; Guan, L.; and Tang, H. 2019b. Recent advances in diversified recommendation. *arXiv preprint arXiv:1905.06589*.

Yang, P.; Zhao, P.; Liu, Y.; and Gao, X. 2018. Robust cost-sensitive learning for recommendation with implicit feedback. In *SDM*, 621–629. SIAM.

Zhang, M., and Hurley, N. 2008. Avoiding monotony: improving the diversity of recommendation lists. In *RecSys*, 123–130. ACM.

Zhao, X.; Zhang, W.; and Wang, J. 2013. Interactive collaborative filtering. In *CIKM*, 1411–1420. ACM.

Zhou, T.; Kuscsik, Z.; Liu, J.-G.; Medo, M.; Wakeling, J. R.; and Zhang, Y.-C. 2010. Solving the apparent diversity-accuracy dilemma of recommender systems. *Proceedings of the National Academy of Sciences* 107(10):4511–4515.