A Framework for Recommending Relevant and Diverse Items

Chaofeng Sha, Xiaowei Wu, Junyu Niu
School of Computer Science, Fudan University
Shanghai Key Laboratory of Intelligent Information Processing
{cfsha,14212010020,jyniu}@fudan.edu.cn

Abstract

The traditional recommendation systems usually aim to improve the recommendation accuracy while overlooking the diversity within the recommended lists. Although some diversification techniques have been designed to recommend top-k items in terms of both relevance and diversity, the coverage of the user's interest is overlooked. In this paper, we propose a general framework to recommend relevant and diverse items which explicitly takes the coverage of user interest into account. Based on the theoretical analysis, we design efficient greedy algorithms to get the near optimal solutions for those NP-hard problems. Experimental results on MovieLens dataset demonstrate that our approach outperforms state-of-the-art techniques in terms of both precision and diversity.

1 Introduction

In traditional recommendation systems, a few top items are recommended to users. The traditional approaches to this top-k recommendation problem focus on the *relevance* of results only: they make use of some standard recommendation algorithm that computes the relevance score of each item with regard to the user individually then just return the top-k items with the highest relevance scores. However, many recent studies have shown theoretically and empirically that it is beneficial to take the *diversity* of results into account as well [Zhang and Hurley, 2008; Zhou *et al.*, 2010; Qin and Zhu, 2013], particularly when different users with diverse interest. Recommending a diverse set of the most relevant items is more likely to ensure that all potential needs of a given user are satisfied. So finding the best set of top-k items for a given user is of crucial importance to recommendation systems.

In order to resolve the user interest ambiguity and avoid the redundancy in the recommendation itemset, a number of diversification techniques [Zhang and Hurley, 2008; Zhou *et al.*, 2010; Qin and Zhu, 2013; Ashkan *et al.*, 2015] have been designed which optimize the top-k items collectively, in terms of both relevance and diversity. Most of these former work, e.g., Entropy Regularized [Qin and Zhu, 2013], defines some objective functions to tradeoff between maximizing relevance and maximizing diversity. Although the work of [Qin and Zhu, 2013] also coined the diversity promotion problem as a submodular function maximization problem with a theoretical guarantee, the intuition behind the diversification term using the entropy of rating of recommended items is not sound. Furthermore, the coverage of user's interest is not explicitly taken into account.

In this paper, we first develop an alternative framework for recommending relevant and diverse items. We formalize the problem as finding a top-k items which maximizes a linear combination of the relevance of the item, the coverage of the user's interest, and diversity between them. We detail three instantiations of the proposed framework. Due to the NP-hardness of the proposed problems, we employ greedy search algorithm with theoretical approximation guarantee of factor for each optimization problem. For one specific case (Case 2 in Section 4), we give a more detailed theoretical analysis and prove the submodularity of the proposed objective function. Compared to the previous approach taking advantage the submodular property of the optimization objective [Qin and Zhu, 2013], the proposed approach significantly improves not only the precision (i.e., the quality of the final result set) but also the diversity between the recommended items. The advantages of our new approach over the existing ones are confirmed by extensive experiments on MovieLens dataset. Our contributions can be summarized as follows:

1. We formalize item recommendation as a combinatorial optimization problem which combines the relevance of items, coverage of the user's interest and diversity between items.
2. We detail the framework under three cases. Based on the theoretical analysis, we design efficient greedy algorithms to get the near optimal solutions for those NP-hard problems.
3. We conduct extensive experiments on MovieLens dataset, which include performance comparisons with state-of-the-art methods. The experimental results demonstrate the effectiveness of our proposed solutions in terms of precision and diversity.

The rest of the paper is organized as follows. The related work is presented in Section 2. Then the latent matrix factorization and submodular function maximization is reviewed in Section 3. Section 4 is devoted to the formulation of the proposed framework and the algorithms. Experimental evaluation using real data are shown in Section 5. We conclude with future work in Section 6.

2 Related Work

The Collaborative Filtering (CF) which establishes the connection between users and items is commonly used in recommendation systems. Two popular approaches are the neighborhood models (user or item based) [X.Su and Khoshgoftaar, 2009] and the latent matrix factorization methods, e.g. Regression- Based Latent Factor Model [Agarwal and Chen, 2009], Probabilistic Matrix Factorization [Salakhutdinov and Mnih, 2008], and Singular Value Decomposition [Koren and Bell, 2011]. In this paper, we diversify items recommended by singular value decomposition.

In recent year, there has been some work on diversifying the recommendation lists to increase users satisfactions. These works can be classified into two categories based on the way how diversity is measured: content-based and temporal-based.

In content-based diversification, which is known as an instance of p-dispersion problem [Drosou and Pitoura, 2010]. Most of the content-based recommendation diversification methods so far are attributes based. In [Ziegler *et al.*, 2005], the diversity of recommendation lists is increased through maximizing the topic attribute difference between items. In [Yu *et al.*, 2009], item explanation is employed to measure the distance between items to achieve diversification instead of item attributes. The work of [Noia *et al.*, 2014] focuses on modeling users' inclination toward selecting diverse items, where diversity is computed by means of content-based item attributes. This modeling is exploited to re-rank the list of Top-N items predicted by a recommendation algorithm, in order to foster diversity in the final ranking. And the state-of-art conditional differential entropy $h(r_S|r_\Omega, V)$ is employed in [Qin and Zhu, 2013] as a diversity regularizer. However the entropy measures the average uncertainty of ratings, r_S, which might not characterize the diversity between items.

There are works to use temporal information for recommendation diversification, e.g. [Lathia *et al.*, 2010] examines the temporal characteristics of item rating patterns to increase recommendation diversity, and [Zhao *et al.*, 2012] increases temporal diversity with product purchase intervals. [Nguyen *et al.*, 2014] studies the "filter bubble" effect, which describe the potential for online personalization to effectively isolate people from a diversity of viewpoints or content, in terms of content diversity at the individual level and exposes users to a slightly narrowed set of items over time.

Besides diversifying the recommendation lists, the aggregate diversity is also introduced in works such as [Yin *et al.*, 2012; Niemann and Wolpers, 2013; Adomavicius and Kwon, 2012]. They treat diversity of item as the aggregate or equivalently average dissimilarity of all pairs of items in the recommended set.

3 Preliminary

Before diving into the proposed framework, we introduce the latent matrix factorization method and submodular function maximization, which our framework and the proposed algorithms are based on.

3.1 Latent Matrix Factorization

Matrix factorization models [Koren *et al.*, 2009] are the most pervasive method employed in recommendation systems, which map both users and items to a joint latent factor space of some dimensionality D, such that user-item interactions are modeled as inner products in that space. Accordingly, each user u is associated with a vector $p_u \in R^D$, and each item i is associated with a vector $q_i \in R^D$. For a given user u, the components of p_u measure the extent of interest the user has on the corresponding factors. For a given item i, the components of q_i measure the extent to which the item possesses those factors. The resulting dot product, $p_u \cdot q_i$, captures the interaction between user u and item i - the user's overall interest in the item's characteristics. This approximates user u's rating of item i, which is denoted by r_{ui}, leading to the estimation

$$\hat{r}_{ui} = p_u \cdot q_i.$$

To learn the factor vectors (p_u and q_i), the regularized squared error is minimized on the itemset with known ratings (training set):

$$\min_{p_u, q_i} \sum_{(u,i)\in R} (r_{ui} - p_u \cdot q_i)^2 + \lambda(||p_u||^2 + ||q_i||^2). \quad (1)$$

Here, R is the set of the (u, i) pairs for which r_{ui} is known (the training set). The parameters can be learned through stochastic gradient descent optimization method [Koren *et al.*, 2009].

3.2 Submodular Function Maximization

A function $f : 2^X \to R$ is submodular if for any $S \subseteq T \subseteq X$ and $i \notin T$, $f(S \cup \{i\}) - f(S) \geq f(T \cup \{i\}) - f(T)$. In other words, submodular functions are characterized by the diminishing return property: the marginal gain of adding an element to a smaller subset of X is higher than that of adding it to a larger subset of X. We are often interested in finding a subset of with cardinality constraint $|S| \leq k$ that maximizes a submodular function, i.e., $\arg\max_S f(S)$ [Krause and Golovin, 2012]. However this problem is in general NP-hard, even for some simple submodular functions such as mutual information [Krause and Guestrin, 2005]. In recent years, submodular functions have been receiving much attention from the Artificial Intelligence community because of the insight about their approximate optimization [Wei *et al.*, 2015; Vanchinathan *et al.*, 2015]. The submodularity of the objective function provides a theoretical approximation guarantee of factor $1 - \frac{1}{e}$ for the (best-first) greedy search algorithm [Nemhauser *et al.*, 1978]:

Theorem 1. Let f be a submodular, (monotone) nondecreasing set function and $f(\emptyset) = 0$. Then, the greedy (search) algorithm finds a set S ($|S| = k$) such that

$$f(S) \geq \left(1 - \frac{1}{e}\right) \max_{S':|S'|\leq k} f(S')$$

4 Proposed Approach

In this section, we give the main contribution of our work: formulating the item recommendation as a combinatorial optimization or search problem, which takes the relevance of items, the coverage of user's interest and diversity between items into account.

4.1 Diversification Framework

For a given user u and the itemset Ω that u rated, we aim to recommend the top-k relevant and diverse item subset $S \subseteq I \setminus \Omega$ with the size constraint $|S| = k$. We leverage the estimated parameter through matrix factorization for the task. One possible strategy is to select the items which user u has interest and are similar to the highly rated ones. However, this simple idea does not account for similarity between the recommended items that could be redundant. Therefore, a natural objective is to maximize the sum of the relevance and dissimilarity of the recommended items which also explicitly takes user's interest into account. Formally, we recommend the itemset $S \subseteq I \setminus \Omega$ ($|S| = k$) which maximizes the following objective function:

$$g(S) = \sum_{i \in S} p_u \cdot q_i + \alpha \sum_{j \in \Omega} w_j \max_{i \in S} q_i \cdot q_j + \beta \sum_{i,j \in S} ||q_i - q_j||, \tag{2}$$

where α and β are the control parameters, and $||q_i - q_j||$ is the Euclidean distance between item vectors q_i and q_j. The weight w_j of item j can be set as 1 or $w_j = e^{\gamma - (r_{max} - r_{uj})}$. The intuition behind w_j is that maybe the user is interested in some kinds (or type, genre) of items while the rated ones are with low quality. Therefore we could introduce the weight w_j to quantify the "importance" of item j.

For ease of discussion, we use the notation $f(S) = \sum_{i \in S} p_u \cdot q_i + \alpha \sum_{j \in \Omega} w_j \max_{i \in S} q_i \cdot q_j$ and $d(S) = \sum_{i,j \in S} ||q_i - q_j||$. The term $f(S)$ in function (2) is used to control the relevance of the selected items and the coverage of the user's interest which is achieve through tradeoff the similarity of recommended items to the rated items. The intuition behind $\sum_{j \in \Omega} w_j \max_{i \in S} q_i \cdot q_j$ is that we recommend the items $q_i \in S$ similar to those rated by the given user ($q_j \in \Omega$) to cover the user's potential interest, which bears resemblance to item-based nearest neighbor methods. Furthermore, it can diversify the recommended items due to the return diminishing property as proved in the following subsection. The dispersion term $d(S)$ in (2) is used to enhance the diversity within the recommended itemset S.

When both parameters α and β are set to 0 ($\alpha = 0, \beta = 0$), the objective function (2) is degenerated to the naive top-k item recommendation problem which just returns the items with the highest estimated rating. It is regarded as one of the baselines in the experimental evaluation. We detail other settings in the following subsections.

4.2 Case 1: $\alpha = 0, \beta > 0$

In this case, the objective function is sum of modular function ($\sum_{i \in S} p_u \cdot q_i$) and dispersion function ($\beta d(S)$). The intuition is that we recommend the items with highly estimated rating and with large distance between them. The optimization problem can be seen as a variant of Maximum Sum Dispersion problem [Hassin *et al.*, 1997] which has been applied in diversifications such as search result diversification [Gollapudi and Sharma, 2009]. Generally, the maximum sum dispersion problem is NP-hard, therefore we resort to an approximation algorithm. In this paper, we employ a greedy algorithm of [Hassin *et al.*, 1997] which is a $\frac{1}{2}$-approximation algorithm. The algorithm is outlined in Algorithm 1, where the distance measure is defined as $d'(i,j) = p_u \cdot q_i + p_u \cdot q_j + \beta ||q_i - q_j||$ which is metric thanks to the fact that the Euclidean distance $||q_i - q_j||$ is a metric.

Algorithm 1 `Greedy Search for Modular-MaxSum Dispersion`

Input: The item matrix Q, the itemset Ω rated by u, and integer k
Output: Item subset $S \subseteq I \setminus \Omega$ with $|S| = k$

$S \leftarrow \Phi$;
for $iter = 1$ to $iter < k/2$ **do**
  $(i, j) \leftarrow \arg\max_{i',j' \in I-S} d'(i', j')$
  $S \leftarrow S \cup \{i, j\}$
end for
if k is odd, add an arbitrary item to S
return S

4.3 Case 2: $\alpha > 0, \beta = 0$

When setting $\beta = 0$, we have that $g(S) = f(S)$. In this case, we aim to recommend the items with high rating and similar to the rated ones. However the optimization problem $\max_{S:|S| \leq K} f(S)$ is also NP-hard, which can be reduced from facility location problem. Fortunately, the simple greedy hill-climbing heuristic mentioned in section 3.2 can be used to get an $1 - \frac{1}{e}$ approximation ratio thanks to the properties of function $f(S)$ established in the following theorem.

Theorem 2. $f(S)$ is monotonic and submodular.

Proof. For any $S \subseteq T$ and $i \notin T$,

$$\begin{aligned} & f(S \cup \{i\}) - f(S) \\ = & \sum_{l \in S \cup \{i\}} p_u \cdot q_l + \alpha \sum_{j \in \Omega} w_j \max_{l \in S \cup \{i\}} q_l \cdot q_j \\ & - \sum_{l \in S} p_u \cdot q_l - \alpha \sum_{j \in \Omega} w_j \max_{l \in S} q_l \cdot q_j \\ = & p_u \cdot q_i + \alpha \sum_{j \in \Omega} w_j \max\{0, q_i \cdot q_j - \max_{l \in S} q_l \cdot q_j\} \end{aligned}$$

As all terms in the last expression are nonnegative, we know that $f(S \cup \{i\}) \geq f(S)$, i.e. the function $f(S)$ is monotonic non-decreasing.

Similarly, we have

$$f(T \cup \{i\}) - f(T) = p_u \cdot q_i + \alpha \sum_{j \in \Omega} w_j \max\{0, q_i \cdot q_j - \max_{l \in T} q_l \cdot q_j\} \tag{3}$$

We know that $\max_{l \in S} q_l \cdot q_j \leq \max_{l \in T} q_l \cdot q_j$ for each item j thanks to the fact that $S \subseteq T$. It is followed that $\sum_{j \in \Omega} w_j \max\{0, q_i \cdot q_j - \max_{l \in S} q_l \cdot q_j\} \geq$

Algorithm 2 `Greedy Search for Submodular Function Maximization`

Input: The item matrix Q, the itemset Ω rated by u, and integer k
Output: Item subset $S \subseteq I \setminus \Omega$ with $|S| = k$

$S \leftarrow \{\arg\max_{i \in I \setminus \Omega} p_u \cdot q_i\}$;
while $|S| < k$ **do**
  $i \leftarrow \arg\max_{i' \in I-S} p_u \cdot q_{i'} + \alpha \sum_{j\in\Omega} w_j \max\{0, q_{i'} \cdot q_j - \max_{l\in S} q_l \cdot q_j\}$
  $S \leftarrow S \cup \{i\}$
end while
return S

$\sum_{j\in\Omega} w_j \max\{0, q_i \cdot q_j - \max_{l\in T} q_l \cdot q_j\}$. Therefore we have that $f(S \cup \{i\}) - f(S) \geq f(T \cup \{i\}) - f(T)$, i.e. the function $f(S)$ is submodular.

Note that although the explicit dispersion between the items is not considered in this case, the submodularity or diminishing property of $f(S)$ can be used to diversify the selected items. The near optimal solution is presented in Algorithm 2, where the item maximizing $p_u \cdot q_i + \alpha \sum_{j\in\Omega} w_j \max\{0, q_i \cdot q_j - \max_{l\in S} q_l \cdot q_j\}$ is selected into S incrementally after the first item.

4.4 Case 3: $\alpha > 0, \beta > 0$

In this case, we enhance the diversity between recommended items through including the sum dispersion of items. The optimization problem $\max_{S:|S|\leq k} g(S)$ is also NP-hard. Fortunately, the approximation algorithm proposed in [Borodin *et al.*, 2012] can be employed in our solution. As stated in [Borodin *et al.*, 2012], the approximation algorithm has $\frac{1}{2}$ approximation ratio when the objective function is a linear combination of a submodular function and the sum of pairwise distance. We have proved the submodularity of function $f(S)$ in the above case. Therefore the objective function proposed in this paper, $g(S) = f(S) + \beta d(S)$, is a variant of the formulation developed in [Borodin *et al.*, 2012].

To be specific, we denote $d(S) = \sum_{\{i,j\}:i,j\in S} ||q_i - q_j||$ and $d(S,T) = \sum_{\{i,j\}:i\in S, j\in T} ||q_i - q_j||$ for any disjoint subsets $S, T \subseteq I$. For any given subset $S \subseteq I$ and any item $i \in I - S$, let $d_i(S) = \sum_{j\in S} ||q_i - q_j||$ be the marginal gain on the distance, $f_i(S) = f(S \cup \{i\}) - f(S)$ be the marginal gain on the weight, and $g_i(S) = f_i(S) + \beta d_i(S)$ be the total marginal gain on the objective function. Let $f'_i(S) = \frac{1}{2} f_i(S)$, and $g'_i(S) = f'_i(S) + \beta d_i(S)$. The final solution is presented in Algorithm 3.

5 Experiments

We evaluate our proposed method on real dataset (MovieLens dataset) compared to the current state-of-the-art diversification methods.

5.1 Data Set

The experiments are carried out on publicly available rating datasets, MovieLens dataset. It consists of 1,000,209 ratings

Algorithm 3 `Greedy Search for Submodular-MaxSum Dispersion`

Input: The item matrix Q, the itemset Ω rated by u, and integer k
Output: Item subset $S \subseteq I \setminus \Omega$ with $|S| = k$

$S \leftarrow \phi$;
while $|S| < k$ **do**
  $i \leftarrow \arg\max_{i\in I-S} g'_i(S)$
  $S \leftarrow S \cup \{i\}$
end while
return S

for 3952 movies by 6040 users of homonym online movie recommender service. All users are selected randomly, and each of them has rated at least 20 movies. After the elimination of the cold-start movies (with less than 10 ratings), we still have 3385 movies and 6040 users. The element of the dataset is represented by a tuple: $t_{u,i} = (u, i, r_{u,i})$, where u denotes userID, i denotes movieID, and $r_{u,i}$, which is an integer score between 1 and 5, denotes the rating of user u for movie i (higher score indicates higher preference).

5.2 Baselines

In order to show the performance improvement of our approach, we compare our approach with the following baselines:
1. PMF: PMF is the standard probabilistic matrix factorization, which is considered as the state-of-art recommendation method. Specifically, we seek to find a set maximizes the following criteria:

$$L(S) = \sum_{i\in S} p_u q_i$$

which is the case that both α and β are set to zero in our framework.
2. PMF+ER: Entropy Regularizer is proposed in [Qin and Zhu, 2013] to promote the diversity within the recommended itemset. More specifically, ER seeks to find a set maximizes following criteria

$$ER(S) = R(S) + \lambda g(S)$$

where $R(S) = \sum_{\omega\in S}(E[r_\omega|r_\Omega, V] - c)$ and $g(S) = h(r_S|r_\Omega, V)$.

5.3 Evaluation Metrics

Following the work of [Zhang and Hurley, 2008; Qin and Zhu, 2013], we split the dataset into a training dataset Y_T and a test dataset Y_P by randomly assigning 50% of tuples in MovieLens to each set. The training dataset Y_T was used to train the PMF model, such that we can get the item vectors. And we define user profile P_u as the set of movies rated by u, and test user profile T_u is the 50% of P_u in test dataset Y_P. Let $|P_u|$ and $|T_u|$ denote the size of the profile.

We use two different metrics to measure the quality of the recommendation list: 1. Precision which is used to measure the accuracy of our proposed approach or the relevance of the

k	D=80			D=100			D=120		
	PMF	PMF+ER	PMF+α+β	PMF	PMF+ER	PMF+α+β	PMF	PMF+ER	PMF+α+β
5	0.19098	0.23910	0.24962	0.16992	0.21203	0.26767	0.16692	0.24511	0.26617
10	0.18647	0.23609	0.27218	0.17970	0.21278	0.30226	0.16992	0.25940	0.25489
15	0.19900	0.25764	0.27820	0.19348	0.24762	0.28872	0.19048	0.26065	0.28822
20	0.20564	0.25639	0.27669	0.19323	0.24549	0.26729	0.18571	0.26015	0.29962
25	0.20361	0.24301	0.26977	0.18887	0.24571	0.26917	0.18135	0.24752	0.30346
30	0.19875	0.23509	0.25714	0.18095	0.23584	0.25965	0.18095	0.23860	0.29449

Table 1: Precision comparison with baselines

k	D=80			D=100			D=120		
	PMF	PMF+ER	PMF+α+β	PMF	PMF+ER	PMF+α+β	PMF	PMF+ER	PMF+α+β
5	0.62417	0.66682	0.71499	0.60356	0.63085	0.76089	0.61843	0.6706	0.79084
10	0.61265	0.65389	0.72330	0.61071	0.64156	0.76655	0.62124	0.69289	0.78897
15	0.6192	0.65975	0.72756	0.62232	0.66056	0.77513	0.63084	0.69546	0.79219
20	0.61800	0.66156	0.72482	0.62866	0.67247	0.77689	0.63591	0.70072	0.79236
25	0.61261	0.65461	0.71684	0.62184	0.67204	0.77753	0.63290	0.69332	0.79139
30	0.60848	0.65218	0.72209	0.61348	0.66615	0.77564	0.63054	0.68832	0.79014

Table 2: Diversity comparison with baselines

recommended items. 2. The diversity between items recommended by diversification methods.

Formally, when a itemset S_u with size $|S_u| = k$ is recommended to user u, the precision is measured as follows:

$$precision = \frac{1}{|U|} \sum_{u \in U} \frac{|T_u \cap S_u|}{|T_u|}$$

where U denotes the set of all users. Note that we do not aim to predict the ratings of movies, but to make a more satisfying top-k recommendation list [Qin and Zhu, 2013]. The precision is a more appropriate metric than the Mean Absolute Error (MAE).

To measure the average dissimilarity of all pairs of elements contained in S_u, the diversity between items is defined as follows [Nguyen *et al.*, 2014]:

$$diversity(S_u) = \frac{2}{k(k-1)} \sum_{i \in S_u} \sum_{j \neq i \in S_u} d(i, j)$$

where $k = |S_u|$, and $d(i, j) = 1 - \frac{q_i q_j}{\|q_i\| \|q_j\|}$ is the distance or dissimilarity between items $i, j \in S_u$.

5.4 Results

For the parameter setting, we set the dimensionality of the latent space $D = 80/100/120$ when training the PMF model. Both the baselines and our approach take the same settings, and we choose to use a learning rate of 20, regularization parameter of 0.1, and a momentum of 0.3. In addition, for "PMF+ER", we set $\lambda = 1$ which is the best result compared to other settings. While for our approach, we set $\alpha = 0.3$ and $\beta = 1.5$. We will examine the impact of parameters, α and β in the next section.

The experimental comparisons between three methods in terms of Precision and Diversity are shown in Table 1 and 2, where our approach is labeled as "PMF+$\alpha + \beta$". In those experiments, we vary the number of recommended items, i.e. we recommend top-5, 10, 15, 20, 25, 30 movies to each user. As excepted that PMF provides the worst performance. Our approach significantly outperforms both "PMF" and "PMR+ER". The results in Table 1 shows that the diversity promoting methods (both ER and β) can match the users' preference. As can be seen rom the Table 2, our method "PMF+α+β" is more effective for capturing the notion of diversity between items compared to "PMF+ER". For individual case of top-k prediction, where $k = 5/10/15/20/25/30$, our approach outperforms both baselines in terms of precision and diversity. The introduction of coverage of user's interest in our objective is beneficial to the promoting the precision and diversity compared to "PMR+ER". As we mentioned in Section 1, the diversity between item characteristic can not be captured by the entropy of the rating scores.

Impact of α and β

In our approach, the parameter α controls the matching with user's kind (or type, genre) interest and the coverage of the user's interest, and β controls the diversity. When $\alpha = \beta = 0$, the method is degenerated to find the movies with the highest rating. If $\alpha = +\infty$ and $\beta = 0$, the model seeks to find the movies that are most similar to user preferred kind (or type, genre). If $\alpha = 0$ and $\beta = +\infty$, the model seeks to find the most diverse movies. And if $\alpha > 0$ and $\beta > 0$, we will get a result which tradeoff between user's preferred kind, the coverage of user's interest and the diversity.

The impacts of α and β on Precision and Diversity are shown in Figure 1 and 2 respectively. Here we also include the precision by baselines PMF and PMF+ER as a control. As can be seen that in every case of top-k prediction (i.e. $k \in 10, 20, 30$), PMF+α+β outperforms PMF+ER and PMF in terms of both precision and diversity. More remarkable, as α stays 0 and β increases, while the prediction precision by PMF+α+β increases at first, then it decreases to steady

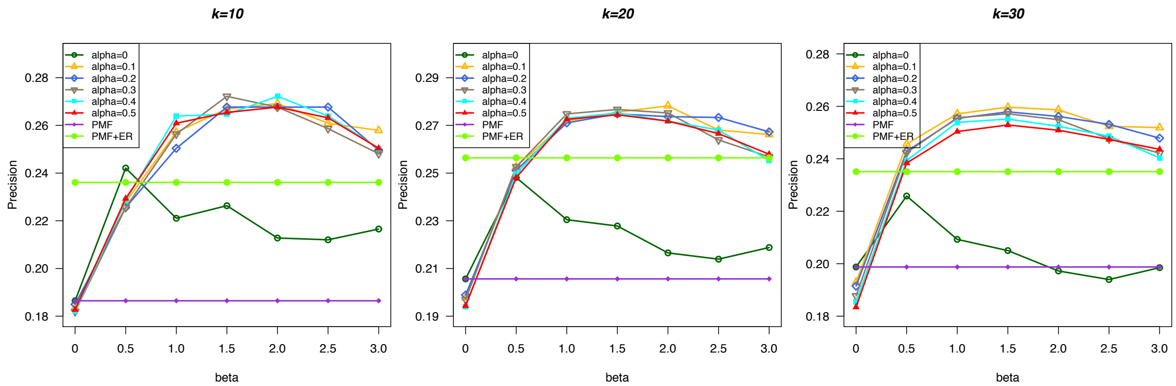

Figure 1: Precision: Impact of α and β

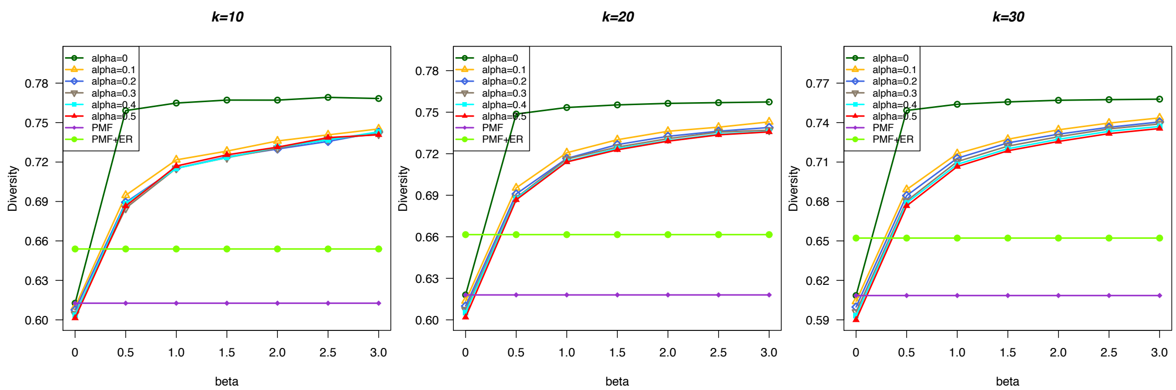

Figure 2: Diversity: Impact of α and β

and with little rebound sometimes. More interestingly, as α varies between 0.1 and 0.5 and β increases, the prediction precision by PMF+α+β increase at beginning, but when the value of β reaches a certain threshold, the precision begins to decrease. Compared with the case that $\alpha = 0$, when α varies between 0.1 and 0.5, then precision increases then stays more smoothly. It is reasonable that the effectiveness is due to the coverage of the user's interest when $\alpha > 0$ and it has the marginal return diminishing property (submodularity). As for the prediction of diversity by PMF+α+β, with the increase of β, it increases at first, and then it is stable. Comparing the case that $\alpha = 0$ and varies between 0.1 and 0.5, when α stays 0, it is more diverse between items. It shows that there is a tradeoff between enhancing the personalization of user and increasing the information of recommendation.

6 Conclusion

In this paper, we propose a general framework to recommend relevant and diverse items. We detail the framework through studying three cases and providing theoretical analysis and designing approximation algorithms for these NP-hard problems. Experimental results on MovieLens dataset demonstrate that our approach outperforms state-of-the-art techniques in terms of both accuracy and diversity. We leave the speedup of the diversification process as pointed out in [Ashkan *et al.*, 2015] and introduction of more explicit user's interest distribution into our framework as the future work.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (NFSC) under Grants No. 61540002 and No. 61572135.

References

[Adomavicius and Kwon, 2012] G. Adomavicius and Y.O. Kwon. Improving aggregate recommendation diversity using ranking-based techniques. *IEEE Transactions on Knowledge and Data Engineering*, 24:896–911, 2012.

[Agarwal and Chen, 2009] D. Agarwal and B.C. Chen. Regression-based latent factor models. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 19–28, 2009.

[Ashkan *et al.*, 2015] A. Ashkan, B. Kveton, S. Berkovsky, and Z. Wen. Optimal greedy diversity for recommendation. In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence*, pages 1742–1748, 2015.

[Borodin *et al.*, 2012] A. Borodin, H.C. Lee, and Y. Ye. Max-sum diversification, monotone submodular functions and dynamic updates. In *PODS*, 2012.

[Drosou and Pitoura, 2010] M. Drosou and E. Pitoura. Search result diversification. *SIGMOD Record*, 39(1):41–47, 2010.

[Gollapudi and Sharma, 2009] S. Gollapudi and A. Sharma. An axiomatic approach for result diversification. In *WWW*, pages 381–390, 2009.

[Hassin *et al.*, 1997] R. Hassin, S. Rubinstein, and A. Tamir. Approximation algorithms for maximum dispersion. *Operations Research Letters*, 21(3):133–137, 1997.

[Koren and Bell, 2011] Y. Koren and R. Bell. Advances in collaborative filtering. *Recommender Systems Handbook*, pages 145–186, 2011.

[Koren *et al.*, 2009] Yehuda Koren, Robert Bell, and Chris Volinsky. Matrix factorization techniques for recommendation systems. *IEEE Computers*, 2009.

[Krause and Golovin, 2012] Andreas Krause and Daniel Golovin. Submodular function maximization. *Chapter in Tractability: Practical Approaches to Hard Problems (to appear), Cambridge University Press*, 2012.

[Krause and Guestrin, 2005] Andreas Krause and Carlos Guestrin. Near-optimal nonmyopic value of information in graphical models. In *Uncertainty in Artificial Intelligence*, pages 324–331, 2005.

[Lathia *et al.*, 2010] N. Lathia, S. Hailes, L. Capra, and X. Amatriain. Temporal diversity in recommender systems. In *SIGIR*, pages 210–217, 2010.

[Nemhauser *et al.*, 1978] G.L. Nemhauser, L.A.Wolsey, and M.L. Fisher. An analysis of approximations for maximizing submodular set functions - i. *Mathematical Programming*, 14(1):265–294, 1978.

[Nguyen *et al.*, 2014] Tien T. Nguyen, Pik-Mai Hui, F. Maxwell Harper, Loren Terveen, and Joseph A. Konstan. Exploring the filter bubble: The effect of using recommender systems on content diversity. In *WWW*, pages 677–686, 2014.

[Niemann and Wolpers, 2013] Katja Niemann and Martin Wolpers. A new collaborative filtering approach for increasing the aggregate diversity of recommender systems. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 955–963, 2013.

[Noia *et al.*, 2014] Tommaso Di Noia, Vito Claudio Ostuni, Jessica Rosati, Paolo Tomeo, and Eugenio Di Sciascio. An analysis of users propensity toward diversity in recommendations. In *RecSys*, 2014.

[Qin and Zhu, 2013] Lijing Qin and Xiaoyan Zhu. Promoting diversity in recommendation by entropy regularizer. In *IJCAI*, pages 2698–2704, 2013.

[Salakhutdinov and Mnih, 2008] R. Salakhutdinov and A. Mnih. Probabilistic matrix factorization. In *Advances in neural information processing systems*, pages 1257–1264, 2008.

[Vanchinathan *et al.*, 2015] Hastagiri Vanchinathan, Andreas Marfurt, Charles-Antoine Robelin, Donald Kossmann, and Andreas Krause. Discovering valuable items from massive data. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2015.

[Wei *et al.*, 2015] Kai Wei, Rishabh Iyer, and Jeff Bilmes. Submodularity in data subset selection and active learning. In *International Conference on Machine Learning (ICML)*, Lille, France, 2015.

[X.Su and Khoshgoftaar, 2009] X.Su and T.M. Khoshgoftaar. A survey of collaborative filtering techniques. *Advances in Artificial Intelligence*, 2009(12), 2009.

[Yin *et al.*, 2012] Hongzhi Yin, Bin Cui, Jing Li, Junjie Yao, and Chen Chen. Challenging the long tail recommendation. *PVLDB*, pages 896–907, 2012.

[Yu *et al.*, 2009] C. Yu, L. Lakshmanan, and S. Amer-Yahia. It takes variety to make a world: diversification in recommender systems. In *EDBT*, pages 368–378, 2009.

[Zhang and Hurley, 2008] Mi Zhang and Neil Hurley. Avoiding monotony: improving the diversity of recommendation lists. In *Proceedings of the 2008 ACM Conference on Recommender Systems, RecSys 2008, Lausanne, Switzerland, October 23-25, 2008*, pages 123–130, 2008.

[Zhao *et al.*, 2012] G. Zhao, M. L. Lee, and W. Hsu. Increasing temporal diversity with purchase intervals. In *ACM SIGIR*, pages 165–174, 2012.

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

[Ziegler *et al.*, 2005] C. Ziegler, S. McNee, J. Konstan, and G. Lausen. Improving recommendation lists through topic diversification. In *WWW*, pages 22–32, 2005.