Set-oriented Personalized Ranking for Diversified Top-N Recommendation

Ruilong Su, Li'Ang Yin, Kailong Chen, Yong Yu
Dept.of Computer Science and Engineering
Shanghai Jiao Tong University
800 Dongchuan Road, Shanghai China, 200240
{rlsu, yinla, chenkl, yyu}@apex.sjtu.edu.cn

ABSTRACT

In this paper, we propose a set-oriented personalized ranking model for diversified top-N recommendation. Users may have various individual ranges of interests. For personalized top-N recommendation task, the combination of relevance and diversity in recommendation results would be desirable. For this purpose, we integrate the concept of diversity into traditional matrix factorization model to construct a set-oriented collaborative filtering model. By optimizing this model with a set-oriented pairwise ranking method, we directly achieve personalized top-N recommendation results which are both relevant and diversified. We also utilize category information explicitly for learning personalized diversity. Experimental results show that our model outperforms traditional models in terms of personalized diversity and maintains good performance on relevance prediction.

Categories and Subject Descriptors

H.3.3 [**Information Systems**]: Information Search and Retrieval—*Information Filtering*

Keywords

Collaborative Filtering, Recommender Systems, Set-oriented Pairwise Ranking, Personalized Diversity

1. INTRODUCTION

Collaborative Filtering (CF) [2] is one of most popular techniques in providing users with personalized suggestions of items or other information. The basic intuition behind CF is that users who had similar ranges of interests in the past are likely to have similar preferences in the future, and the more similar they were in the past, the more likely they will make similar choices in the future.

Top-N recommendation results are often presented to a target user in the form of a rank list according to their predicted relevance scores. In recent years, diversity of the results list has been regarded as another important factor that influences the satisfaction of users to a recommender system significantly. A recommendation list which contains different types of relevant items would be more desirable. In order to achieve this goal, most traditional methods rerank the relevance prediction results to balance relevance and diversity [8, 9]. We believe that relevance and diversity should be considered together. As personalized relevance can be predicted through learning, personalized diversity can be also predicted directly. Unlike relevance, diversity is a set-oriented concept. To quantify this concept, we construct a set-oriented CF model. We optimize this model for ranking by using a set-oriented pairwise ranking method and directly achieve personalized top-N recommendation results which are both relevant and diversified.

The paper is organized as follows: In Section 2, we review some related work briefly. Then in Section 3, we present the details of our set-oriented personalized ranking model. Our experimental results and analysis are reported in Section 4. Finally, Section 5 concludes the paper.

2. RELATED WORK

During the past years, collaborative filtering has been used as an effective technique for recommender systems [2]. In collaborative filtering, latent factor models usually have promising performance on predictive accuracy and scalability. Most latent factor models are based on factorizing the user-item rating matrix, which are known as Singular Value Decomposition (SVD) models [4].

Ranking-oriented models have been proposed recently in order to tackle the task of top-N recommendation more practically. A generic optimization criterion BPR was proposed in [5]. The experiments empirically showed that for the task of personalized ranking, learning a CF model with BPR outperforms other learning methods.

Diversity is regarded as another important aspect for the recommendation quality besides relevance. Previous studies showed that diversity of a recommendation list may hamper relevance to some degree, but will improve user satisfaction as a whole [3]. For traditional work about diversity, the concept of user is usually a generic *average* user, thus the diversified top-N recommendation results are non-personalized. In recent years, many diversification methods have been proposed [8, 9], including some adaptive diversification methods [6]. But most of these methods are two-step methods that consider relevance and diversity separately. In contrast with this, we combine relevance and diversity in a unified model which adaptively balances them for users with different interests in one step.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
RecSys'13, October 12–16, 2013, Hong Kong, China.
Copyright 2013 ACM 978-1-4503-2409-0/13/10 ...$15.00.
http://dx.doi.org/10.1145/2507157.2507207.

3. SET-ORIENTED PERSONALIZED RANKING (SPR)

In this section, we present our set-oriented personalized ranking model. The objective of this model is to directly recommend personalized item sets of size N which are both relevant and diversified to users with different ranges of interests. Key notations are shown in Table 1.

3.1 Set-oriented Formalization

Let U be the set of users and I be the set of items. Ratings of U on I are scored in a matrix $R(m \times n)$. Entry r_{ui} in R denotes the rating that user u gives to item i and reflects the preference of user u for item i. Let I_u be the set of items whose ratings are observed in matrix R for user u. We define a subset $I_u^+ \subseteq I_u$ containing items that user u likes.

$$I_u^+ := \{i | i \in I_u, r_{ui} \geq r^+\}$$

Where r^+ represents a rating threshold. We assume that a user prefers all items in I_u^+ over all items in $I_u^- := I \setminus I_u^+$. Previous pairwise ranking approaches use item pairs as training data. In order to quantify the recommendation diversity, in our work we use item set pairs as training data. We define T_u^+ as a candidate positive item set of size N.

$$T_u^+ := \{i | i \in I_u^+\}$$

We denote the set of T_u^+ as $\mathcal{T}_u^+$. Similarly, we define a common item set of size N as:

$$T_u := \{i | i \in I_u\}$$

We also denote the set of T_u as $\mathcal{T}_u$. A training pair (u, S_i, S_j) for user u is constructed by selecting two item sets S_i and S_j from $\mathcal{T}_u^+$ and $\mathcal{T}_u$ respectively, accordingly the set-oriented training collection is formalized as follows:

$$D_S := \{(u, S_i, S_j) | S_i \in \mathcal{T}_u^+ \wedge S_j \in \mathcal{T}_u \wedge (Con1 \vee Con2)\}$$

In which,

$$\begin{aligned} Con1 &:= (r_{s_i} > r_{s_j}) \\ Con2 &:= (r_{s_i} = r_{s_j}) \wedge (d_{s_i} > d_{s_j}) \end{aligned}$$

Where r_s denotes the relevance score of a given item set S and d_s denotes the diversity score of this set. The set S_i is preferred to S_j by considering the two conditions above: 1) the relevance score r_{s_i} of S_i is higher than r_{s_j} of S_j; 2) the diversity score d_{s_i} of S_i is higher than d_{s_j} of S_j, if the relevance scores of S_i and S_j are equal. We will discuss how to calculate the r_s and d_s later.

3.2 Set-oriented Collaborative Filtering Model

As a famous class of latent factor models, SVD models factorize the user-item rating to the inner product of user latent factor p_u and item latent factor q_i, where both of p_u and q_i are f-dimension. In the basic SVD model, the rating that user u gives to item i is predicted as follows:

$$\hat{r}_{ui} = \mu + b_u + b_i + p_u^{\mathrm{T}} q_i \tag{1}$$

In order to learn the diversity together with relevance, we introduce a new concept named "Set Diversity Bias" (SDB) representing the diversification degree of an item set. By combining SDB and the basic SVD model, we propose a set-oriented CF model to directly predict the score of user

Table 1: Key Notations

Representations	Descriptions
u, i, c	user u, item i, category c
U, I, C	user set, item set, category set
μ	overall average rating
b_u, b_i	users/items bias
$b_d(S_k)$	set diversity bias (SDB)
$bias$	μ, b_u, b_i, $b_d(S_k)$
p_u, q_i	latent factors for user-item affinity
v_i	latent factors for item-item similarity
p_c, w_i	latent factors for category-item relation

u on a given item set S_k. The score is decided by both relevance and diversity of S_k. It is formulated as this:

$$\hat{r}_{u,S_k} = \mu + b_u + \sum_{i \in S_k} b_i + p_u^{\mathrm{T}} (\sum_{i \in S_k} q_i) + \lambda b_d(S_k) \tag{2}$$

where $b_d(S_k)$ denotes the diversity bias of S_k and λ is a constant to control the weight of diversity in the model.

3.3 Set-oriented AUC Optimization

Many traditional personalized ranking methods maximize the metric of area under the ROC curve (AUC) [1]. In order to learn the set-oriented CF model (2), we maximize a set-oriented AUC. With the notation of D_S, it can be formalized as:

$$AUC(u) := \frac{1}{|D_S|} \sum_{(u,S_i,S_j) \in D_S} \delta(\hat{r}_{u,S_i} - \hat{r}_{u,S_j}) \tag{3}$$

Where $\hat{r}_{u,S_i}$ is the predicted preference score of user u on item set S_i. A value of $(\hat{r}_{u,S_i} - \hat{r}_{u,S_j}) > 0$ means the model predicts that user u prefers item set S_i to item set S_j. We use logistic loss function as $\delta(x)$:

$$\delta(x) = \frac{1}{1 + e^{-x}} \tag{4}$$

Which is a surrogate loss function and corresponds to Bayesian Personalized Ranking (BPR) [5]. Now we can derive our loss function for set-oriented AUC optimization as follows:

$$l(\theta) = \sum_{(u,S_i,S_j) \in D_S} ln(1 + e^{-(\hat{r}_{u,S_i} - \hat{r}_{u,S_j})}) + \lambda_r ||\theta||^2 \tag{5}$$

Where θ represents the unknown parameters in (2) and λ_r is a constant for regularization. We solve this loss function with a stochastic gradient descent algorithm. In every single training step, we use the bootstrap sampling to randomly select a triple (u, S_i, S_j) from D_S for updating our model.

3.4 Set Diversity Bias (SDB)

For the effective learning of diversity, set diversity bias should be able to reflect the diversity of an item set reasonably and easy to learn. A well-known approach proposed is to model diversity as the *average dissimilarity* of all pairs of items in the set [7]. Recently, a new idea has been proposed that the diversity of an item set is associated with the latent factors of items in it [6]. Inspired by this idea, we define v_i as the latent factors of item i for item-item similarity. For any pair of item i and item j, the more similar v_i and v_j are, the greater the inner product of v_i and v_j is. Besides, the categories of items are one kind of important and explicit information related to the diversity of an item set. Let $|C_i|$ be the number of categories item i contains. We also integrate it into SDB to normalize the latent factors, since an item often contains more than one category. With the notations above, our $b_d(S_k)$ is formalized as:

$$b_d(S_k) = -\sum_{i \in S_k} \sum_{j \in S_k, j \neq i} \frac{1}{|C_i||C_j|} v_i^{\mathrm{T}} v_j \tag{6}$$

We also define p_c as the latent factors of category c and w_i as the latent factors of item i for category-item relation, which are estimated by the logistic regression model below.

$$\min_{p_c, w_i} \sum_{i \in I} \sum_{c \in C_i} (1 - \frac{1}{1 + e^{-p_c^{\mathrm{T}} w_i}})^2 \tag{7}$$

The result of this model is that items with common categories have similar w_i, thus we introduce a L2-norm regularizing term $||v_i - w_i||^2$ to help learn SDB.

4. EXPERIMENTAL EVALUATION

In this section, we present a series of experiments to evaluate our proposed set-oriented personalized ranking model.

4.1 Experimental Setup

4.1.1 Dataset

We use the publicly available dataset MovieLen-1M in our experimental evaluation. In our work, we regard items with ratings larger than or equal to 4 as relevant. We randomly split relevant/irrelevant items of each user into the training set and the test set with the ratio of 4:1. Then D_S is constructed on the training set as described in section 3.1. During the training, we define the relevance score r_s of a given item set S as the number of relevant items in it, while the diversity score d_s of set S is measured by $\alpha NDCG_h$@N in (9).

4.1.2 Evaluation Metrics

In our experiments, we adopt the *precision*@N to evaluate the relevance of recommendation sets. The common metric *precision*@N is defined as:

$$precision@N = \frac{H_u}{N} \tag{8}$$

Where H_u is the number of relevant items in a recommendation set for user u. The diversity of recommendation sets can be evaluated by the metric $\alpha NDCG$@N [6]. We use the *highest* $\alpha NDCG@N$ ($\alpha NDCG_h$@N) and the *average* $\alpha NDCG@N$ ($\alpha NDCG_a$@N) instead, since the items in the set have many potential rankings. Let $L_u := \{l_1, l_2, \ldots, l_{N!}\}$ be the set of all potential inner rankings in a given item set of size N for user u, we formulate $\alpha NDCG_h$@N and $\alpha NDCG_a$@N as:

$$\alpha NDCG_h@N = \max_{l_p \in L_u} \alpha NDCG@N \tag{9}$$

$$\alpha NDCG_a@N = \frac{1}{|L_u|} \sum_{l_p \in L_u} \alpha NDCG@N \tag{10}$$

During training, we utilize $\alpha NDCG_h$ to measure the diversity of item sets. We obtain an approximate value of it with greedy algorithm in O(N^2), since the time complexity of calculating exact $\alpha NDCG_h$@N is O($N|L_u|$). While in the test, we utilize $\alpha NDCG_a$@N to evaluate the *statistical* performance of the recommendation sets of size N. For each user u, the recommendation set is approximate optimal recommendation set constructed by greedy selection strategy in O($a_u N^2$), where a_u is the amount of items in the test set for user u.

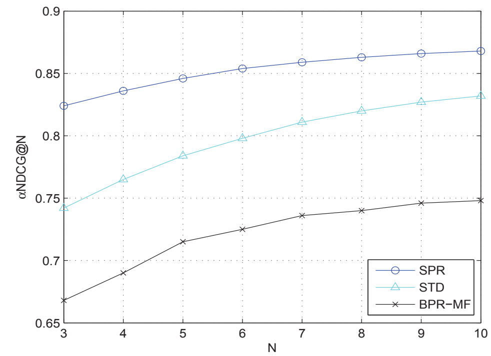

Figure 1: $\alpha NDCG$@N of SPR, STD and BPR-MF on MovieLen-1M

4.1.3 Compared Models

In this paper, we compare our model with several others [5, 9], which are called BPR-MF (Bayesian Personalized Ranking based on Matrix Factorization) and STD (Sequential Topic Diversification) separately in this paper. We refer to the parameters recommended for BPR-MF in [5] and STD in [9].

4.2 Results and Discussion

4.2.1 Performance on Diversity

Figure 1 compares the performance of our model with baselines in terms of diversity on MovieLen-1M. BPR-MF gives the worst performance on diversity, since it only focuses on providing the most relevant results for users without taking the diversity into account. STD gives higher performance than BPR-MF through re-ranking the prediction results based on topic(category) diversification. Our proposed SPR consistently outperforms BPR-MF and STD regarding $\alpha NDCG$@N at any value of $N \in [3, 10]$. Note that for our model, we use $\alpha NDCG_a$@N to evaluate the *statistical* performance of the recommendation sets. Another observation is that our model achieves significant improvement when the size N of recommendation set is small. This result empirically shows that our proposed model is suitable to diversify the top-N recommendation results for a small value of N.

4.2.2 Diversity vs. Relevance

In this section, we conduct experiments in the scenario of top-3 recommendation to explore the relation between relevance and diversity under various settings of λ in (2). Experimental results are shown in Figure 2. In our set-oriented CF model, λ is a constant to balance relevance and diversity for recommendation. We can observe that as the value of λ grows, $\alpha NDCG_a$@3 gradually increases while *precision*@3 decreases. This is consistent to the common intuition that the improvement of diversity weakens the relevance of results. By adjusting λ carefully, a reasonable balance of recommendation relevance and diversity can be achieved.

Table 2: Example top-4 recommendation results for users with *Focused* interests

Type	User with *Focused* interest (userID: 4651)	
ID	**Movies user has seen**	**Categories**
1147	When We Were Kings (4)	7
1192	Paris Is Burning (4)	7
1856	Kurt & Courtney (5)	7, 12
2129	The Saltmen of Tibet (5)	7
ID	**Recommendation Set**	**Categories**
116	Anne Frank Remembered (4)	7
134	Sonic Outlaws (4)	7
162	Crumb (4)	7
2330	Hands on a Hard Body (4)	7

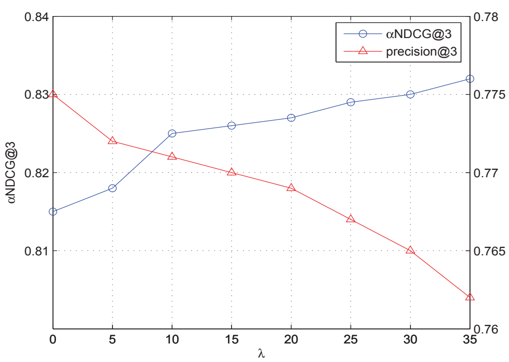

Figure 2: Diversity vs. Relevance

4.2.3 *Personalized Diversity*

In this section, we demonstrate that recommendation results given by our model achieves good personalized diversity. In Movielen-1M, we regard users who are related to 15 or more categories as "Broad" users and users who are related to 10 or less categories as "Focused" users. We also use numerical id (1-18) to represent different categories in Movielen-1M, such as 1 presents "Action", 7 presents "Documentary" and 17 presents "War".

We conduct experiments for top-4 recommendation. The results are shown in Table 2 and Table 3. In the two tables, movie ratings given by a user are presented in the brackets after movie names. Our model recommends movie sets with category ranges corresponding to different tastes of users. Table 2 shows a "focused" user(id: 4651) case. This user prefers documentary movies suggested by his rating history. Our recommendation set for this user focuses on the same movie type. The relevance of this set is very high with all the movie ratings above 4. Table 3 illustrates a case of the user(id: 406) who has broad interests. The recommendation set for this user covers broad types to meet his taste and also has good performance on relevance. Note that though "Alien" got a rating score 3, it is regarded as a diverse one containing 4 categories related to this user.

We also explore novelty of the recommendation sets for "Broad" users in top-3 recommendation. For this purpose, we consider the number of unique categories that each movie contains in a recommendation set. According to our statistical results, the expectation and variance of this value is 2.3 and 1.3 respectively. This observation means that each movie contains at least one different category from all the others in a recommendation set. This result further supports that our model has good performance in terms of personalized diversity.

Table 3: Example top-4 recommendation results for users with *Broad* interests

Type	User with *Broad* interest (userID: 406)	
ID	**Movies user has seen**	**Categories**
110	Braveheart (4)	1, 8, 17
589	Terminator 2: Judgment Day (4)	1, 15, 16
1183	The English Patient (4)	8, 14, 17
3175	Galaxy Quest (4)	2, 5, 15
ID	**Recommendation Set**	**Categories**
1097	E.T. (4)	4, 8, 9, 15
1210	Star Wars: Episode VI (4)	1, 2, 14, 15, 17
1214	Alien (3)	1, 11, 15, 16
2000	Lethal Weapon (4)	1, 5, 6, 8

5. CONCLUSION

In this paper, we propose a set-oriented personalized ranking model for diversified top-N recommendation. Our experimental results show that our model outperforms traditional models in terms of personalized diversity and maintains good performance on relevance prediction. In future work, we plan to provide a more thoughtful analysis for the potential relation between relevance and diversity in personalized top-N recommendation according to our experimental results.

6. REFERENCES

[1] A. P. Bradley. The use of the area under the roc curve in the evaluation of machine learning algorithms. *Pattern recognition*, 30(7):1145–1159, 1997.

[2] D. Goldberg, D. Nichols, B. M. Oki, and D. Terry. Using collaborative filtering to weave an information tapestry. *Communications of the ACM*, 35(12):61–70, 1992.

[3] N. Hurley and M. Zhang. Novelty and diversity in top-n recommendation–analysis and evaluation. *ACM Transactions on Internet Technology (TOIT)*, 10(4):14, 2011.

[4] Y. Koren. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 426–434. ACM, 2008.

[5] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme. Bpr: Bayesian personalized ranking from implicit feedback. In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*, pages 452–461. AUAI Press, 2009.

[6] Y. Shi, X. Zhao, J. Wang, M. Larson, and A. Hanjalic. Adaptive diversification of recommendation results via latent factor portfolio. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*, pages 175–184. ACM, 2012.

[7] B. Smyth and P. McClave. Similarity vs. diversity. In *Proceedings of the International Conference on Case-Based Reasoning and Development (ICCBR)*, volume 2080, pages 347–361. Springer, 2001.

[8] S. Vargas, P. Castells, and D. Vallet. Intent-oriented diversity in recommender systems. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in information retrieval*, pages 1211–1212. ACM, 2011.

[9] C. Ziegler, S. McNee, J. Konstan, and G. Lausen. Improving recommendation lists through topic diversification. In *Proceedings of the 14th international conference on World Wide Web*, pages 22–32. ACM, 2005.