Improving Topic Diversity in Recommendation Lists: Marginally or Proportionally?

Xiaolu Xing, Chaofeng Sha, and Junyu Niu

Shanghai Key Laboratory of Intelligent Information Processing,
School of Computer Science, Fudan University,
Shanghai, China
{14212010021,cfsha,jyniu}@fudan.edu.cn

Abstract. Diversifying the recommendation lists in recommendation systems could potentially satisfy user's needs. Most diversification techniques are designed to recommend the top-k relevant and diverse items, which take the coverage of the user preferences into account. The relevance scores are usually estimated by methods such as latent matrix factorization. While in this paper, we model the users' interests with the topic distributions on the rated items. And then we investigate how to improve the topic diversification within the recommendation lists. We first estimate the topic distributions of users and items through training Latent Dirichlet Allocation (LDA) on the rating set. After that we propose two topic diversification methods based on submodular function maximization and proportionality respectively. Experimental results on MovieLens and FilmTrust datasets demonstrate that our approach outperforms state-of-the-art techniques in terms of distributional diversity.

Keywords: Recommender system, Diversity, LDA

1 Introduction

Most traditional recommendation systems usually recommend items with high predicted scores to users by using some standard recommendation algorithms. However, many recent studies have shown theoretically and empirically that it is more beneficial to take the diversity within the recommendation lists into account as well [1, 7, 16], particularly when different users with diverse and ambiguous interests. Recommending a diverse set of the most relevant items is more likely to satisfy all potential needs of a given user. So finding a top-k itemset 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 lists, a number of diversification frameworks [1, 7, 16, 12, 9] have been proposed, which optimize the top-k items collectively in terms of both relevance and diversity. Most of these former works, e.g., Entropy Regularized [16], define some objective functions to balance between maximizing relevance and maximizing diversity. These methods estimate the relevance score through methods such as Latent Matrix Factorization [13]. However, we should be aware that most of the fundamental methods like LMF are proposed to achieve predictive ratings, which are not so suitable for getting the top-k lists that meet all of the potential needs of users.

Inspired by search result diversification methods from information retrieval field [3, 14], we also propose a two-stage diversification framework. In the first stage, we estimate the topic distributions of users and items through training LDA on rating set as we mentioned above. And in the second stage, we employ two methods to "cover" the topic distributions of users and items. In the second method, we use the election-based approach proposed in [14] to diversify recommendation lists proportionally, which has advantages on exploiting the two topic distributions.

The contributions of the paper can be summarized as follows: 1. We model each user's rating set as a document and train LDA on the "document" set to get the topic distributions, which are incorporated into the proposed objective functions. 2. We propose two diversification methods based on submodular function maximization and proportionality respectively. 3. We conduct extensive experiments on MovieLens dataset and FilmTrust dataset, which include performance comparisons with state-of-the-art methods. The experimental results demonstrate the effectiveness of our proposed solutions.

2 Related Work

Our work is related to the search result diversification in the information retrieval field and item diversification in the recommendation systems, and also related to the submodular function maximization. We give a brief review of those works in this section.

In the diversification based on item similarity, There is a trade-off between the utility and diversity of items. In the most popular framework MMR [4], the item i is greedily or marginally added into the result set S through maximize $(1-\lambda)w_i - \lambda \max_{j \in S} sim(i,j)$. The conditional differential entropy $h(r_S|r_\Omega, V)$ is employed in [16] as a diversity regularizer to the utility function $\sum_{i \in S} p_u^T q_i$. The adaptive attribute-based diversification model proposed in [10] follows the MMR framework. When computing similarity, they consider the diversification attribute of each user with respect to different item attributes. The propensity towards diversity is measured by entropy for each user. In [15], the authors propose a general framework which tradeoffs the utility of the items, the coverage of the user preferences and the diversity between items.

In [8], the diversity of recommendation lists is improved by maximizing the explicit topic attribute differences among items. In our first method, we follow the framework of IA-Select [3] which maximizes a probabilistic coverage function defined through the distribution of explicit topics. A probabilistic framework xQuAD (eXplicit Query Aspect Diversification) is proposed in [18] for Web search result diversification, which explicitly accounts for the various aspects of an under-specified query. The ranking algorithm is designed to satisfy the uncovered aspect by the selected documents. The explicit relevance model is also studied in [17]. In the diversity-weighted utility maximization (DUM) algorithm proposed in [9], the item is greedily added into to the recommendation list which try to maximize the objective function $\sum_{i=1}^{k}[f(A_i) - f(A_{i-1})]w_i$.

Our second diversification method follows the election-based approach proposed in [14] which diversifies the search result by proportionality. For each position in the recommendation list, We iteratively determine the topic that best maintains the overall proportionality, and then select the best item on this topic for this position. Therefore, proportional topics in the recommendation list could satisfy the preferences of the user.

3 Proposed Methods

In this section, we introduce a two-stage diversification framework. In the first stage, we estimate the topic distributions through training LDA on the transformed rating set. Then we propose two diversification methods based on the inferred distributions.

3.1 Problem Statement

Before delving into our methods, we introduce notations used in the paper. The set of users and items are denoted as U and V. The rating set is denoted as R. The set of topics we infer from LDA is denoted as T. The notation u, i, and t denote an user, an item, and a topic respectively. The diversification problem in this paper is to recommend an itemset S with size constraint $|S| = k$ to user u with diverse topics.

3.2 Topic Distribution Estimation

To improve the topic diversity within the recommendation lists, we should estimate the topic distributions of users and items. In our framework, we retort to Latent Dirichlet Allocation. We treat each user as a document and the items rated by users as words in documents. In our method, we transform the rating values of items to the number of occurrences of the word, i.e., the tuple $t_{u,i} = (u, i, r_{u,i})$ will be transformed into $\hat{t}_{u,i} = (u, i, i, i...)$ where the number of occurrence of i is $r_{u,i}$.

After the transformation, we use Latent Dirichlet Allocation [6] to estimate two distributions: user-topic distributions and item-topic distributions, which are denoted as θ and φ. Let θ_u denotes the topical distribution of the user u and $\theta_{u,t}$ denotes the probability of that the user u chooses the topic t. Similarly, let φ_t denotes the distribution of topic t over the set of items and $\varphi_{t,i}$ denotes the probability of that the item i appears in the topic t. We employ symmetric Dirichlet priors $Dir(\alpha)$ and $Dir(\beta)$ with the hyper-parameters of α and β on θ and φ respectively. Then the user preferences can be computed by using θ and φ as follows:

$$P_{u,i} = \sum_{t \in T} \theta_{u,t} \varphi_{t,i}, \tag{1}$$

We treat $P_{u,i}$ as the propensity of u selecting i. In other words, it is treated as user u's preferences.

Then we can recommend the top-k unrated items for u by reordering u's preferences $[P_{u,i} \; for \; i \in V]$. Actually this is one of baseline method in our experiment. For ease of use, we also introduce the item-topic distribution ϕ as a reverse of φ, then $\phi_{i,t}$ means the probability that item i covers topic t.

3.3 Diversify Marginally

Inspired by IA-Select [3], we recommend an itemset $S(|S| = k)$ to user u which maximizes the following objective function:

$$f(S) = R(S) + \lambda \sum_{t \in T} \theta_{u,t} \left(1 - \prod_{i \in S} (1 - \phi_{i,t}) \right), \tag{2}$$

Algorithm 1 `LDA-Greed`

Input: The itemset V and an integer k
Output: Subset of items $S \subseteq V$ with $|S| = k$

$S \leftarrow \{\arg\max_{i \in V} P_{u,i} + \lambda \sum_{t \in T} \theta_{u,t}\phi_{i,t}\}$;
while $|S| < k$ **do**
  $i^* \leftarrow \arg\max_{i \in V-S} P_{u,i} + \lambda \sum_{t \in T} \theta_{u,t}\phi_{i,t} \prod_{i' \in S}(1 - \phi_{i',t})$
  $S \leftarrow S \cup \{i^*\}$
  $V \leftarrow V - \{i^*\}$
end while
return S

where $R(S) = \sum_{i \in S} P_{u,i}$.

It is easy to verify that this problem is NP-hard (reduction from the Set Cover problem). Fortunately, thanks to the well-known result of [2], we can resort to the submodularity of the probabilistic coverage function. The submodularity of the objective function provides a theoretical approximation guarantee of factor $1 - \frac{1}{e}$ for the greedy search algorithm. Formally,

Theorem 1. Let f be a submodular, monotone set function and $f(\emptyset) = 0$. Then, the greedy search algorithm finds a set S ($|S| = k$) such that $f(S) \geq \left(1 - \frac{1}{e}\right) \max_{S':|S'| \leq k} f(S')$.

Theorem 2. The function defined in (2) is submodular and monotone.

Now we present the greedy algorithm outline in Algorithm 1, where we diversify the recommendation lists marginally.

3.4 Diversify Proportionally

In this section, we present a method to diversify items proportionally, inspired by the election-based approach which aims to diversify search result[14]. The main idea behind the method is to find a representative itemset S from the unrating items, which has the same topic proportionality to the topic distribution inferred for user u, i.e., $\{s_t\}_{t \in T} \propto \theta_u$, where $\{s_t\}_{t \in T}$ is the topic distribution of items contained in S, and s_t is calculated as follows: $s_t = \sum\limits_{i \in S} \frac{\phi_{i,t}}{\sum\limits_{t' \in T} \phi_{i,t'}}$.

To achieve this proportionality, when we add an item into S, we first choose a topic which has the most lack of proportionality in S. We use $quotient[t]$ to denote the proportional lack of topic t in S which is calculated as $quotient[t] = \frac{\theta_{u,t}}{2s_t+1}$.

The topic t^* is chosen as follows: $t^* \leftarrow argmax_{t \in T}\ quotient[t]$. Then we choose the optimal item with respect to topic t^*. The direct way is to select the item i with the highest $\phi_{i,t*}$. However, adding items into S not only affects the concerned topic t^*, but also affects several other topics covered by these items. So we introduce the parameter α to balance these effects: $i^* \leftarrow argmax_{i \in V-S} \quad \alpha \cdot quotient[t^*] \cdot \phi_{i,t^*} + (1-\alpha) \cdot \sum\limits_{t \neq t^*} quotient[t] \cdot \phi_{i,t}$. The adopted diversification by proportionality algorithm is presented in Algorithm 2.

Algorithm 2 LDA-PE

Input: The itemset V and an integer k
Output: Subset of items $S \subseteq V$ with $|S| = k$

$s_t \leftarrow 0, \forall t \in T$
while $|S| < k$ **do**
  for $t \in T$ **do**
    $quotient[t] = \frac{\theta_{u,t}}{2s_t+1}$,
  end for
  $t^* \leftarrow argmax_{t \in T}\ quotient[t]$.
  $i^* \leftarrow argmax_{i \in V-S} \quad \alpha \cdot quotient[t^*] \cdot \phi_{i,t^*} + (1-\alpha) \cdot \sum\limits_{t \neq t^*} quotient[t] \cdot \phi_{i,t}$.
  $S \leftarrow S \cup \{i^*\}$
  $V \leftarrow V - \{i^*\}$
  for $t \in T$ **do**
    $s_t \leftarrow s_t + \frac{\phi_{i^*,t}}{\sum\limits_{t' \in T} \phi_{i^*,t'}}$
  end for
end while
return S

4 Experiments

In this section, we verify the performances of the proposed diversification frameworks on two data sets.

4.1 Datasets and Comparison Methods

The experiments are evaluated on the publicly accessible datasets: MovieLens(1M) and FilmTrust. The details of two data sets are shown in Table 1.

Table 1. Statistics of Data Sets

Dataset	Users	Items	Ratings	Density
MovieLens	6040	3952	1000209	4.47%
FilmTrust	1508	2071	35497	1.14%

We evaluate the LDA-PE and LDA-Greed methods proposed above in the experiments, and compare them with the following methods: SVD, LDA, MMR, DUM [9], and PMF++(PMF+α+β) [15]. As LDA cannot directly generate a top-k recommendation itemset, we rank the items by calculating $P_{u,i} = \sum\limits_{t \in T} \theta_{u,t} \varphi_{t,i}$. Similarly we get the ranked lists by sorting with the ratings predicted by SVD.

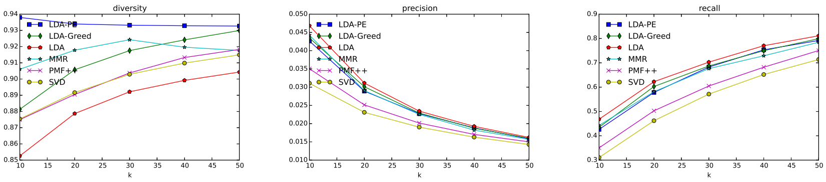

Fig. 1. Results on MovieLens

4.2 Evaluation Metrics

Diversity. The diversity is measured based on the item similarities in the recommendation lists. A recommendation list is more diverse when it consists of dissimilar items. The diversity is defined as $Diversity = 1 - \frac{\sum_{i_1, i_2 \in S, i_1 \neq i_2} sim(i_1, i_2)}{k(k-1)/2}$, where $sim(i_1, i_2)$ is the cosine similarity between i_1 and i_2, and is generated from the item-topic distribution ϕ in LDA.

Precision and Recall. We follow the evaluation protocol proposed in [19].

4.3 Experiment Design

Since DUM can only recommend a small number of items(less than 10), we first conduct other methods in MovieLens and FilmTrust datasets with $k \in \{10, 20, 30, 40, 50\}$, and then conduct DUM with LDA-PE and LDA-Greed in just MovieLens dataset with $k \in \{2, 4, 6, 8, 10\}$. All the methods take use of the "hit" protocol mentioned above. All the parameters are tuned manually. As LDA-PE, LDA-Greed are all based on LDA method, we run the LDA method first and make the further study on the results of LDA method. In addition, we also test the topic dimensionality d_T in LDA by studying the performances of LDA-Greed and LDA-PE on Movielens with $d_T \in \{25, 50, 75, 100\}$,

4.4 Results

For the parameter setting, we use $\alpha = 50.0/d_T$, $\beta = 0.01$ in LDA, and $\lambda = 0.02$, $\gamma = 0.005$ in LMF. These two methods are the basis of other methods. In addition, we set $\lambda = 0.5$ in LDA-PE, and $\lambda = 0.8$ in LDA-Greed to get the best performance. Topic dimensionality in LDA we set for MovieLens is $d_T = 50$, which we will explain in next subsection, and similarly we set $d_T = 25$ for FilmTrust.

Figure 2 shows the performances of our methods and the baseline methods in terms of diversity, precision and recall metrics with $k \in \{10, 20, 30, 40, 50\}$ on MovieLens dataset. As we can see, the diversities of all the methods except MMR increase as k grows, which means that for these methods, the larger recommended lists are more likely to have high diversity to satisfy all potential needs of users. Compared to other methods, LDA-PE has the best diversity. And LDA-Greed also has good performance, although its diversity is lower than MMR when $k <= 30$, but after that is higher than MMR. In particular, all other methods show an upward trend, except that LDA-PE's diversity curve remains stable declining, indicating that LDA-PE is more suitable for

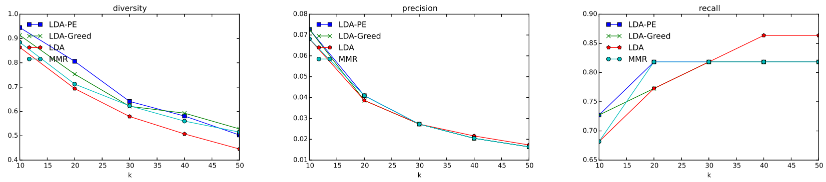

Fig. 2. Results on FilmTrust

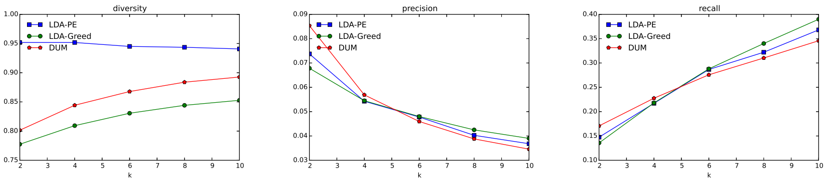

Fig. 3. Compare to DUM

low-capacity lists than others. LDA has the best precision and recall, because it is designed not to predict the ratings of users but to capture the preferences of users, and our choice of precision and recall metrics just meets this feature. The precision and recall of LDA-PE and LDA-Greed are below LDA and MMR for the penalize of diversity promotion.

Figure 3 shows the performances of methods on FilmTrust dataset. Due to the sparsity of FilmTrust dataset and the atypical strategy of calculating precision and recall in our approach, SVD and PMF++ do not have good precision and recall even though they have good Root Mean Square Error(RMSE)(below 1), so we just present the results of other methods here. As we can see, the performance of the diversity of all methods is rapidly deteriorating when k grows. We find that it's because the average number of users' ratings in FilmTrust is too small (about 20), so when k is too high, it's hard to get good results. Similar to Figure 2, LDA-PE also has the best performance in terms of diversity, and LDA-Greed is slightly lower than LDA-PE. The precision and recall of these methods are broadly similar.

Figure 4 shows the comparison between DUM and our two proposed methods on the MovieLens dataset with $k \in \{2, 4, 6, 8, 10\}$. We can see that the diversity of LDA-PE shows a big advantage at this small range of k, while LDA-Greed has the worst performance, which indicates that LDA-Greed is not so suitable for a small k. As for precision and recall, DUM is the best when k is less than 6, but LDA-PE and LDA-Greed perform better later.

5 Conclusion

In this paper, we take a further adoption of LDA method to the top-k recommendation, which can better reflect user preferences. We propose two diversification methods based on margin and proportion respectively, and make a comparison between the two methods. Experimental results on the MovieLens and FilmTrust datasets demonstrate that our methods is superior to the most advanced methods in terms of distributional diversity, while the proportion-based method has better performance.

References

1. Mi Zhang, Neil Hurley: Avoiding monotony: improving the diversity of recommendation lists. Lausanne, Switzerland, October 23-25 (2008)
2. G.L. Nemhauser, L.A.Wolsey, M.L. Fisher: An analysis of approximations for maximizing submodular set functions - I. Mathematical Programming. Mathematical Programming, vol. 14, pp. 265–294 (1978)
3. Rakesh Agrawal and Sreenivas Gollapudi and Alan Halverson and Samuel Ieong: Diversifying Search Results. WSDM 2009, pp. 5–14 (2009)
4. Jaime G. Carbonell, Jade Goldstein: The Use of MMR, Diversity-Based Reranking for Reordering Documents and Producing Summaries. SIGIR 1998, Melbourne, Australia (1998)
5. Younghoon Kim, Kyuseok Shim: TWILITE: A recommendation system for Twitter using a probabilistic model based on latent Dirichlet allocation. Inf. Syst., vol. 42, pp. 59-77 (2014)
6. David M. Blei, Andrew Y. Ng, Michael I. Jordan: Latent Dirichlet Allocation. Journal of Machine Learning Research, vol. 3, pp. 993-1022 (2003)
7. T. Zhou, Z. Kuscsik, J-G. Liu, M. Medo, J. R. Wakeling, Y-C. Zhang: Solving the apparent diversity-accuracy dilemma of recommender systems. PNAS, vol. 107, 2010.
8. C. Ziegler, S. McNee, J. Konstan, G. Lausen: Improving Recommendation Lists Through Topic Diversification. WWW, pp. 22-32 (2005)
9. A. Ashkan, B. Kveton, S. Berkovsky, Z. Wen: Optimal Greedy Diversity for Recommendation. IJCAI 2015, pp. 1742-1748.
10. Tommaso Di Noia, Vito Claudio Ostuni, Jessica Rosati, Paolo Tomeo, Eugenio Di Sciascio: An analysis of users' propensity toward diversity in recommendations. RecSys'14.
11. Le Wu, Qi Liu, Enhong Chen, Nicholas Jing Yuan, Guangming Guo, Xing Xie: Relevance Meets Coverage: A Unified Framework to Generate Diversified Recommendations. ACM TIST, vol. 7, pp. 39 (2016)
12. A. Ashkan, B. Kveton, S. Berkovsky, Z. Wen: Diversified utility maximization for recommendations. RecSys Poster Proceedings (2014)
13. Y. Koren, R. Bell: Advances in Collaborative Filtering. Recommender Systems Handbook, pp. 145–186 (2011)
14. Van Dang, W. Bruce Croft: Diversity by proportionality: an election-based approach to search result diversification. SIGIR '12, Portland, OR, USA, August 12-16 (2012)
15. Chaofeng Sha, Xiaowei Wu, and Junyu Niu. A Framework for Recommending Relevant and Diverse Items. IJCAI 2016, New York, NY, USA, 9-15 July (2016)
16. Lijing Qin and Xiaoyan Zhu. Promoting Diversity in Recommendation by Entropy Regularizer. IJCAI, pp. 2698–2704 (2013)
17. Saul Vargas, Pablo Castells, David Vallet: Explicit relevance models in intent-oriented information retrieval diversification. SIGIR '12, Portland, OR, USA, August 12-16 (2012)
18. Rodrygo L. T. Santos, Craig Macdonald, Iadh Ounis: Exploiting query reformulations for web search result diversification. WWW 2010.
19. Nicola Barbieri, Giuseppe Manco: An Analysis of Probabilistic Methods for Top-N Recommendation in Collaborative Filtering. ECML PKDD 2011.