Cross-Region Collaborative Filtering for New Point-of-Interest Recommendation

Ning Zheng
School of Software
Tsinghua University
Beijing, China
thu.zhengning@gmail.com

Xiaoming Jin
School of Software
Tsinghua University
Beijing, China
xmjin@tsinghua.edu.cn

Lianghao Li
Department of Computer Science and Engineering
Hong Kong University of Science and Technology
lianghaoli@yahoo.com

ABSTRACT

With the rapid growth of location-based social networks (LBSNs), Point-of-Interest (POI) recommendation is in increasingly higher demand these years. In this paper, our aim is to recommend new POIs to a user in regions where he has rarely been before. Different from the classical memory-based recommendation algorithms using user rating data to compute similarity between users or items to make recommendation, we propose a cross-region collaborative filtering method based on hidden topics mined from user check-in records to recommend new POIs. Experimental results on a real-world LBSNs dataset show that our method consistently outperforms naive CF method.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: Information filtering

Keywords

cross-region, collaborative filtering, location based social network

1. INTRODUCTION

Location-based social networks (LBSNs), such as Foursquare, Gowalla and Facebook Places, are a kind of work that allow users to check-in and share POIs with their friends. With the prevalence of location-based social networks, plenty of locations are being explored rapidly. For such applications, it is crucial to accurately recommend interesting POIs to users.

Though it is believed that to a certain degree the movement and mobility patterns of human beings are full of uncertainty and variability, experiments show that people always visit locations in one or two certain regions. Figure 1 shows the users' check-ins in Queens, Brooklyn, Bronx and Manhattan Districts of New York City. The first 4,824 locations are in Queens District. The next 4,267 are in Brooklyn. The following 3,201 are in Bronx. And the last 4,654 are in Manhattan. The density of sub-regions *Queens*, *Brooklyn*, *Bronx* and *Manhattan* in bottom-left to upper-right diagonal is much higher than the density of other regions in the figure, meaning that the users tend to visit a specific district and rarely visit locations in other districts at a certain time.

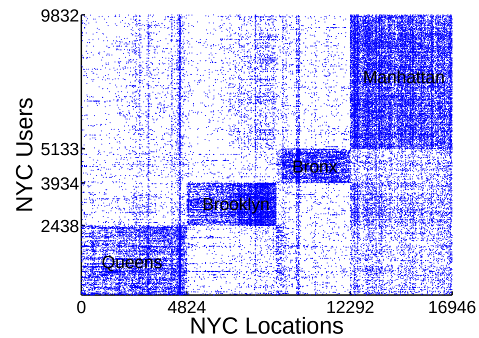

Figure 1: User check-in distribution in Queens, Brooklyn, Bronx and Manhattan districts of New York City.

One of the most widely used recommendation algorithms is memory-based collaborative filtering (CF), which uses user rating data to compute similarity between users or items for making recommendation. However, when recommending POIs in a new region for users, memory-based CF is hard to provide POI suggestions based on existing user-item rating matrix, for the like-minded users tend to visit the same one or two specific regions and rarely visit POIs in other regions. To tackle this problem, we propose the cross-region topic-based collaborative filtering (CRTCF) for new POIs recommendation.

2. OUR SOLUTION

In this section, we will first introduce some definitions. Then we will discuss how to mine hidden topics from user check-in records. Finally, we will present how to use the mined hidden topics to recommend new POIs to an user who rarely come to that region.

2.1 Definition

Let U, P and C respectively denote the user set, the POI set and the check-in set, which keep track of check-in activities in LBSNs. Let $c_{i,j} = k$ denote that user $u_i \in U$ check-ins at POI $p_j \in P$ for k times. Let F denote the friend set and f_i denote all friends of user u_i.

In this paper, we treat different districts of a city as different regions, each of which consists of a number of POIs. Among all the regions of one city, we regard the region where user u_i check-ins most as the *major region* of that user.

2.2 Cross-Region Collaborative Filtering

As we discussed before, the classical collaborative filtering algorithms may fail to recommend accurate new POIs out of major region to users. Suppose there are two users, u_i and u_j, having the same major region. Though their user similarity is high, it is difficult to recommend POIs of other regions to user u_j based on user u_i's check-in records, because both of them have few check-ins out of their major region. On the contrary, if the two users do not have the same major region, the user similarity between them is low, which limits the effectiveness of collaborative filtering.

Although regions may different in their locations and POIs, they may share some similar check-in patterns and POIs, such as restaurants, shopping malls and bars. We can regard these similar patterns in various regions as they are from the same topic of a latent space. Specifically, we mine the hidden topics from different regions and then use the mined topics as features to calculate the user similarity in classical collaborative filtering algorithm.

We mine the hidden topics from the user check-in record using Latent Dirichlet Allocation (LDA) [1], a mixed-membership model which has been successfully used to detect patterns and clusters in text. In our paper, we regard each user as a document and the set of POIs that a user has checked into as its words. Considering some users do not have many check-ins, in this paper we add their friends' check-ins to extend theirs document to a certain degree, since friends tend to have similar tastes, interests and life styles. Thus by using LDA, we can group POIs which tend to occur simultaneously into the same topic and then assigns to each user a topic distribution to represent how often they check into POIs in each topic. Specifically, for user u_i, we represent the topic distribution of u_i as $\vec{\theta^i} = \{\theta_1^i, \theta_2^i, ..., \theta_K^i\}$, where K is the number of topics predefined in LDA.

CRTCF makes recommendation of new POIs in a new region for an active user by first finding his friends who have similar topic distribution, namely similar pattern distribution and then taking a weighted combination of their check-in numbers. More formally let u_i be the active user and p_k be the POI of interest, which have not been visited by u_i before, out of his major region. Then the predicted rating of u_i to p_k, $r_{i,k}$ can be calculated as follow:

$$r_{i,k} = \frac{\sum_{j \in f_i} c_{j,k} * w_{i,j}}{\sum_{j \in f_i} w_{i,j}} \tag{1}$$

where $w_{i,j}$ is the cosine similarity between topic distribution vector of θ^i and θ^j. Then we recommend the POIs which have the highest predicted rating to the user.

3. EXPERIMENT

In this section, we used the Gowalla dataset [3] for experiments. The Gowalla data set are collected from public check-in data from February 2009 to October 2010. Our experiment focused on the POI recommendation in Queens, Brooklyn, Bronx and Manhattan Districts in New York City. There is a total number of 2,532 users, 15,192 undirected friendships and 88,464 check-ins. The experiment was conducted on each user separately. For each user, we chose POIs in his major region (district) as training data and the remaining POIs in other regions (districts) as test data. Not surprisingly, a high proportion of 80.29% of POIs is in the major region and the remaining 19.71% is in other regions.

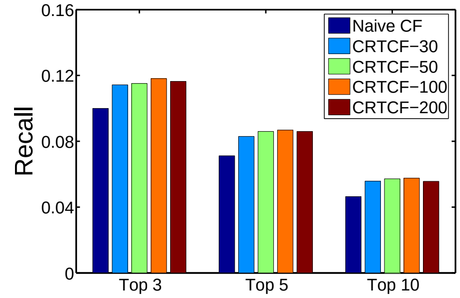

Figure 2: Recall of top-N recommendation for naive CF and CRTCF

And in our experiment, we only recommend POIs which is out of the major region. We adopt *Recall* as our evaluation criteria, which are measured by the number of hits within the top-N POIs that were recommended.

To test the effectiveness of our method, we compare our method with naive CF [2]. Figure 2 shows the experiment results for *Recall* of top-N (N = 3, 5, 10) recommendation of naive CF and our method under various number of hidden topics. As illustrated in Figure 2, CRTCF consistently outperforms naive CF in all experiment settings. For example, when using 100 topics, our method improves the recall of top-N recommendation by 21%. The experiment results verify the effectiveness of our method for cross-region POI recommendation.

4. CONCLUSION

In this work, we propose a novel probabilistic method to recommend new POIs to a user in regions where he/she has rarely been before. Different from the conventional recommendation algorithms which only uses user rating data to compute similarity between users or items, our method mines the shared check-in patterns for users from different regions and then utilizes the shared patterns to further explore more similar user across regions. By using the proposed cross-region topic-based collaborative filtering method, we effectively transfer knowledge across regions to recommend for an user in a new region. The experiment results on a real-world application dataset show that our method significantly improve the recommendation performance compared with the baselines.

5. ACKNOWLEDGMENT

The work was supported by National Natural Science Foundation of China (60973103, 90924003).

6. REFERENCES

[1] D. Blei, A. Ng, and M. Jordan. Latent dirichlet allocation. *the Journal of machine Learning research*, 3:993–1022, 2003.

[2] J. S. Breese, D. Heckerman, and C. Kadie. Empirical analysis of predictive algorithms for collaborative filtering. In *Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence*, pages 43–52. Morgan Kaufmann Publishers Inc., 1998.

[3] E. Cho, S. A. Myers, and J. Leskovec. Friendship and mobility: user movement in location-based social networks. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1082–1090. ACM, 2011.