PGRank: Personalized Geographical Ranking for Point-of-Interest Recommendation

Haochao Ying[1], Liang Chen[2], Yuwen Xiong[1], Jian Wu[1]
[1] Zhejiang University, Hangzhou, China
{haochaoying, orpine, wujian2000}@zju.edu.cn
[2] RMIT University, Melbourne, Australia
liang.chen@rmit.edu.au

ABSTRACT

Point-of-interest (POI) recommendation has become more and more important, since it could discover user behavior pattern and find interesting venues for them. To address this problem, we propose a rank-based method, PGRank, which integrates user geographical preference and latent preference into Bayesian personalized ranking framework. The experimental results on a real dataset show its effective.

Keywords

POI Recommendation; Geographical Preference; Rank

1. INTRODUCTION

With the rapid grows of location-based social network, users post their physical locations by checking-in points-of-interest (POIs), such as museum, restaurants, etc. A vast amount of check-in data creates personalized POI recommendation, which helps users discover new POIs and advertisers find targeted users. However, unlike traditional recommendation based on explicit feedback data (e.g., movie rating 1-5), the check-in data (e.g., the frequency of check-in) is implicit feedback.

Point-wise regression and pair-wise ranking preference algorithm are two major paths to solve implicit feedback recommendation. The point-wise regression algorithm supposes that users don't like all unobserved items and optimizes the absolute rating scores, while pair-wise ranking algorithm assumes that users' preference of observed items is stronger than unobserved items and then directly transfer the prediction to ranking. Recently, GeoMF [1] incorporates geographical information into weighted matrix factorization model, which is a point-wise regression model. However, the assumption that users don't like all unvisited POIs may be not reasonable.

In this paper, we propose a hybrid pair-wise POI recommendation approach, named personalized geographical ranking (PGRank), which integrates POI coordinates into Bayesian Personalized Ranking framework. Specifically, user latent preference and geographical preference, extracted from matrix factorization and cluster techniques, are utilized together to improve the POI recommendation accuracy. Comprehensive experiments are conducted based on foursquare dataset to evaluate our approach. The results provide good demonstration in the superiority over other competitors.

WWW'16 Companion, April 11–15, 2016, Montréal, Québec, Canada.
ACM 978-1-4503-4144-8/16/04.
DOI: http://dx.doi.org/10.1145/2872518.2889378 .

2. METHODOLOGY

Problem Definition. Suppose that frequent matrix $\mathbf{R} \in \mathbb{R}^{n \times m}$ represents check-in records of users, where n and m denote the size of users and POIs, respectively. Then given user u, our goal is to recommend new POIs which he/she never checked but will be like to visit in the future.

Personalized Geographical Ranking. Unlike recommendation task based on explicit rating, the check-in data represents implicit feedback of users, which means users prefer visited POIs than unvisited. Therefore, we can exploit pair-wise ranking model to generate top-N recommendation for each user. In this paper, we adapt Bayesian Personalized Ranking (BPR) criterion to our problem [3]. Otherwise, we intuitively assume that a POI preference by one user is proportional to the check-in frequency. Formally, if user i has visited more times in POI j than POI k, the likelihood of this preference is $p(f_{ij} \succeq f_{ik}) = I(x_{ij} > x_{ik})\sigma(f_{ij} - f_{ik})$ where x_{ij} is the frequency of check-in, $I(x)$ is the indicator function, $\sigma(x) = \frac{1}{1+e^{-x}}$, and f_{ij} denotes the ranking score of POI j for user i. Obviously, how to compute the ranking score function determines the performance of recommendation.

In this paper, we jointly integrate user latent preference as well as geographical preference to predict ranking score. First, we tackle the latent preference as traditional matrix factorization. Particularly, we suppose that there is latent factor vector $\mathbf{u}_i \in \mathbb{R}^{k \times n}$ and $\mathbf{v}_j \in \mathbb{R}^{k \times m}$ respectively associated with user i and POI j, where k means the number of latent factor. Therefore, the ranking score based on user latent preference is $f_{ij}^1 = \mathbf{u}_i^T \mathbf{v}_j$. Secondly, it has been proved that the context information, geographical coordinates, significantly affects users' choice of POIs. Furthermore, the geographical distribution of POIs by a user always exhibits multi-region phenomenon. The user behavior pattern may vary dramatically in different regions [4]. To capture this property, we cluster the geographical space into l regions. For simplify, k-means algorithm is used in the experiments. Then, the ranking score based on the geographical preference (i.e., region preference) is $f_{ij}^2 = r_{ic_j}$, where c_j means the region includes POI j. Thus, the final ranking score function is as follows:

$$f_{ij} = \mathbf{u}_i^T \mathbf{v}_j + r_{ic_j} \tag{1}$$

Then, the likelihood of all users is $p(\succeq |\Theta) = \prod_{i \in \mathcal{U}, j \in \mathcal{V}} p(f_{ij} \succeq f_{ik})$, where $\mathcal{U}$, $\mathcal{V}$ is the set of users and POIs, respectively. $\Theta = \{\mathbf{U}, \mathbf{V}, \mathbf{r}\}$ is the set of all parameters.

With the prior distribution of $p(\Theta)$, we can exploit maximum a posterior (MAP) to optimize Θ for the correct personalized ranking. Note that the prior distribution is a form of regularization which reduce the complexity of model and avoid overfitting (e.g., Gaussian prior distribution in this paper). Through the logarithm, the final objective function is as follows:

$$\begin{aligned}
\Theta &= \arg\max_{\Theta} \log \prod_{i \in \mathcal{U}, j \in \mathcal{V}} p(f_{ij} \succ f_{ik}) p(\Theta) \\
&= \arg\max_{\Theta} \sum_{i \in \mathcal{U}} \frac{1}{1 + e^{(f_{ij} - f_{ik})}} \\
&- \frac{\beta_1}{2} \Big(\sum_{i \in \mathcal{U}} ||\mathbf{u}_i||_F^2 + \sum_{j \in \mathcal{V}} ||\mathbf{v}_j||_F^2\Big) - \frac{\beta_2}{2} \sum_{i \in \mathcal{U}, j \in \mathcal{V}} ||r_{ic_j}||_F^2
\end{aligned} \tag{2}$$

where β_1, β_2 are fixed regularization parameters to balance the bias and variance of model.

Optimization. Analogous to BPR, we utilize sample technology to accelerate optimization process and Stochastic Gradient Descent to update Θ. Then we sort the ranking score function in Eq.1 and generate the top-k recommendation list for each user.

3. EXPERIMENTS

Dataset. To demonstrate the performance of the proposed model, we use the Foursquare check-in dataset during the period of Aug. 2010 and Jul. 2011 in Singapore [5]. The preliminary statistics shows that this dataset contains 2,321 users and 5,596 POIs with the sparsity of 98.51%. Each record comprises user ID, POI ID, POI coordinate and time. Similar to [5], we randomly mask off 62.5% of check-in as training set, 12.5% as tuning set to choose parameters, and the remaining 25% as testing data for each user.

Baseline Methods. In the experiments, we compare the following approachs: (1)PMF: probabilistical matrix factorization is a basic point-wise model [2]. (2)GeoMF: a weighted matrix factorization model with geographical information [1]. (3)BPR: a traditional pair-wise framework for recommendation [3]. (4)PGRank: Personalized Geographical Ranking is our proposed model described in Section 2.

Results. We employ two metrics, i.e., precision and recall, to evaluate the performance of different methods, which are widely utilized for POI recommendation. We set the learning rate as 0.01, max iteration as 300, $k = 200$ for our model. The number of regions is 150 and the hyper-parameters are $\beta_1 = 0.02$, $\beta_2 = 0.12$.

Figure 1 summarizes the performance of different approaches. We can observe that the methods (GeoMF, BPR and PGRank) are much better than PMF, both in terms of precision and recall. This indicates that traditional rating-based matrix factorization may not suit the scenario of POI recommendation. Because PMF is adept in explicit feedback data, while the check-in data is implicit feedback. It could also be found that rank-based method, BPR, consistently performs better than GeoMF, although BPR don't consider any context information in POI recommendation. Specifically, BPR

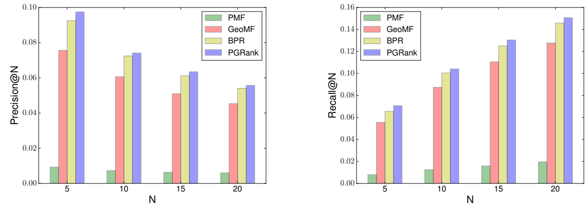

Figure 1: Performance Comparison on Foursquare

improves 20% and 16.3% in precision@20 and recall@20, respectively. The reason may be that directly optimizing ranking is more appropriate than optimizing the absolute scores in implicit feedback data.

As a whole, our proposed approach, PGRank, outperform all of the other methods. Compared with GeoMF which also considers geographical information, PGRank improves GeoMF by 28.8% and 27% in precision@5 and recall@5, respectively. Furthermore, our model improves BPR by 5.3% and 7.8% in precision@5 and recall@5,respectively. This indicates that it is effective to fuse user region preference into traditional BPR framework.

4. CONCLUSIONS AND FUTURE WORK

In this paper, we propose a pair-wise POI recommendation approach, PGRank, which extracts the user latent preference and geographical preference under Bayesian personalized ranking framework. The experiments demonstrate the effectiveness of our method. In the future work, we will incorporate more context information into our model, such as time stamp of check-in, categories of POIs, social relationship of users and so on.

5. ACKNOWLEDGMENTS

This research was partially supported by the Natural Science Foundation of China under grant of No. 61379119, Science and Technology Program of Zhejiang Province under grant of No. 2013C01073, the Open Project of Qihoo360 under grant of No. 15-124002-002, the Fundamental Research Funds for the Central Universities.

6. REFERENCES

[1] D. Lian, C. Zhao, X. Xie, G. Sun, E. Chen, and Y. Rui. Geomf: joint geographical modeling and matrix factorization for point-of-interest recommendation. In *SIGKDD*, pages 831–840. ACM, 2014.

[2] A. Mnih and R. Salakhutdinov. Probabilistic matrix factorization. In *NIPS*, pages 1257–1264, 2007.

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

[4] W. Wang, H. Yin, L. Chen, Y. Sun, S. Sadiq, and X. Zhou. Geo-sage: A geographical sparse additive generative model for spatial item recommendation. In *SIGKDD*, pages 1255–1264. ACM, 2015.

[5] Q. Yuan, G. Cong, Z. Ma, A. Sun, and N. M. Thalmann. Time-aware point-of-interest recommendation. In *SIGIR*, pages 363–372. ACM, 2013.