Collaborative Filtering Meets Next Check-in Location Prediction

Defu Lian
University of Science and Technology of China
liandefu@mail.ustc.edu.cn

Vincent W. Zheng
Advanced Digital Sciences Center, Singapore
vincent.zheng@adsc.com.sg

Xing Xie
Microsoft Research Asia, Beijing, China
xingx@microsoft.com

ABSTRACT

With the increasing popularity of Location-based Social Networks, a vast amount of location check-ins have been accumulated. Though location prediction in terms of check-ins has been recently studied, the phenomena that users often check in novel locations has not been addressed. To this end, in this paper, we leveraged collaborative filtering techniques for check-in location prediction and proposed a short- and long-term preference model. We extensively evaluated it on two large-scale check-in datasets from Gowalla and Dianping with 6M and 1M check-ins, respectively, and showed that the proposed model can outperform the competing baselines.

Keywords

Location Prediction; Collaborative Filtering; LBSNs

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval—*clustering, information filtering*

General Terms

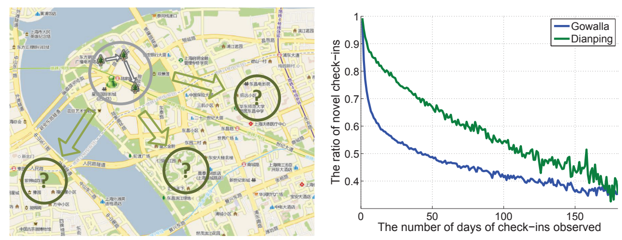
Algorithms, Design, Experimentation

1. INTRODUCTION

With the increasing popularity of Location-based Social Networks (LBSNs), a vast amount of location check-ins have been accumulated. In this paper, we are interested in predicting user's future check-in locations based on these data. In particular, we attempt to determine which Point Of Interest (POI), such as a clothing store or a western restaurant, a user will check in next. One of its typical scenarios is shown in Figure 1(a).

Though the next check-in location prediction problem has been recently studied [1, 3, 4], the phenomena that users often check in novel locations has not been addressed. According to our observations, shown in Figure 1(b), users checked in over 35% *novel POIs* each day on average even after half a year, where *novel POIs* are those POIs that users have not checked in before. The check-ins at *novel POIs* bring challenges to the prediction models which heavily depend on the feature of user's individual check-in frequency at POIs [1, 3, 4] since this feature is zero at *novel POIs*.

WWW 2013 Companion, May 13–17, 2013, Rio de Janeiro, Brazil.
ACM 978-1-4503-2038-2/13/05.



(a) (b)

Figure 1: (a)A typical scenario for next check-in location prediction. Given three successive check-ins(tree icons) of a user(head icon), we predict her next location.(b)The ratio of check-ins at *novel POIs* given the number of days of check-ins observed.

To deal with this problem, we leveraged collaborative filtering techniques to resort to the similar users' patterns. In particular, we proposed a factorizing approach, which we named as a short- and long-term preference model (SLoP), for dimension reduction and encoded users' check-in patterns in a low dimension latent space. In short-term preference model, we extended sequential collaborative filtering (SCF) [5] to not only model the transition between the POIs but also consider various features including dynamic user's preferences, spatio-temporal constraints. In addition, since users often checked in several POIs in a short time, similar to the successive check-ins in Figure 1(a), we considered these successive check-ins as groups instead of single check-in as basic units of SCF. However, in the short-term preference we simply modeled user's local preference that depended on contextual information such as time and previous check-in locations, thus we leveraged the long-term preference, which learned the personalized pairwise preference between POIs, for its supplement.

2. COLLABORATIVE FILTERING FOR LOCATION PREDICTION

We assume check-in history $\mathcal{G}_u$ of user u is represented as a sequence of check-in groups in chronological order, i.e. $G_u^{1:n} = \{G_u^1, \cdots, G_u^n\}$, where n is the total number of check-in groups. Then the next check-in location prediction problem is formalized as $Pr(i \in G_u^{n+1}|t, G_u^{1:n})$, that is the probability of POI i belonging to the next check-in group.

In short-term preference, since we only consider current check-in group and the time of next check-in as contextual information, i.e., $Pr(i \in G_u^{n+1}|t, G_u^{1:n}) = Pr(i \in G_u^{n+1}|t, G_u^n)$, it is defined as

$$\hat{Pr}(i \in G_u^{n+1}|t, G_u^n) \propto (p_u + w_{h(t)}) \cdot q_i + \frac{f(\delta d, \delta t)}{|G_n^u|} \sum_{k \in G_n^u} r_k \cdot q_i$$

where the probability is considered as some real value that can be factorized. p_u, $w_{h(t)} \in \mathbb{R}^F$ can denote user's and temporal preference on some intrinsic POI categories, respectively, and $q_i, r_k \in \mathbb{R}^F$ can denote the possibility of POI i belonging to the corresponding POI categories. Thus $(p_u + w_{h(t)}) \cdot q_i$ represents dynamic user's preference, which means user's preference is varied with time, where $h(t)$ maps time t to an hour of the week. And $\frac{1}{|G_n^u|} \sum_{k \in G_n^u} r_k \cdot q_i$ models the transition from previous check-in locations to next. $f(\delta d, \delta t) = I_{\{\delta t < \Delta T\}} e^{-\beta \delta d}$ places an spatio-temporal constraint on POIs' transition, where δt and δd is the time interval and distance between the next check-in group and the current one. This constraint indicates the larger influence of more adjacent check-in groups. For learning these parameters, we follow [5] to perform stochastic gradient descent on an objective function based on the pairwise preference between user's check-in POIs and non check-in POIs plus a Frobenius norm of parameters to avoid over-fitting.

In long-term preference, it learns the personalized pairwise preference between POIs without considering the influence of time and previous check-ins, i.e., $Pr(i \in G_{n+1}^u|t, G_{1:n}^u) = Pr(i \in G_{n+1}^u)$, it is represented as

$$\hat{Pr}(i \in G_{n+1}^u) \propto p_u \cdot q_i$$

where $p_u, q_i \in \mathbb{R}^F$ share similar meaning to that in the short-term preference. However we learn them by performing stochastic gradient descent on a different ranking objective function which considers not only the pairwise preference between user's check-in POIs and non check-in POIs but also the pairwise preference between the check-in POIs with different frequency.

Since our goal is to perform POIs ranking, we don't calculate their real probability but simply consider the short and long-term preference as two scores for POIs. Then we blend them in a linear way as our SLoP model to get final scores for POIs.

3. PERFORMANCE EVALUATION

We evaluated on two large-scale check-in datasets from Gowalla [2] and Dianping, with 6M check-ins at 1,280,969 POIs from 107,092 users and 1M check-ins at 150,094 POIs from 20,429 users, respectively. These check-ins were preprocessed by first filtering the users with fewer than 10 days of check-in history and then grouping successive check-ins in a short time. Then they were split into a training portion (80%) and a testing portion (20%). Since we aimed to perform POIs ranking, in order to evaluate the performance of ranking algorithms, we exploited Accuracy at position k (Acc@k). Acc@k was $\frac{1}{k}$ if the POI of a check-in was returned at $j \leq k$ position and 0 otherwise. Finally, we also evaluated the ranking performance on the check-ins at *novel POIs*, thus we reported Acc@k at both all check-ins and the check-ins at *novel POIs*. Due to the space limit, we only showed Acc@10.

We compared SLoP with the following baselines: 1) UMostFreq, which predicted next location as user's most frequented location; 2) UTMostFreq, similar to UMostFreq, but also

Table 1: Comparison with baselines.

Approaches	Acc@10_Dianping		Acc@10_Gowalla	
	novel	*all*	*novel*	*all*
UMostFreq	0.0000	0.2093	0.0000	0.3032
UTMostFreq	0.0004	0.1490	0.0007	0.2159
MostFreq	0.0108	0.0262	0.0394	0.0842
SHM	0.0000	0.2157	0.0000	0.3065
SLoP	0.0016	0.2199	0.0037	0.3120

considered the time (hour of week); 3) MostFreq, which predicted next location as the most frequented location. 4) SHM [3], which took into account both user's check-in frequency and transition between POIs. The comparing results were shown in Table 1. From this figure, we observed that: 1) UMostFreq performed better than UTMostFreq and MostFreq. Their reasons were that UTMostFreq encountered over-fitting due to the insufficiency of individual check-in history and that without distinguishing users lost significant individual patterns 2) SHM outperformed the above three baselines since it already exploited these information. 3) Although MostFreq showed a good performance on the check-ins at *novel POIs*, it didn't perform well on other check-ins. However, according to the performance gap between MostFreq and SLoP on the check-ins at *novel POIs*, we observed that there was still improving space to predict the check-ins at *novel POIs*. 4) SLoP outperformed all baselines. This was because SLoP not only considered those information but also leveraged collaborative filtering for the check-ins at *novel POIs*. However, the improvement over SHM was not large. The reason may lies in the following two reasons. First, sparse user-POI frequency matrix resulted in a low performance on the check-ins at *novel POIs* and thus only brought small improvement on all check-ins. Second, the fuse of short- and long-term preference deceased the performance on the check-ins at *novel POIs* since the long-term preference played a dominating role in SLoP.

4. CONCLUSIONS

In this paper, we studied the next check-in prediction problem and proposed a CF-based algorithm. By evaluating on two large-scale check-in datasets, our proposed model – SLoP outperformed four baselines.

5. REFERENCES

[1] J. Chang and E. Sun. Location3: How users share and respond to location-based data on social. In *Proc. of ICWSM'11*, 2011.

[2] E. Cho, S. Myers, and J. Leskovec. Friendship and mobility: user movement in location-based social networks. In *Proc. of KDD'11*, pages 1082–1090, 2011.

[3] H. Gao, J. Tang, and H. Liu. Exploring social-historical ties on location-based social networks. In *Proc. of ICWSM'12*, 2012.

[4] A. Noulas, S. Scellato, N. Lathia, and C. Mascolo. Mining user mobility features for next place prediction in location-based services. In *Proc. of ICDM'12*, pages 1038–1043. IEEE, 2012.

[5] S. Rendle, C. Freudenthaler, and L. Schmidt-Thieme. Factorizing personalized markov chains for next-basket recommendation. In *Proc. of WWW'10*, pages 811–820. ACM, 2010.