Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence

Fused Matrix Factorization with Geographical and Social Influence in Location-Based Social Networks

Chen Cheng[1], Haiqin Yang[1], Irwin King[2,1], Michael R. Lyu[1]

[1]Computer Science and Engineering
The Chinese University of Hong Kong
Shatin, N.T., Hong Kong
{ccheng, hqyang, king, lyu}@cse.cuhk.edu.hk

[2]AT&T Labs Research
525 Market Street
San Francisco, CA, USA
irwin@research.att.com

Abstract

Recently, location-based social networks (LBSNs), such as Gowalla, Foursquare, Facebook, and Brightkite, etc., have attracted millions of users to share their social friendship and their locations via check-ins. The available check-in information makes it possible to mine users' preference on locations and to provide favorite recommendations. Personalized Point-of-interest (POI) recommendation is a significant task in LBSNs since it can help targeted users explore their surroundings as well as help third-party developers to provide personalized services. To solve this task, matrix factorization is a promising tool due to its success in recommender systems. However, previously proposed matrix factorization (MF) methods do not explore geographical influence, e.g., multi-center check-in property, which yields suboptimal solutions for the recommendation. In this paper, to the best of our knowledge, we are the first to fuse MF with geographical and social influence for POI recommendation in LBSNs. We first capture the geographical influence via modeling the probability of a user's check-in on a location as a Multi-center Gaussian Model (MGM). Next, we include social information and fuse the geographical influence into a generalized matrix factorization framework. Our solution to POI recommendation is efficient and scales linearly with the number of observations. Finally, we conduct thorough experiments on a large-scale real-world LBSNs dataset and demonstrate that the fused matrix factorization framework with MGM utilizes the distance information sufficiently and outperforms other state-of-the-art methods significantly.

Introduction

Recently, with the rapid development of mobile devices and ubiquitous Internet access, location-based social services become prevalent. Online location-based social networks (LBSNs), such as Gowalla, Foursquare, Facebook, and Brightkite, etc., have attracted millions of users to share their social friendship, experience and tips of Point-of-interest (POI) via check-ins. These information embeds abundant hints of users' preference on locations. The information not only can be utilized to help a specific user to explore new places of the city, but also can help third-parties such as advertisers to provide specific advertisements for the recommended positions. Hence, personalized POI recommendation becomes a significant task in LBSNs.

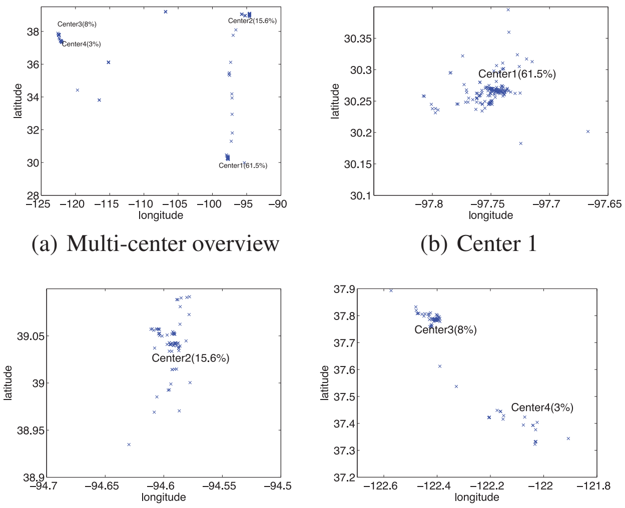
	l_1	l_2	l_3	l_4	l_5	l_6	$\cdots$	$l_{\lvert\mathcal{L}\rvert-1}$	$l_{\lvert\mathcal{L}\rvert}$
u_1	?	?	164	?	1	?	$\cdots$	?	1
u_2	40	2	?	?	?	1	$\cdots$	?	?
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ldots$	$\vdots$	$\vdots$
$u_{\lvert\mathcal{U}\rvert-1}$	?	?	1	1	?	?	$\cdots$	2	?
$u_{\lvert\mathcal{U}\rvert}$	?	2	?	?	1	?	$\cdots$	?	10

Figure 1: User-location check-in frequency matrix.

In the LBSNs, the check-in data contain the following unique characteristics:

- **Frequency Data and Sparsity**. An observed entry in the user-location matrix indicates the frequency of a user visiting a place. Hence, only positive data are available in the task, see Fig. 1 for an illustration. The density of the dataset is about 2.08×10^{-4}, which makes the POI recommendation task very tough.
- **Multi-centers and Normal Distribution**. Users tend to check in around several centers, where the check-in locations follow a Gaussian distribution at each center as shown in Fig. 2 for a typical user's check-in behavior.
- **Inverse Distance Rule**. Although each user contains different personalized taste for POI, the probability of visiting a place is inversely proportional to the distance from its nearest center (see Fig. 3(a) for more information). This implies that if a place is too far away from the location a user lives, although he/she may like that place, he/she would probably not go there.
- **Friendship Influence**. The average overlap of a user's check-ins to his/her friends' check-ins is about 9.6%. This implies that social influence exists, but the effect may be limited.

In this paper, in order to provide more accurate and efficient POI recommendation, we propose a novel fused matrix factorization (MF) framework to take into account the above four factors. Our contributions are threefold. First, we mine the dataset and extract the characteristics of the crawled large-scale data from an LBSNs website. Second, based on the data properties, we model the probability of a user's check-in on a location as a Multi-center Gaussian



(a) Multi-center overview (b) Center 1

(c) Center 2 (d) Center 3 and 4

Figure 2: A typical user's multi-center check-in behavior.

Model (MGM). This is different from an early POI recommendation work in LBSNs (Ye et al. 2011), which assumes a power-law distribution of the check-in probability with respect to the distance within the whole check-in history. Third, we consider the social influence, and more importantly, we utilize the inverse distance rule and incorporate multi-center geographical influence into the fused MF framework. Our comparison on the large-scale real-world LBSNs data shows that fusing MF with MGM can achieve significantly better performance than other state-of-the-art methods.

Related Work

Location-based service (LBS) research becomes prevalent (Lu, Tseng, and Yu 2011; Yang et al. 2011a; Yang, King, and Lyu 2011; Zheng et al. 2011) due to a wide range of potential applications, e.g., personalized marketing strategy analysis, personalized behavior study, context-aware analysis, etc. In particularly, POI recommendation has attracted much research interest in recent years (Kang, Kim, and Cho 2006; Horozov, Narasimhan, and Vasudevan 2006; Zheng et al. 2009; 2010a; Leung, Lee, and Lee 2011). In the following, we review several main approaches in collaborative filtering communities.

One line of research is to solve POI recommendation based on the extracted stay points from GPS trajectory logs of several hundred monitored users (Zheng et al. 2009; 2010a; 2010b; Leung, Lee, and Lee 2011; Zheng and Xie 2011; Cao, Cong, and Jensen 2010). In (Zheng et al. 2010a), three matrices, location-activity, location-feature, and activity-activity, are constructed and a collective matrix factorization method is proposed to mine POI and activities. A tensor factorization is conducted on the user-location-activity relationship to be solved the same problem using the same dataset (Zheng et al. 2010b). In (Leung, Lee, and Lee 2011), a Collaborative Location Model (CLM) is proposed to incorporate activity to facilitate the recommendation.

The other line of work centers on POI recommendation based on the LBSNs data (Ye, Yin, and Lee 2010; Ye et al. 2011). A pioneer work of POI recommendation in LBSNs debuts in (Ye, Yin, and Lee 2010). The work has been extended and further studied in (Ye et al. 2011). More specifically, geographical influence is considered by assuming a power-law distribution between the check-in probability and the distance along the whole check-in history (Ye et al. 2011). However, they ignore users' multi-center check-in behavior. Moreover, the proposed memory-based CF method has to compute all pairwise distances of the whole visiting history. This is time consuming which makes it impossible to solve large-scale datasets.

In summary, GPS data are usually in small-scale, about one or two hundred users, but the data are very dense. Contrarily, LBSNs data are in large-scale, but very sparse (Noulas et al. 2011; Scellato et al. 2011). To solve large-scale recommendation problems, matrix factorization is a promising tool due to its success in Netflix competition (Bell, Koren, and Volinsky 2007; Koren 2009). However, previously proposed methods do not employ MF in LBSNs POI recommendation and do not explore details of the geographical characteristics in the LBSNs data. The above overlook and the significance of POI recommendation in LBSNs motivate us to conduct the work in this paper.

Table 1: Basic statistics of the Gowalla dataset.

#U	#L	#E
53,944	367,149	306,958
#$\widetilde{U}$	#$\widetilde{L}$	#$\widetilde{E}$
51.33	7.54	11.38
#max. U	#max. L	#max. E
2,145	3,581	2,366

Check-in Data Characteristics

The check-in data we explore are crawled from one of the most popular online LBSNs — Gowalla [1]. Gowalla is an LBSN website created in 2009 for users to check in to locations through mobile devices. We collect a complete snapshot, including users' profile, users' check-in locations, check-in time stamps, users' friend list, and location details, from Gowalla during the period from February 2009 to September 2011 via the provided public API. To reduce noise data, we remove users with less than 10 check-ins and locations with less than 20 visits. We then create a LBSNs dataset, whose basic statistics are summarized in Table 1. Details of the data are depicted in the following:

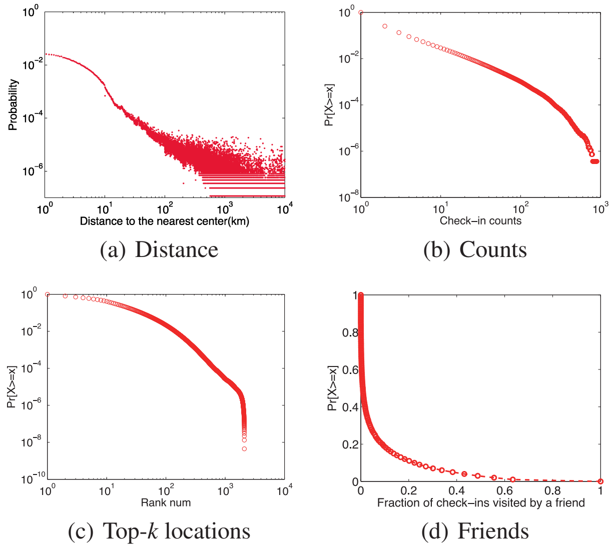
- The dataset consists of 4,128,714 check-ins from 53,944 users on 367,149 locations, and totally 306,958 edges in the whole users' social graph. The density of the dataset is about 2.08×10^{-4}.
- The average number of visited locations of a user is 51.33. The average number of visited users for a location is 7.54.

[1] http://gowalla.com/

The average number of friends for a user is 11.38.

- The maximum number of locations for a user is 2,145. The maximum number of visited users for a location is 3,581. The maximum number of friends for a user is 2,366.

In the following, we further study the location distribution, frequency distribution, and the social relationship among users' check-ins.



(a) Distance

(b) Counts

(c) Top-*k* locations

(d) Friends

Figure 3: Check-ins probability vs. distance, counts, top-*k* locations, common check-ins of friends.

Location Distribution

Figure 2(a) shows the longitude and latitude of a typical user's check-in locations, where the locations form four centers. Figure 2(b)-2(d) further show the details of each center. This observation yields our assumption different from the power-law distribution on users' check-in history in (Ye et al. 2011). In addition, our statistic is also a little different from the two states ("home" and "office") check-in behavior mentioned in (Cho, Myers, and Leskovec 2011). After examining the comments of locations, we find that other than the centers of "home" and "office" (counting above half of a user's check-ins), other centers count at least 10% of check-ins. These centers may be a user's usual business travel places, e.g., an office of a branch of a large company, or vocation places, which provide abundant information that need to be differentiated.

Frequency Distribution

Figure 3(b) plots the Complementary Cumulative Distribution Function (CCDF) for the number of each user's check-in numbers at each location. It is shown that about 74% of locations are only visited once and only about 3% of locations are visited more than 10 times. This means that users usually visit several important places, e.g., home, office, and some stores or bars, with very high frequency, while most of other places are seldom visited. Overall, these places are around several centers. Figure 3(c) further shows the CCDF function of top-*k* frequently visited locations. The most visited location accounts for about 18.8% of all users' check-ins. The top-10 most visited locations account for 68% of all check-ins and the ratio increases to 80.5% for the top-20 most visited locations, following the *Pareto principle* (aka. 80-20 rule) (Hafner 2001).

Social Influence

In the dataset, we find that the average overlap of a user's check-ins to his/her friends' check-ins is only 9.6%. This indicates that less than 10% of a user's check-ins are also visited by his/her friends, which is similar to the statistic reported in (Cho, Myers, and Leskovec 2011). Figure 3(d) plots the CCDF of the fraction of a user's check-ins that are visited by his/her friends. It is known that about 38% of users, their check-in locations are no checked in by their friends, while almost 90% of users contain less than 20% of common check-ins with their friends. The statistics are a little different from those in (Cho, Myers, and Leskovec 2011), but the overall trend is similar. These observations imply that social relationship has limited effect on users' check-ins, but they still cannot be ignored.

Recommendation with Social and Geographical Information

The problem of personalized POI recommendation is defined as follows: given a partially observed user-location check-in frequency matrix with users in $\mathcal{U}$ and locations in $\mathcal{L}$, users' social relationship $\mathcal{F}$, and the longitude and latitude of check-in locations, the task is to recommend top-*k* locations to a user that he/she does not visit before. To solve this problem, we first propose a personalized Multi-center Gaussian Model (MGM) to capture the geographical influence on a user's check-ins. We then depict the matrix factorization, consider the social information, and propose a fused MF framework to include geographical influence.

Multi-center Gaussian Model (MGM)

A significant characteristic of check-in locations is that they are usually located around several centers as shown in Fig. 2. The second characteristic of check-in locations is that the probability of a user visiting a location is inversely proportional to the distance from its nearest center; see Fig. 3(a).

These two characteristics indicate that geographical information plays strong influence on users' check-in behavior. Based on statistics from Fig. 2 and Fig. 3(a), we adopt Gaussian distribution to model users' check-in behavior and propose the Multi-center Gaussian Model (MGM). That is, the probability of a user u, visiting a POI l, given the multi-center set C_u, is defined by:

$$P(l|C_u) = \sum_{c_u=1}^{|C_u|} P(l \in c_u) \frac{f_{c_u}^{\alpha}}{\sum_{i\in C_u} f_i^{\alpha}} \frac{\mathcal{N}(l|\mu_{c_u}, \Sigma_{c_u})}{\sum_{i\in C_u} \mathcal{N}(l|\mu_i, \Sigma_i)}. \quad (1)$$

Here, l denotes the longitude and latitude of a position, C_u is the set of centers for the user u. For each center, calculating Eq. (1) consists of the multiplication of three terms:

- $P(l \in c_u) \propto 1/dist(l, c_u)$ determines the probability of the location l belonging to the center c_u, which is inversely proportional to the distance between the location l and the center c_u. This is based on the *inverse distance rule* observed from Fig. 3(a).
- The second term denotes the normalized effect of check-in frequency f_{c_u}, on the center c_u. The parameter $\alpha \in (0,1]$ is introduced to maintain the frequency aversion property, where very high check-in frequency does not play too significant effect.
- The third term denotes the normalized probability of a location belonging to the center c_u, where $\mathcal{N}(l|\mu_{c_u}, \Sigma_{c_u})$ is the probability density function of the Gaussian distribution, μ_{c_u} and Σ_{c_u} correspond to the mean and covariance matrices of regions around the center c_u.

Now, the problem turns to finding the centers. Here, we propose a greedy clustering algorithm among the check-ins due to the Pareto principle (Hafner 2001); see statistics in Fig. 2 and Fig. 3(c)). More advance techniques to calculate data similarity can be referred to (Yang et al. 2011b). We scan from the most visited POI and combine all other visited check-in locations, whose distance is less than d kilometers from the selected POI, into a region. If the ratio of the total check-in number of this region to the user's total check-in amount is greater than a threshold θ, we set these check-in positions as a region and determine its center. Algorithm 1 shows the procedure of discovering multiple centers. In our experiment, by trial on the training dataset, we set θ to 0.02, the the distance threshold d to 15 and the frequency control parameter α to 0.2.

Algorithm 1 Multi-center Discovering Algorithm

```
for all user $i$ in the user set $\mathcal{U}$ do
    Rank all check-in locations in $|\mathcal{L}|$ according to visiting frequency
    $\forall l_k \in L$, set $l_k.center = -1$;
    Center_list = $\emptyset$; center_no = 0;
    for $i = 1 \to |L|$ do
        if $l_i.center == -1$ then
            center_no++; Center = $\emptyset$; Center.total_freq = 0;
            Center.add($l_i$); Center.total_freq += $l_i$.freq;
            for $j = i + 1 \to |L|$ do
                if $l_j.center == -1$ and $dist(l_i, l_j) \leq d$ then
                    $l_j.center =$ center_no; Center.add($l_j$);
                    Center.total_freq += $l_j$.freq;
                end if
            end for
            if Center.total_freq $\geq |u_i|$.total_freq * $\theta$ then
                Center_list.add(Center);
            end if
        end if
    end for
    RETURN Center_list for user $i$;
end for
```

Matrix Factorization

Matrix Factorization (MF) is one of the most popular methods for recommender systems (Salakhutdinov and Mnih 2007; 2008; Bell, Koren, and Volinsky 2007; Koren 2009). Given the partial observed entries in a $|\mathcal{U}| \times |\mathcal{L}|$ frequency matrix F, the goal of MF is to find two low-rank matrices $U \in \mathbb{R}^{K \times |\mathcal{U}|}$ and $L \in \mathbb{R}^{K \times |\mathcal{L}|}$ such that $F \approx U^T L$. The predicted probability of a user u, like a location l, is determined by

$$P(F_{ul}) \propto U_u^T L_l. \tag{2}$$

Probabilistic Matrix Factorization (PMF) PMF is one of the most famous MF models in collaborative filtering (Salakhutdinov and Mnih 2007). It assumes Gaussian distribution on the residual noise of observed data and also places Gaussian priors on the latent matrices U and V. The corresponding objective function of PMF for the frequency data is defined as follows:

$$\min_{U,L} \sum_{i=1}^{|\mathcal{U}|} \sum_{j=1}^{|\mathcal{L}|} I_{ij}(g(F_{ij}) - g(U_i^T L_j))^2 + \lambda_1 \|U\|_F^2 + \lambda_2 \|L\|_F^2, \tag{3}$$

where $g(x) = 1/(1+\exp(-x))$ is the logistic function, I_{ij} is the indicator function which equals to 1 if user i checks in the location j and equals 0 otherwise. $\|\cdot\|_F$ denotes the Frobenius norm.

Note: Since the observed frequency data is all positive, we sample the same number of unobserved data from the rest matrix and deem them as the frequency to 0. This is a standard way to solve the one-class problem in CF (Pan et al. 2008; Pan and Scholz 2009).

PMF with Social Regularization (PMFSR) We further include the social information into the PMF as (Ma et al. 2008; Zhou et al. 2009; Ma et al. 2011c; 2011b) to improve the model performance. We adopts the PMF with Social Regularization (PMFSR) (Ma et al. 2011b), whose objective function is defined as follows:

$$\begin{aligned}\min_{U,L} \Omega(U,L) &= \sum_{i=1}^{|\mathcal{U}|} \sum_{j=1}^{|\mathcal{L}|} I_{ij}(g(F_{ij}) - g(U_i^T L_j))^2 \\ &+ \beta \sum_{i=1}^{|\mathcal{U}|} \sum_{f \in \mathcal{F}(i)} Sim(i,f) \|U_i - U_f\|_F^2 \\ &+ \lambda_1 \|U\|_F^2 + \lambda_2 \|L\|_F^2, \end{aligned} \tag{4}$$

where $\mathcal{F}(i)$ is the set of friends for user u_i, and $Sim(i,f)$ is the similarity between user u_i and his friend u_f.

Probabilistic Factor Models (PFM) Since PMF outputs poor performance in our preliminary results (see Fig. 4), we turn to Probabilistic Factor Models (PFM) (Chen et al. 2009; Ma et al. 2011a), which can model the frequency data directly. PFM places Beta distributions as priors on the latent matrices U and V, while defines a Poisson distribution on the frequency. This leads to seeking U and V by minimizing

$\Psi(U, L; F)$, which is defined by:

$$\begin{aligned}\Psi(\cdot,\cdot;\cdot) \;=\; & \sum_{i=1}^{|\mathcal{U}|}\sum_{k=1}^{K}((\alpha_k-1)\ln(U_{ik}/\beta_k)-U_{ik}/\beta_k) \\ & +\sum_{j=1}^{|\mathcal{L}|}\sum_{k=1}^{K}((\alpha_k-1)\ln(L_{jk}/\beta_k)-L_{jk}/\beta_k) \\ & +\sum_{i=1}^{|\mathcal{U}|}\sum_{j=1}^{|\mathcal{L}|}(F_{ij}\ln(U^TL)_{ij}-(U^TL)_{ij})+c, \quad (5)\end{aligned}$$

where $\alpha = \{\alpha_1, \ldots, \alpha_K\} > \mathbf{0}_K, \beta = \{\beta_1, \ldots, \beta_K\} > \mathbf{0}_K$ are parameters for Beta distributions, and c is a constant term.

Fusion Framework

The matrix factorization methods only model users' preference on locations. They do not explore the geographical influence. As observed from Fig. 3(a), users tend to check in locations around their centers. Hence, we fusion users' preference on a POI and the probability of whether a user will visit that place together to determine the probability of a user u visits a location l, which is defined as follows:

$$P_{ul} = P(F_{ul}) \cdot P(l|C_u), \qquad (6)$$

where $P(l|C_u)$ is calculated by Eq. (1) via the MGM and $P(F_{ul})$ encodes users' preference on a location determined by Eq. (2).

Complexity Analysis

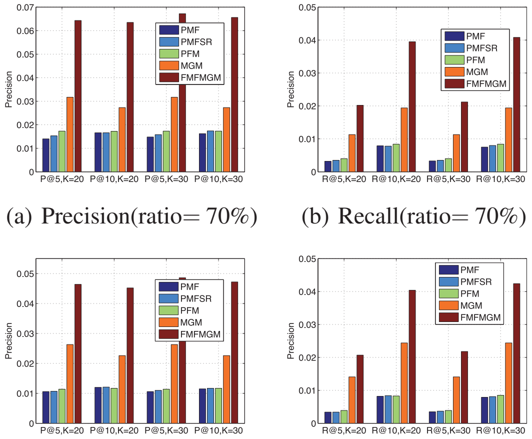
The computation cost consists of the calculation of matrix factorization models and calculating the probability of a user visiting a POI. The training time for the matrix factorization models scales linearly with the number of observations (Salakhutdinov and Mnih 2007; Ma et al. 2011b). For the probability computation, the cost is to calculate the centers. This also scales linearly with the number of observations. Hence, our proposed method is efficient and can scale up to very large datasets.

Experiments

The experiments address the following questions: 1) How does our approach compare with the baseline and the state-of-the-art algorithms? 2) What is the performance on users with different check-in frequency? This is a scenario for cold-start users whose check-ins are few.

Setup and Metrics

The experimental data include user-location check-in records, users' friendship list, and geographical information (longitude and latitude of check-in locations). We split the crawled Gowalla dataset into two non-overlapping sets: a training set and a test set, where the proportion of training data is test on 70% and 80%, respectively. Here, training data 70%, for example, means we randomly select 70% of the observed data for each user as the training data to predict the remaining 30% data. The random selection was carried out 5 times independently, and we report the average results. The hyperparameters are tuned on the training dataset.



(a) Precision(ratio= 70%) (b) Recall(ratio= 70%)

(c) Precision(ratio= 80%) (d) Recall(ratio= 80%)

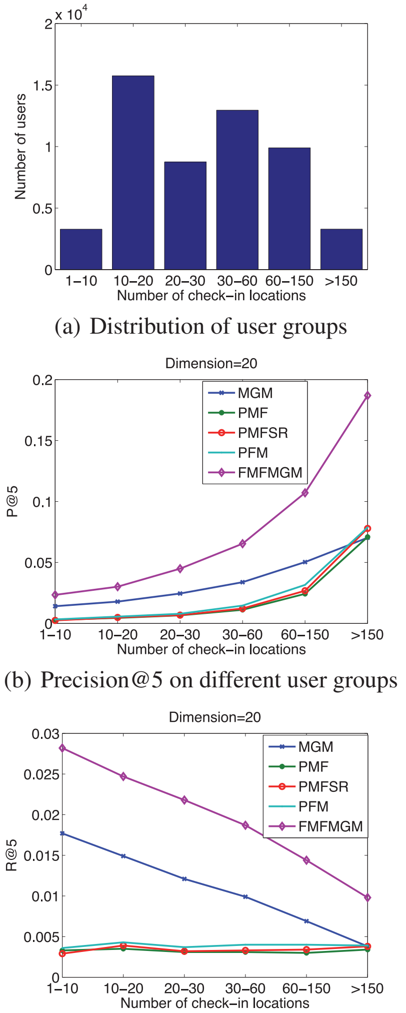
Figure 4: Performance Comparison

POI recommendation is to recommend the top-N highest ranked positions to a targeted user based on a ranking score from a recommendation algorithm. To evaluate the model performance, we are interested in finding out how many locations in the test set are recovered in the returned POI recommendation. Hence, we use the Precision@N and Recall@N as the metrics to evaluate the returned ranking list against the check-in locations where users actually visit. These two metrics are standard metrics to measure the performance of POI recommendation (Ye et al. 2011). Precision@N defines the ratio of recovered POI to the N recommended POI and Recall@N defines the ratio of recovered POI to the size of test set. In the experiment, N is set to 5 and 10, respectively.

Comparison

In the experiments, the compared approaches include:

1. **Multi-center Gaussian Model (MGM)**: this method recommends a position based on the probability calculated by Eq. (1).
2. **PMF**: this is a well-known method in matrix factorization (Salakhutdinov and Mnih 2007). Its objective function is shown in Eq. (3).
3. **PMF with Social Regularization (PMFSR)**: this method is proposed to include the social friendship under the PMF framework (Ma et al. 2011b). Its objective function is shown in Eq. (4).
4. **Probabilistic Factor Models (PFM)**: this method is a promising method to model frequency data (Ma et al. 2011a). Its objective function is shown in Eq. (5).
5. **FMF with MGM (FMFMGM)**: this is the fused matrix factorization framework with the Multi-center Gaussian Model (FMFMGM). The users' preference on locations is calculated by the PFM model. Here, we select PFM because PFM can model the frequency data better than PMF.



(a) Distribution of user groups

(b) Precision@5 on different user groups

(c) Recall@5 on different user groups

Figure 5: Performance comparison on different user groups

Figure 4 reports the average of five run results on the top 5 and top 10 recommendation by the compared models using 20 and 30 as the number of latent feature dimensions, respectively. The results show that:

- MGM and FMFMGM outperform PMF, PMFSR, and PFM, significantly in all metrics. For example, FMFMGM attains 0.0643 and MGM attains 0.0317 P@5 when the latent dimension is 20 and 70% of data are using for training, while PFM, the best model without considering location information, achieves 0.0173 for the counter part. This implies that geographical influence plays a significant role in POI recommendation. By utilizing the geographical influence, we can provide much more accurate POI recommendation to targeted users.
- FMFMGM achieves much significantly better performance, at least 50%, than the MGM. That is, for case of the latent dimension being 30 and 80% of data for training, the performance increases from 0.0141 for MGM to 0.0218 for FMFMGM. This verifies that the probability of a user visiting a POI is controlled by both the user's personal preference and the personal check-in location constraints. By utilizing users' personalized tastes captured by MF models, we can attain more accurate prediction.
- PMFSR attains a little better results than those of PMF. This shows that social influence is not so important in POI recommendation and it also coincides the fact that friends share very low, only 9.6% common POI.

Performance on Different Users

One challenge of the POI recommendation is that it is difficult to recommend POI to those users who have very few check-in history. In order to compare our method with the other methods thoroughly, we first group all the users based on the frequency of observed check-ins in the training set, and then evaluate the model performances within different user groups. Figure 5 shows the compared performance on different user groups. Here, users are grouped into 6 types: "1-10", "10-20", "20-30", "30-60","60-150", and ">150", which denotes the frequency range of users' check-ins in the training data.

Figure 5(a) summarizes the distribution on different ranges of users' check-in frequency in 70% of the training data. It is shown that the number of users' check-in locations mainly lies in the range of 10 and 150. From Fig. 5(b) and Fig. 5(c), we observe that our FMFMGM method consistently outperforms other compared methods in terms of P@5 and R@5. When users' check-in frequency is small, MGM outperforms PMF, PMFSR, and PFM. But when users' check-in frequency becomes larger, PMF, PMFSR, and PFM performs better than MGM. It is reasonable since when users' check-in frequency is small, especially for cold-start users, it is difficult to learn users' preferences, and thus geographical information plays more influence on the prediction. When more check-in information is available, both users' preferences and geographical influence can be learned more accurately, but users' preferences dominate the geographical influence. More importantly, when combining them together, our FMFMGM method performs much better than other methods.

Conclusions and Future Work

In this paper, we have detailedly investigated the characteristics of the large-scale check-in data from a popular LBSNs website, Gowalla. Based on the extracted properties of the data, we propose a novel Multi-center Gaussian Model to model the geographical influence of users' check-in behavior. We then consider users' social information, and propose a fused matrix factorization method to include the geographical influence of users' check-in locations. Results from extensive experiments show that our proposed method outperforms other state-of-the-art approaches significantly.

There are several directions worthy of considering for future study: 1) how to model extremely sparse frequency data, e.g., by designing more subtle sampling techniques, to improve MF methods; 2) how to include other information, e.g., location category, and activity, into our fused framework; 3) how to incorporate temporal effect on POI recommendation to capture the change of users' preference.

Acknowledgments

The work described in this paper was fully supported by two grants from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. CUHK 413210 and CUHK 415311) and two grants from Google Inc. (one for Focused Grant Project “Mobile 2014” and one for Google Research Awards).

References

Bell, R. M.; Koren, Y.; and Volinsky, C. 2007. Modeling relationships at multiple scales to improve accuracy of large recommender systems. In *KDD*, 95–104.

Cao, X.; Cong, G.; and Jensen, C. 2010. Mining significant semantic locations from gps data. *Proceedings of the VLDB Endowment* 3(1-2):1009–1020.

Chen, Y.; Kapralov, M.; Pavlov, D.; and Canny, J. 2009. Factor modeling for advertisement targeting. In *NIPS*, 324–332.

Cho, E.; Myers, S. A.; and Leskovec, J. 2011. Friendship and mobility: user movement in location-based social networks. In *KDD*, 1082–1090.

Hafner, A. 2001. Pareto’s principle: The 80-20 rule. *Retrieved December* 26:2001.

Horozov, T.; Narasimhan, N.; and Vasudevan, V. 2006. Using location for personalized poi recommendations in mobile environments. In *Proceedings of the International Symposium on Applications on Internet*, SAINT ’06, 124–129. Washington, DC, USA: IEEE Computer Society.

Kang, E.-y.; Kim, H.; and Cho, J. 2006. Personalization method for tourist point of interest (POI) recommendation. In Gabrys, B.; Howlett, R.; and Jain, L., eds., *Knowledge-Based Intelligent Information and Engineering Systems*, volume 4251 of *Lecture Notes in Computer Science*. Berlin, Heidelberg: Springer Berlin / Heidelberg. chapter 48, 392–400.

Koren, Y. 2009. Collaborative filtering with temporal dynamics. In *KDD*, 447–456.

Leung, K. W.-T.; Lee, D. L.; and Lee, W.-C. 2011. Clr: a collaborative location recommendation framework based on co-clustering. In *SIGIR*, 305–314.

Lu, E. H.-C.; Tseng, V. S.; and Yu, P. S. 2011. Mining cluster-based temporal mobile sequential patterns in location-based service environments. *IEEE Trans. Knowl. Data Eng.* 23(6):914–927.

Ma, H.; Yang, H.; Lyu, M.; and King, I. 2008. Sorec: social recommendation using probabilistic matrix factorization. In *Proceedings of the 17th ACM conference on Information and knowledge management*, 931–940. ACM.

Ma, H.; Liu, C.; King, I.; and Lyu, M. 2011a. Probabilistic factor models for web site recommendation. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information*, 265–274. ACM.

Ma, H.; Zhou, D.; Liu, C.; Lyu, M.; and King, I. 2011b. Recommender systems with social regularization. In *Proceedings of the fourth ACM international conference on Web search and data mining*, 287–296. ACM.

Ma, H.; Zhou, T.; Lyu, M.; and King, I. 2011c. Improving recommender systems by incorporating social contextual information. *ACM Transactions on Information Systems (TOIS)* 29(2):9.

Noulas, A.; Scellato, S.; Mascolo, C.; and Pontil, M. 2011. An empirical study of geographic user activity patterns in foursquare. *ICWSM11*.

Pan, R., and Scholz, M. 2009. Mind the gaps: weighting the unknown in large-scale one-class collaborative filtering. In *KDD*, 667–676.

Pan, R.; Zhou, Y.; Cao, B.; Liu, N. N.; Lukose, R. M.; Scholz, M.; and Yang, Q. 2008. One-class collaborative filtering. In *ICDM*, 502–511.

Salakhutdinov, R., and Mnih, A. 2007. Probabilistic matrix factorization. In *NIPS*.

Salakhutdinov, R., and Mnih, A. 2008. Bayesian probabilistic matrix factorization using markov chain monte carlo. In *ICML*, 880–887.

Scellato, S.; Noulas, A.; Lambiotte, R.; and Mascolo, C. 2011. Socio-spatial properties of online location-based social networks. *Proceedings of ICWSM* 11.

Yang, H.; Chen, S.; Lyu, M. R.; and King, I. 2011a. Location-based topic evolution. In *Proceedings of the 1st international workshop on Mobile location-based service*.

Yang, H.; Xu, Z.; Ye, J.; King, I.; and Lyu, M. R. 2011b. Efficient sparse generalized multiple kernel learning. *IEEE Transactions on Neural Networks* 22(3):433–446.

Yang, H.; King, I.; and Lyu, M. R. 2011. *Sparse Learning Under Regularization Framework*. LAP Lambert Academic Publishing, first edition.

Ye, M.; Yin, P.; Lee, W.-C.; and Lee, D. L. 2011. Exploiting geographical influence for collaborative point-of-interest recommendation. In *SIGIR*, 325–334.

Ye, M.; Yin, P.; and Lee, W. 2010. Location recommendation for location-based social networks. In *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 458–461. ACM.

Zheng, Y., and Xie, X. 2011. Learning travel recommendations from user-generated gps traces. *ACM TIST* 2(1):2.

Zheng, Y.; Zhang, L.; Xie, X.; and Ma, W.-Y. 2009. Mining interesting locations and travel sequences from gps trajectories. In *WWW*, 791–800.

Zheng, V. W.; Zheng, Y.; Xie, X.; and Yang, Q. 2010a. Collaborative location and activity recommendations with gps history data. In *WWW*, 1029–1038.

Zheng, V.; Cao, B.; Zheng, Y.; Xie, X.; and Yang, Q. 2010b. Collaborative filtering meets mobile recommendation: A user-centered approach. In *Proceedings of the 24rd AAAI Conference on Artificial Intelligence*.

Zheng, Y.; Zhang, L.; Ma, Z.; Xie, X.; and Ma, W.-Y. 2011. Recommending friends and locations based on individual location history. *TWEB* 5(1):5.

Zhou, T. C.; Ma, H.; King, I.; and Lyu, M. R. 2009. Tagrec: Leveraging tagging wisdom for recommendation. In *Proceedings of CSE*, 194–199.