iGSLR: Personalized Geo-Social Location Recommendation - A Kernel Density Estimation Approach

Jia-Dong Zhang Chi-Yin Chow
Department of Computer Science, City University of Hong Kong, Hong Kong
jzhang26@student.cityu.edu.hk chiychow@cityu.edu.hk

ABSTRACT

With the rapidly growing location-based social networks (LBSNs), personalized geo-social recommendation becomes an important feature for LBSNs. Personalized geo-social recommendation not only helps users explore new places but also makes LBSNs more prevalent to users. In LBSNs, aside from *user preference* and *social influence*, *geographical influence* has also been intensively exploited in the process of location recommendation based on the fact that geographical proximity significantly affects users' check-in behaviors. Although geographical influence on users should be personalized, current studies only model the geographical influence on all users' check-in behaviors in a universal way. In this paper, we propose a new framework called iGSLR to exploit personalized social and geographical influence on location recommendation. iGSLR uses a kernel density estimation approach to personalize the geographical influence on users' check-in behaviors as individual distributions rather than a universal distribution for all users. Furthermore, *user preference*, *social influence*, and *personalized geographical influence* are integrated into a unified geo-social recommendation framework. We conduct a comprehensive performance evaluation for iGSLR using two large-scale real data sets collected from Foursquare and Gowalla which are the two of the most popular LBSNs. Experimental results show that iGSLR provides significantly superior location recommendation compared to other state-of-the-art geo-social recommendation techniques.

Categories and Subject Descriptors

H.2.8 [**Database Management**]: Database Applications-Spatial databases and GIS; H.3.3 [**Information Search and Retrieval**]: Information Filtering

General Terms

Algorithms, Experimentation.

Keywords

Location-based social networks, location recommendation, personalized geographical influence, kernel density estimation, social influence

1. INTRODUCTION

With the advancement of mobile devices, wireless communication and location acquisition technologies, location-based social networks (LBSNs), such as Foursquare, Gowalla, and Facebook places, have attracted millions of users. In an LBSN (Figure 1), users can establish social links and share their experiences of visiting some specific locations, also known as *points-of-interest* (POIs), e.g., restaurants, stores, and museums. These visits are also known as *check-in* activities that reflect users' preferences on locations. It is an essential task of LBSNs to utilize user preferences and other information (e.g., social friendships) to make location recommendation to users, which not only helps users explore new places but also makes LBSNs more attractive to users.

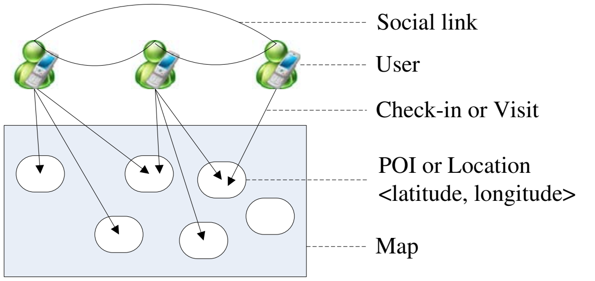


Figure 1: A location-based social network

A promising way for location recommendation is to apply the conventional *collaborative filtering* (CF) techniques. CF treats user preferences derived from check-in activities as a *user-location rating matrix* in which each entry represents the frequency of a user visiting a location. CF has been widely employed for location recommendation (e.g., [1, 2, 6, 14, 24, 25, 26]). Moreover, in terms of the argument that friends are more likely to share common interests [11, 12, 15], CF techniques have also exploited social friendships as a *user-user similarity matrix* to improve the quality of location recommendation (e.g., [1, 2, 6, 14, 24, 25, 26]); this kind of CF is also known as *social collaborative filtering* (SCF). However, the improvement of SCF could be considerably limited, because in general users with social friendships only share less than 10% commonly visited locations [2, 3, 24].

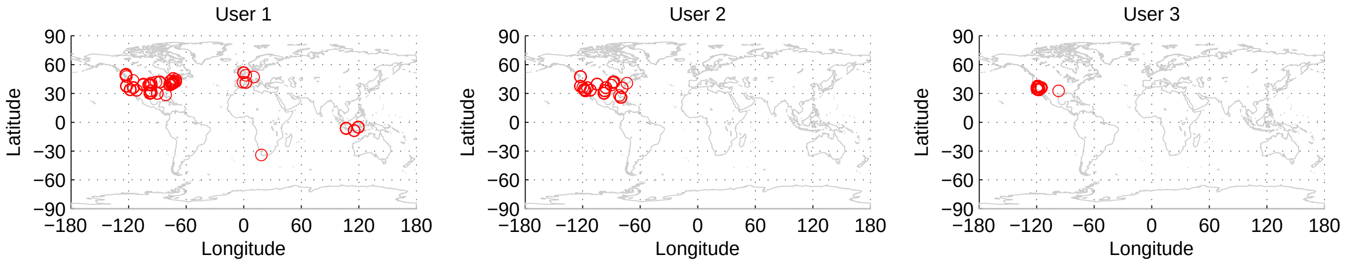


Figure 2: Distributions of personal check-in locations

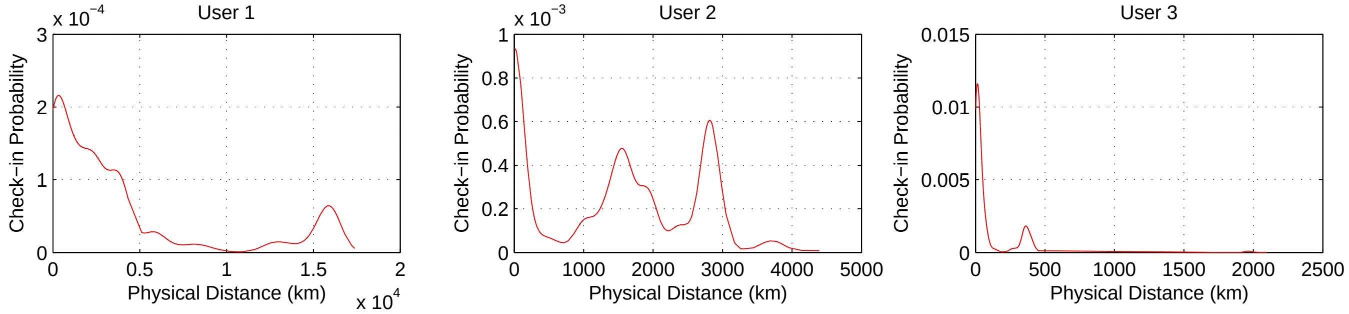


Figure 3: Personal check-in probabilities over geographical distances

Therefore, some researchers have turned to explore two kinds of geographical influence on users. A simple way is to utilize **the geographical influence of users**, i.e., the distance between the residences of users with social friendships, to adjust their similarity weight in SCF, because nearby friends share more commonly visited locations than others [6, 14, 24, 26]. However, users often travel from one place to another and hence their static residences may not reflect their actual geographical positions. As a result, the improvement on the quality of location recommendation is also very limited by augmenting the geographical information of users' residences.

A better way is to exploit **the geographical influence of locations**, i.e., the distance between every pair of locations visited by the same user, to model all users' check-in behaviors. On the one hand, some researchers [25] assume that the distance of visited locations follows a *power-law distribution* (PD), where the model parameters are derived from the whole check-in history database. On the other hand, the work [2] clusters the whole check-in history database to find the most popular POIs as centers and assumes that the distance between visited locations and their centers follows a *multi-center Gaussian model* (MGM). In these two studies, the obtained distribution is used to deduce the probability of a user visiting a new location; this probability is fused with the predicted rating of the user to the new location given by CF or SCF. Both have shown that recommending locations based on fused ratings can enhance the quality of location recommendation to some extent.

Nevertheless, the geographical influence of locations is universally modeled as a common distribution for all users in [2, 25]. On the contrary, in reality the geographical influence of locations on users' check-in behaviors should be unique and personal. For instance, indoorsy persons like visiting POIs around their living areas while outdoorsy persons prefer traveling around the world to explore new POIs. Thus, we argue that the geographical influence of locations on individual users' check-in behaviors should be personalized when recommending locations for users. In fact, personalization is one of the most essential requirements of recommendation that can help alleviate the problem of information overload and is an important enabler of the success of e-business [18].

Real-world motivating examples. To observe users' unique check-in behaviors, a spatial analysis is conducted on two publicly available real data sets collected from Foursquare [6] and Gowalla [3], which are the two of the most popular LBSNs. Specifically, we focus on three users with the largest number of visited locations in each data set and have observed similar findings in two data sets. Here, due to space limitation we only show the three users' check-in locations in the Foursquare data set in Figure 2. The geographical influence of locations on these three users' check-in behaviors is unique: User 1 travels around the world, e.g., North America, Europe, South Africa, and South Asia; User 2 moves around in the United States of America; and User 3 usually visits POIs around her living area, i.e., Los Angeles. To further understand the geographical influence on these three users' check-in behaviors, Figure 3 depicts their individual check-in distribution over the distance between every pair of POIs visited by the same user. Their distance distributions are also unique, so it is undesirable to model them as a universal distribution, e.g., PD [25] and MGM [2].

In this paper, we propose a new personalized geo-social location recommendation framework called iGSLR. In iGSLR, we are motivated to explore **the personalized geographical influence of locations** on users' check-in behaviors. (1) We design a new method to model the personalized geographical influence of locations as an individual distance distribution for each user rather than a common distribution for all users. (2) At the same time, we do not have any assumption about the form of a distance distribution, because we model it based on a nonparametric method, i.e., the popular kernel density estimation, an attractive feature of which is that it can be used with arbitrary distributions [21].

Furthermore, in iGSLR we develop a unified geo-social recommendation framework to integrate *the personalized geographical influence of locations* with *user preference*, *social influence*, and *the geographical influence of users*. We handle the geographical influences of users and locations in different steps since they are essentially distinct. The distance between the residences of users is highly related to users and only implies the correlation between users. In contrast, the distance between every pair of locations visited by a user is greatly related to the check-in behavior of the user and reflects the correlation between locations. Therefore, the latter should be more powerful than the former when used for location recommendation. Specifically, we implement the fusion framework through two steps. (1) We first integrate *user preference*, *social influence*, and *the geographical influence of users* by deriving the similarity measure between users based on social friendships and the geographical influence of users as an input of CF, called **Input-Fusion**. (2) We then apply a product or sum rule to combine the rating and the probability of a user visiting a location that are predicted based on the Input-Fusion step and *the personalized geographical influence of locations*, respectively, called **Output-Fusion**.

The main contributions of this paper can be summarized as follows:

- We explore *the personalized geographical influence of locations* on a user's check-in behavior through learning an individual distance distribution from the user's check-in history based on kernel density estimation. The proposed method does not need any assumption about the form of distribution made in previous studies. Thus, the obtained personalized distance distribution can more accurately estimate the probability of a user checking in at a new location. (Section 3)

- We design a unified geo-social recommendation framework fusing *user preference*, *social influence*, *the geographical influence of users*, and *the personalized geographical influence of locations*. To the best of our knowledge, this is the first study of considering the geographical influence of both users and locations when recommending locations. (Section 4)

- We conduct extensive experiments to evaluate the performance of iGSLR using two large-scale real data sets collected from Foursquare and Gowalla. We also test the performance of iGSLR for cold-start users with only a few check-in records and its performance in the problem of data sparsity. Experimental results show that iGSLR outperforms the state-of-the-art geo-social recommendation techniques including PD [25] and MGM [2]. (Sections 5 and 6)

The remainder of this paper is organized as follows. Section 2 highlights related work and some background on CF techniques. Section 3 describes how to use kernel density estimation to model *the personalized geographical influence of locations* on users' check-in behaviors. We then present our unified location recommendation framework that integrates *user preference*, *social influence*, *the geographical influence of users*, and *the personalized geographical influence of locations* in Section 4. In Sections 5 and 6, we evaluate the performance of iGSLR and analyze experimental results, respectively. Finally, we conclude this paper in Section 7.

2. RELATED WORK

In this section, we highlight related work in recommender systems and present some background on collaborative filtering and social collaborative filtering.

2.1 Recommender Systems

Recommendation techniques. Recommender systems apply knowledge discovery techniques to the problem of making personalized recommendations about information, products, or services in which users are likely to be interested. The techniques used by recommender systems can be generally classified into three main categories: content-based, collaborative filtering, and hybrid recommendation techniques. Among these methods, collaborative filtering (CF) requires a user-item rating matrix (i.e., user preferences) as the input and is the most popular and promising technology to enable personalization in recommender systems [8]. The CF techniques can be divided into *model-based* and *memory-based*. Memory-based methods can be further grouped into user-based CF [10] and item-based CF [19].

Social recommendation. With the rapid growth of social networks, like Facebook and Twitter, social network information, e.g., personal profile content, tags and friendships, has been utilized to improve the quality of recommender systems. This is because such information can be considered as *implicit* user feedbacks to alleviate the data sparsity problem of *explicit* user feedbacks (i.e., the rating matrix) in traditional CF systems [9]. In particular, recommendation techniques integrating social friendships into CF have been widely studied, including model-based methods [5, 7, 15, 16, 20, 22, 23] and memory-based methods [9, 11, 12]. The rationale behind these methods is that friends are more likely to share common interests and thus the social influence should be considered when making recommendation.

Geo-social recommendation. Recently with the emergence of LBSNs, like Foursquare, Gowalla, and Facebook places, recommending locations (i.e., POIs) for users becomes prevalent. For example, some studies provide POI recommendations using GPS trajectory data [13, 27, 28, 29, 30]. However, these techniques have not leveraged any geographical influence when generating recommendation. In reality, POIs are totally different from other non-spatial items, such as books, music and movies in conventional recommender systems, because physical interactions are required for users to visit POIs [25]. Thus, the geographical influence of users and locations plays a significant role in users' check-in behaviors [2, 25].

To exploit geographical influence for improving the quality of location recommendation, some techniques [6, 14, 24, 26] employed the geographical influence of users to update their similarity weights but no consideration for the geographical influence of locations. Some researchers [1] viewed POIs as ordinary non-spatial items and considered the geographical influence of locations by predefining a range; POIs only within this range will be possibly recommended to users. Furthermore, other techniques [2, 25] explored the geographical influence of locations by modeling the distance between two POIs visited by the same user as a common distribution for all users, e.g., a power-law distribution [25] or a multi-center Gaussian distribution [2]. Nonetheless, in practice geographical influence of locations should be unique for each user.

Thus, we consider that the geographical influence of locations on users' check-in behaviors should be personalized during the recommendation process. In this paper, we are motivated to explore *the personalized geographical influence of locations* by modeling the influence as a personalized distance distribution for each user based on kernel density estimation. Furthermore, we integrate the *the personalized geographical influence of locations* with *user preference*, *social influence*, and *the geographical influence of users* for location recommendation through a unified framework.

2.2 Collaborative Filtering

Collaborative filtering (CF) aims at looking for patterns of agreement among users' ratings for items. The intuition behind CF is that if a user has agreed with her neighbors in the past, she will continue to do so for future items [17]. Let U be a set of users, L be a set of locations (i.e., POIs), and R be a user-location rating matrix derived from check-in activities, where each entry $r_{i,j}$ denotes the frequency of user u_i visiting location l_j. Given a certain entry $r_{i,j} = 0$ (i.e., u_i has not visited location l_j), the rating of u_i to unvisited location l_j, denoted as $\hat{r}_{i,j}$, can be predicted using (i) the user-based CF method [10]:

$$\hat{r}_{i,j} = \frac{\sum_{u_k \in U \wedge k \neq i} CosSim(u_i, u_k) \cdot r_{k,j}}{\sum_{u_k \in U \wedge k \neq i} CosSim(u_i, u_k)}, \tag{1}$$

where $CosSim(u_i, u_k)$ in Equation (1) is the cosine similarity measure between users u_i and u_k computed by:

$$CosSim(u_i, u_k) = \frac{\sum_{l_j \in L} r_{i,j} \cdot r_{k,j}}{\sqrt{\sum_{l_j \in L} r_{i,j}^2}\sqrt{\sum_{l_j \in L} r_{k,j}^2}}, \tag{2}$$

or (ii) the item-based CF method [19] ("item" means "location" in the case of LBSNs):

$$\hat{r}_{i,j} = \frac{\sum_{l_k \in L \wedge k \neq j} r_{i,k} \cdot CosSim(l_k, l_j)}{\sum_{l_k \in L \wedge k \neq j} CosSim(l_k, l_j)}, \tag{3}$$

where $CosSim(l_k, l_j)$ in Equation (3) is the cosine similarity measure between locations l_k and l_j computed by:

$$CosSim(l_k, l_j) = \frac{\sum_{u_i \in U} r_{i,k} \cdot r_{i,j}}{\sqrt{\sum_{u_i \in U} r_{i,k}^2}\sqrt{\sum_{u_i \in U} r_{i,j}^2}}. \tag{4}$$

2.3 Social Collaborative Filtering

The user-based and item-based CF techniques do not consider the social influence among users. But in the real world, users usually turn to friends to seek recommendations for books, movies, or POIs. This is because a user's preference can be influenced by her close friends or a group of friends that are likely to share some common interests. For example, friends often go to some places like movie theaters or restaurants together, or a user may travel on spots highly recommended by her friends. Based on these observations, some social collaborative filtering (SCF) methods have been proposed [12, 15]:

$$\hat{r}_{i,j} = \frac{\sum_{u_k \in U \wedge k \neq i} SocSim(u_i, u_k) \cdot r_{k,j}}{\sum_{u_k \in U \wedge k \neq i} SocSim(u_i, u_k)}, \tag{5}$$

where $SocSim(u_i, u_k)$ in Equation (5) is the similarity measure between users u_i and u_k derived from the social influence among users rather than the user-location rating matrix as in the user-based and item-based CF methods.

A simple but effective method can be used to derive a measure of the social similarity between users u_i and u_k [12]:

$$SocSim(u_i, u_k) = \frac{|F(u_i) \cap F(u_k)|}{|F(u_i) \cup F(u_k)|}, \tag{6}$$

where $F(u_i)$ denotes the set of users having social friendships with user u_i.

3. MODELING PERSONALIZED GEOGRAPHICAL INFLUENCE WITH KERNEL DENSITY ESTIMATION

The state-of-the-art studies have shown that the geographical proximity of locations significantly influences users' check-in behaviors. Ye et al. [25] observed that users tend to visit locations close to their homes or offices and also may be interested in exploring the nearby places of their visited locations. They assumed that the geographical distance between two locations visited by the same user following a power-law distribution. Cheng et al. [2] argued that users tend to visit locations around centers (i.e., the most popular POIs) and assumed that the check-in locations follow a Gaussian distribution at each center. They adopted the multi-center Gaussian model to form the distribution of the distance between a visited location and its center. However, these studies apply a *universal* distance distribution for all users to represent **the geographical influence of locations**, which is the geographical influence of the distance between every pair of locations visited by the same user (as defined in Section 1), on users' check-in behaviors.

Kernel density estimation. The experiment results (depicted in Section 1) inspire us to study the *personalized geographical influence of locations* on an individual user's check-in behavior. Concretely, we model the personalized distribution of the distance between any pair of locations visited by the user using kernel density estimation, since it can be used with arbitrary distributions and without the assumption that the form of the distance distribution is known. The kernel density estimation process consists of two steps: *distance sample collection* and *distance distribution estimation.*

Step 1: Distance sample collection. We can acquire a sample for a user by computing the distance between every pair of locations that have been checked in by the user, because each check-in in LBSNs is associated with its user's identity and position (i.e., latitude and longitude coordinates). In particular, for a cold-start user with only one check-in, we instead employ the distance between the only visited location and her residence as the sample.

Step 2: Distance distribution estimation. Let D be the distance sample for a certain user that is drawn from some distribution with an unknown density f. Its kernel density estimator $\hat{f}$ over distance d using D is given by:

$$\hat{f}(d) = \frac{1}{|D|h} \sum_{d' \in D} K\left(\frac{d - d'}{h}\right), \tag{7}$$

where $K(\cdot)$ is the kernel function and h is a smoothing parameter, called the bandwidth. In this paper we apply the most popular normal kernel:

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \tag{8}$$

and the optimal bandwidth [21]:

$$h = \left(\frac{4\hat{\sigma}^5}{3n}\right)^{1/5} \approx 1.06\hat{\sigma}n^{-1/5}, \tag{9}$$

where $\hat{\sigma}$ is the standard deviation of the sample in D.

Probability of a new location. After we find a distance distribution based on kernel density estimation, we design a method based on Equation (7) to derive the probability of a user u_i visiting a new location l_j given u_i's set of visited locations $L_i = \{l_1, l_2, \ldots, l_n\}$. First, we compute the distance of every pair of locations in L_i and l_j as follows:

$$d_{ij} = distance(l_i, l_j), \forall l_i \in L_i. \tag{10}$$

Each d_{ij} is then used to derive a probability based on Equation (7) as follows:

$$\hat{f}(d_{ij}) = \frac{1}{|D|h} \sum_{d' \in D} K\left(\frac{d_{ij} - d'}{h}\right). \tag{11}$$

Finally, the probability of u_i visiting a new location l_j can be obtained by taking the mean probability as follows:

$$p(l_j|L_i) = \frac{1}{n} \sum_{i=1}^{n} \hat{f}(d_{ij}). \tag{12}$$

Eventually, we can exploit *the personalized geographical influence of locations* to make location recommendation for u_i by returning the top-k locations l_j with the highest probability $p(l_j|L_i)$ according to Equation (12). However, we should integrate it with other information including user preference, social friendships, and geographical information of users to obtain better quality of location recommendation. To this end, we will describe a unified framework integrating all these factors together in Section 4.

4. A UNIFIED FRAMEWORK

In this section, we develop a unified framework integrating *user preference*, *social influence*, *the geographical influence of users*, and *the personalized geographical influence of locations* to enhance the quality of location recommendation. The framework incorporates two steps: (1) **Input-Fusion** for integrating *user preference*, *social influence*, and *the geographical influence of users*; and (2) **Output-Fusion** for integrating the output of the Input-Fusion step and *the personalized geographical influence of locations*.

Note that the geographical influences of users and locations are handled by two different steps, as they essentially differ from one another: the distance between the residences of users is related to users and indicates the correlation between users, whereas the distance between locations visited by a user is related to her check-in behavior and reflects the correlation between locations.

4.1 Input-Fusion

In the Input-Fusion step, the similarity measure between two users u_i and u_k is derived based on their social friendship and geographical residence distance. Formally, let $F(u_i)$ be a set of users having social friendships with u_i. If $u_k \in F(u_i)$, the similarity between users u_i and u_k is measured as follows [26]:

$$SGSim(u_i, u_k) = 1 - \frac{distance(u_i, u_k)}{\max\limits_{u_f \in F(u_i)} distance(u_i, u_f)}, \tag{13}$$

where $distance(u_i, u_k)$ returns the geographical distance between their residences. Otherwise, $SGSim(u_i, u_k) = 0$.

We then use the result of Equation (13) as an input for Equation (1) to predict the rating of a user u_i to a new location l_j (i.e., l_j has not been visited by u_i) using the following equation:

$$\hat{r}_{i,j} = \frac{\sum_{u_k \in F(u_i)} SGSim(u_i, u_k) \cdot r_{k,j}}{\sum_{u_k \in F(u_i)} SGSim(u_i, u_k)}. \tag{14}$$

Note that Equation (14) integrates user preference, social influence, and the geographical influence of the distance between two users' residences into the process of location recommendation.

4.2 Output-Fusion

In the Output-Fusion step, the goal is to further fuse the rating given by Equation (14) with the probability derived from *the personalized geographical influence of locations* (i.e., the personalized distribution of the geographical distance between every pair of locations visited by the same user) in Equation (12). First, the rating is transformed into a probability using

$$\hat{p}_{i,j} = \frac{\hat{r}_{i,j}}{\max_{l_j \in L - L_i}\{\hat{r}_{i,j}\}}, \tag{15}$$

where $\max_{l_j \in L - L_i}\{\hat{r}_{i,j}\}$ is a normalization term. Then, the normalized probability in Equation (15) is combined with the probability in Equation (12) into a unified measure through two popular rules: the product rule

$$\hat{s}_{i,j} = \hat{p}_{i,j} \cdot p(l_j|L_i), \tag{16}$$

or the sum rule

$$\hat{s}_{i,j} = \frac{\hat{p}_{i,j} + p(l_j|L_i)}{2}. \tag{17}$$

Note that in the sum rule, we assign the same weight for two probabilities instead of using weighting parameters that cost effort to find out their optimal settings and usually suffer from over-fitting.

5. EXPERIMENTAL EVALUATION

In this section, we describe our experiment settings for evaluating the performance of iGSLR against the state-of-the-art location recommendation techniques. Specifically, our evaluation focuses on three aspects of iGSLR: (1) the accuracy of iGSLR (i.e., precision and recall); (2) how well iGSLR deals with cold-start users who have only a few check-in POIs; and (3) how well iGSLR deals with the data sparsity problem.

5.1 Dataset Description

We use two publicly available large-scale real check-in data sets[1] that were crawled from Foursquare [6] and Gowalla [3],

[1] The large-scale real check-in data sets used for our experiments can be downloaded from `http://www.public.asu.edu/~hgao16/Publications.html` and `http://snap.stanford.edu/data/loc-gowalla.html`.

Table 1: Statistics of the two data sets

	Foursquare	Gowalla
Number of users	11,326	196,591
Number of locations (POIs)	182,968	1,280,969
Number of check-ins	1,385,223	6,442,890
Number of social links	47,164	950,327
User-location matrix density	2.3×10^{-4}	2.9×10^{-5}
Avg. No. of visited POIs per user	42.44	37.18
Avg. No. of check-ins per location	2.63	3.11

respectively. The statistics of the data sets are shown in Table 1. We split each data set into the training set and the testing set in terms of the check-in time rather than using a random partition method, because in practice we can only utilize the past check-in data to predict the future check-in events. Unless otherwise specified, the 90% of check-in data with earlier timestamp are used as the training set and the remaining check-in data are used as the testing set. In the experiments, the training set is used to learn the recommendation models of the evaluated techniques described in Section 5.2 to predict the testing data.

5.2 Evaluated Recommendation Techniques

The recommendation techniques implemented in our experiments are listed below.

- User-based CF (denoted by U) and Location-based CF (denoted by L): they only consider *user preference* for location recommendation (Section 2.2).
- Social CF (denoted by S): it integrates *user preference* with *social influence* into location recommendation (Section 2.3).
- Social & Geographical CF (denoted by SG): it integrates *user preference*, *social influence*, and *the geographical influence of users* in location recommendation based on Equation (14) [26].
- Output-Fusion using Equation (16) (**Default**) or (17), where $p(l_j|L_i)$ is implemented based on our kernel density estimation (iGSLR) (Section 3), power-law distribution (PD) [25], and multi-center Gaussian model (MGM) [2]. These three techniques fuse *user preference*, *social influence*, *the geographical influence of users*, and *the geographical influence of locations*.

Noth that: (1) in iGSLR the *the geographical influence of locations* is personalized, as presented in Section 3; (2) the model parameters of PD are obtained using maximum likelihood estimation; and (3) the centers of MGM are discovered by the mean-shift clustering algorithm [4].

5.3 Performance Metrics

In general, recommendation techniques compute a score for each candidate item (i.e., a location or POI in this paper) regarding a target user and return POIs with the **top-k** highest scores as a recommendation result to the target user. To evaluate the quality of location recommendation, it is important to find out how many locations actually visited by the target user in the testing data set are discovered by the recommendation technique. For this purpose, we employ two standard metrics: *precision* and *recall* [2, 25]:

- Precision defines the ratio of the number of discovered POIs to the k recommended POIs.
- Recall defines the ratio of the number of discovered POIs to the number of **positive POIs**, which have been visited by the target user in the testing set.

Precision and recall are averaged over all users to obtain the overall performance for various values of k from 5 to 50.

In addition, we evaluate how well these recommendation algorithms deal with cold-start users who have visited only a few locations in the training set. These techniques are also evaluated with smaller numbers of visited locations in the training set, called **given-n**, from 1 to 10. Note that precision and recall are averaged over the value of k, i.e., from 5 to 50.

6. EXPERIMENTAL RESULTS

This section analyzes our extensive experiment results. We first compare our iGSLR against the state-of-the-art geo-social CF techniques including PD [25] and MGM [2] in terms the overall performance (Section 6.1) and the performance on cold-start users (Section 6.2). We then study the effect of data sparsity (Section 6.3) and the number of positive locations (Section 6.4). Finally, we investigate two fusion rules: the product rule and the sum rule (Section 6.5).

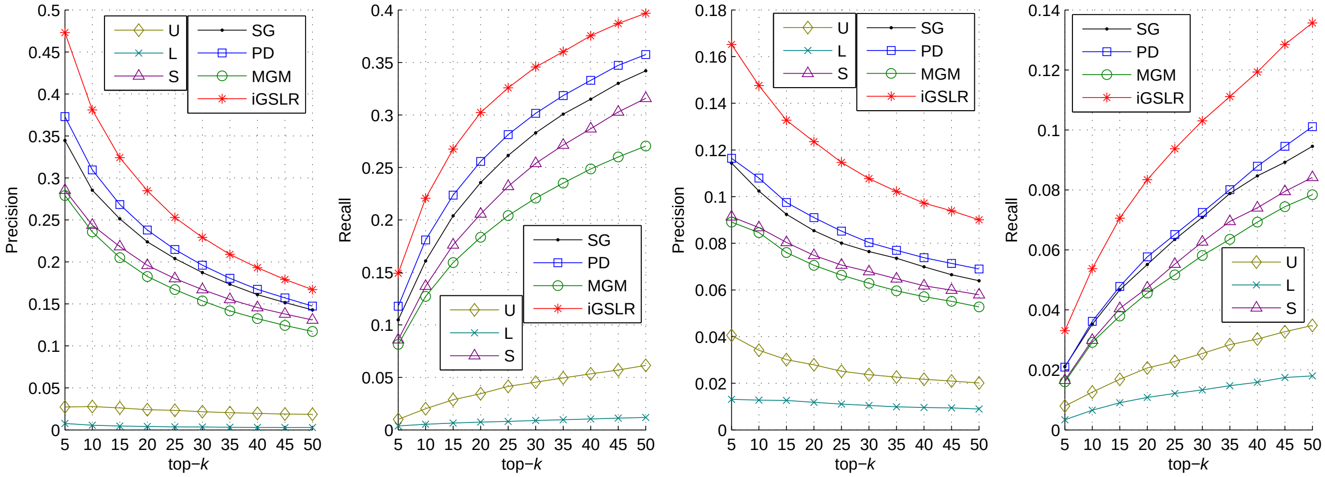
6.1 Comparison of Overall Performance

Figure 4 depicts the overall performance of a variety of recommendation techniques (listed in Section 5.2) on two large-scale real data sets collected from Foursquare and Gowalla. As a whole, **our proposed method iGSLR always exhibits the best accuracy** based on precision and recall for all the values of k, where k denotes the number of recommended locations. The details are demonstrated as follows.

U and L. Both U and L return the most *inaccurate* locations in terms of precision and miss almost all locations actually visited by target users in terms of recall in Figure 4. Such poor performance is caused by the lack of considering social and geographical influence and the sparsity problem of the large-scale real data set (i.e., its density is only 2.32×10^{-4} or 2.9×10^{-5}, as shown in Table 1).

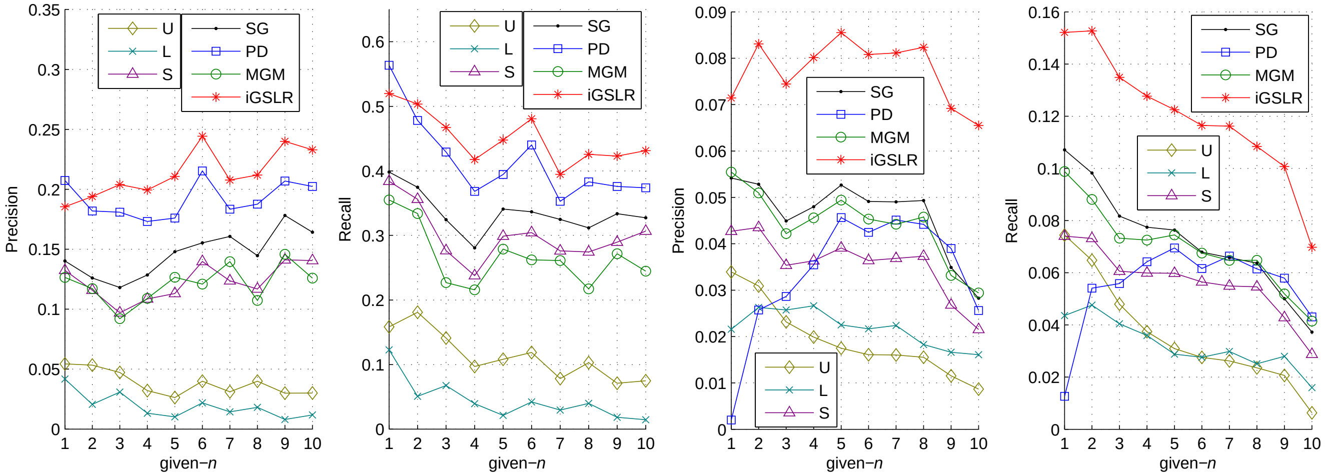
S and SG. (1) With the consideration of social friendships, S significantly enhances the quality of location recommendation in comparison to U and L in terms of precision and recall for all the values of k. These results show that social influence benefits location recommendation. (2) Furthermore, SG is superior to S by using the additional residence information of users to adjust the similarity between users with social friendships. The result shows the geographical influence of users also improves the quality of location recommendation, because people would like to visit locations with their nearby friends.

iGSLR, PD, and MGM. To enhance the quality of location recommendation, iGSLR, PD [25], and MGM [2] integrate *the geographical influence of locations* with SG using the Output-Fusion method via the product rule in Equation (16) (refer to Section 6.5 for the comparison between the product and sum rules), where $p(l_j|L_i)$ is obtained through our personalized distance distributions proposed in this paper, power-law distribution, and multi-center Gaussian model, respectively. (1) Compared to the baseline SG, iGSLR and PD improve the quality of recommendation by fusing the geographical influence of locations in terms of precision and recall. More im-



(a) Foursquare (Precision) (b) Foursquare (Recall) (c) Gowalla (Precision) (d) Gowalla (Recall)

Figure 4: Overall performance of various recommendation methods



(a) Foursquare (Precision) (b) Foursquare (Recall) (c) Gowalla (Precision) (d) Gowalla (Recall)

Figure 5: Performance of various recommendation methods on cold-start users

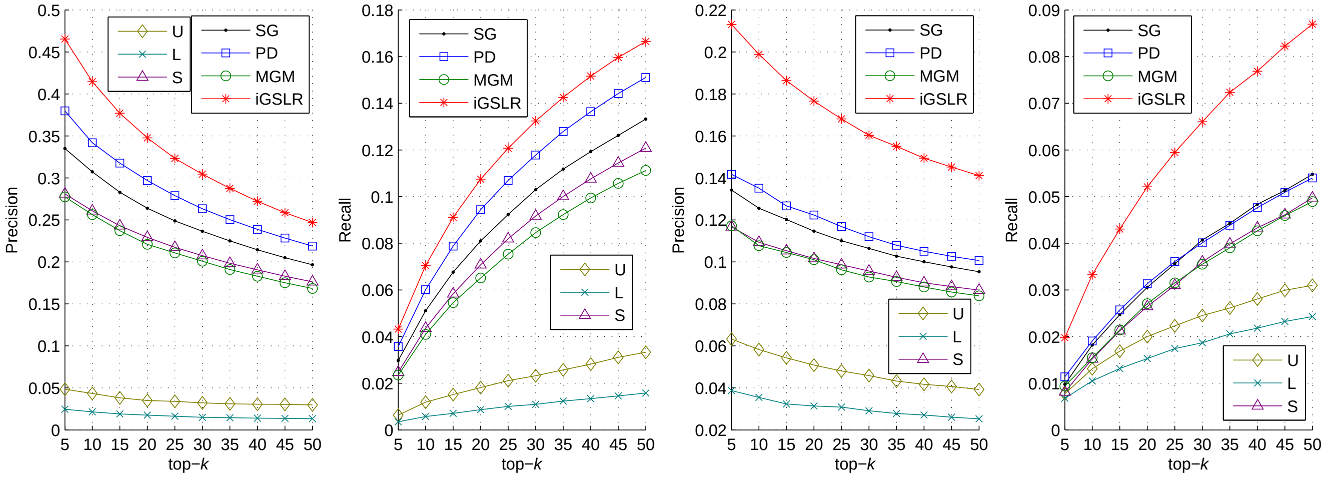
portantly, the improvement of iGSLR is significantly higher than that of PD on both data sets. (2) On the contrary, MGM obviously performs worse than SG, and even worse than S. The reason is that MGM not only models the geographical influence of locations as a universal distribution for all users but also considers the distance between a location and a center instead of between every pair of locations visited by the same user. As a result, the obtained distribution $p(l_j|L_i)$ of MGM is actually independent of user u_i, i.e., L_i, and hence it can be simplified as $p(l_j)$.

6.2 Performance on Cold-Start Users

Figure 5 depicts the performance of a variety of recommendation techniques on cold-start users based on two data sets, where n denotes the number of locations that a user has visited in the training set. (1) On the Foursquare data set, the performance of iGSLR is worse than PD only for users that have only one check-in record. This is because iGSLR cannot acquire a personal distance sample for these users, as it needs to compute the distance between every pair of locations visited by the same user. Fortunately, when users have checked in more than one location, iGSLR always performs better than PD in terms of precision and recall in Figures 5(a) and 5(b). (2) On the Gowalla data set with one-order-of-magnitude lower density, iGSLR still outperforms the other recommendation methods to a large extent, whereas PD deteriorates dramatically, even worse than SG and MGM. (3) Since it is important for LBSNs to provide good recommendation for cold-start users, iGSLR is better than other state-of-the-art geo-social recommendation techniques for LBSNs to attract new users.

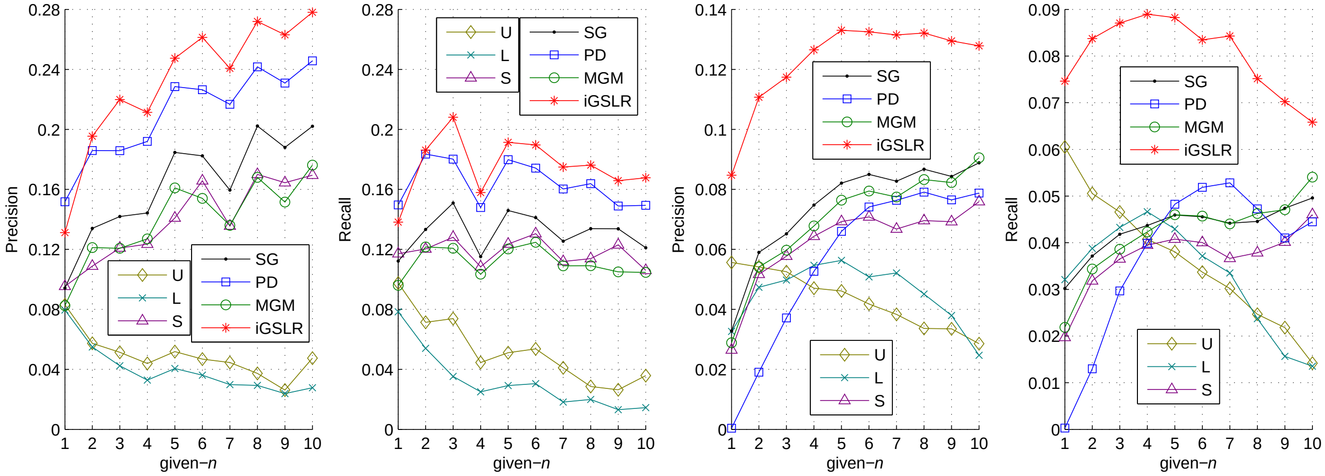
6.3 Effect of Data Sparsity

We further study how iGSLR deals with the data sparsity problem by using only 50% of the check-in data as the training sets. This experiment is in line with the previous work [25]. Figures 6 and 7 depict the overall performance



(a) Foursquare (Precision) (b) Foursquare (Recall) (c) Gowalla (Precision) (d) Gowalla (Recall)

Figure 6: Effect of data sparsity on overall performance with 50% of check-in data as the training set



(a) Foursquare (Precision) (b) Foursquare (Recall) (c) Gowalla (Precision) (d) Gowalla (Recall)

Figure 7: Effect of data sparsity on cold-start users with 50% of check-in data as the training set

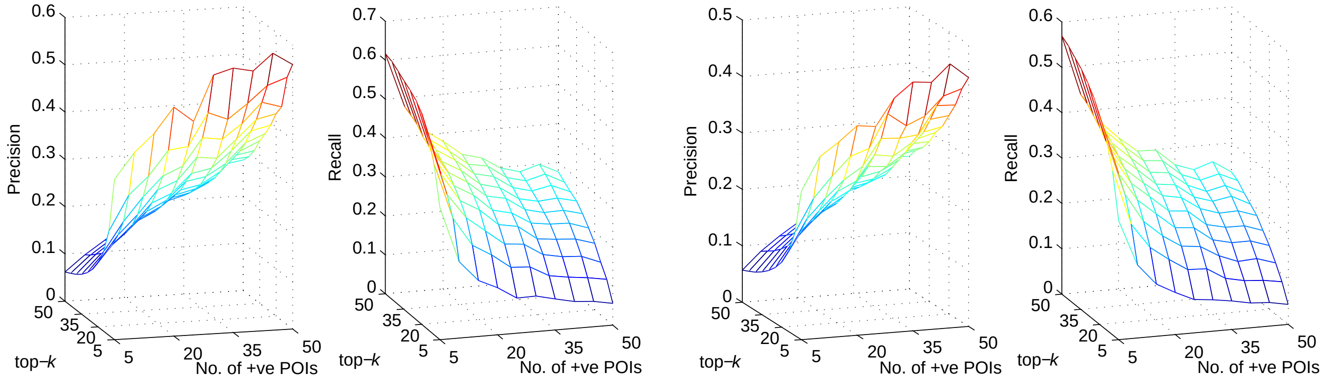
and the performance on cold-start users, respectively.

Compared to Figures 4 and 5, the recall of these evaluated recommendation techniques for both top-k and given-n deteriorates as the proportion of the training data decreases. This is because the density of the training set becomes sparser and the number of positive POIs of a user (i.e., the POIs visited by a user in the testing set) becomes relatively larger (note that this number is the denominator of recall). Interestingly, their precisions increases to some extent. Our explanation is that, with the decrease of the proportion of the training data, the lower density decrease the precision, but the larger number of positive POIs in the testing set increases the prior possibility of any recommended location being a positive POI and then contributes to the improvement on precision.

Most importantly, when confronting more severe problems of data sparsity, iGSLR shows much better overall precision and recall than the second best result given by PD [25] in Figure 6. Furthermore, iGSLR also gives better location recommendation for cold-start users (Figure 7). These results further verify the superiority of exploiting the *personalized* geographical influence of locations for location recommendation proposed in this paper over the *universalized* geographical influence of locations adopted by PD and MGM.

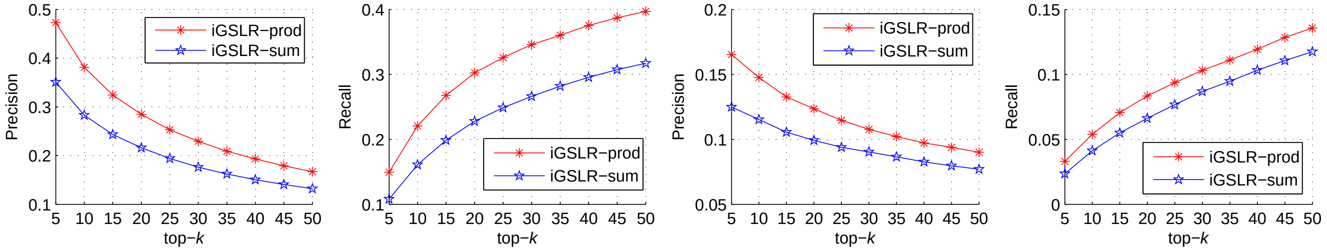
6.4 Effect of the Number of Positive Locations

As mentioned in Section 6.3, the number of positive (+ve) POIs affects the quality of location recommendation. We here discuss its effect on the performance of iGSLR and PD [25] only for the Foursquare data set due to space limitation, as shown in Figure 8. For both iGSLR and PD at the same value of k, their precision increases but their recall decreases as the number of the positive POIs gets larger. This is because the raise of the number of the positive POIs means that the recommendation techniques can more easily discover a location that a target user would like to visit



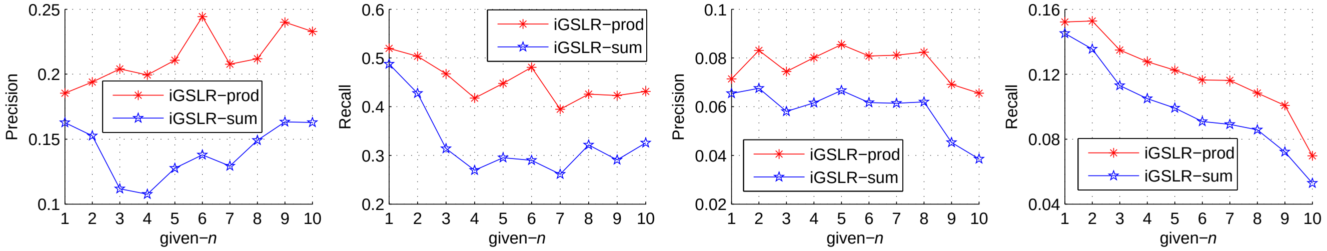
(a) iGSLR (b) PD

Figure 8: Effect of the number of positive (+ve) locations on iGSLR and PD



(a) Foursquare (Precision) (b) Foursquare (Recall) (c) Gowalla (Precision) (d) Gowalla (Recall)

Figure 9: Comparison of two fusion rules (product and sum) based on the overall performance



(a) Foursquare (Precision) (b) Foursquare (Recall) (c) Gowalla (Precision) (d) Gowalla (Recall)

Figure 10: Comparison of two fusion rules (product and sum) based on the performance on cold-start users

but it is hard to discover all of them. More significantly, iGSLR always shows better precision and recall than PD. For instance, at $k = 5$, the precision of iGSLR is higher than 33.9%, which means a target user can expect to be satisfied with one recommendation result out of three recommendation results generated by iGSLR. However, PD has to return four recommendation results to satisfy a user, because the precision of PD at $k = 5$ is 26.3%.

6.5 Comparison of Fusion Rules

Heretofore, our iGSLR always apply the product rule (Equation (16)) to fuse the personalized geographical influence of locations with other information. Here, we compare the performance of iGSLR using the product rule with that using the sum rule (Equation (17)), as shown in Figures 9 and 10. The product rule is consistently superior to the sum rule according to both top-k (overall performance) and given-n (performance on cold-start users) on two data sets. We argue that the probability $p(l_j|L_i)$ derived from the personalized geographical influence of locations is intrinsically different from the normalized probability $\hat{p}_{i,j}$ from the estimated rating $\hat{r}_{i,j}$. Thus, it is inappropriate to apply the sum rule to add two different things together; instead it is more reasonable to employ the product rule, where the probability $p(l_j|L_i)$ can be viewed as a weight to adjust the estimated rating $\hat{r}_{i,j}$.

7. CONCLUSION AND FUTURE WORK

In this paper, we have explored the geographical influence on users' check-in behaviors in location-based social networks (LBSNs). Aiming at overcoming the limitation that the state-of-the-art techniques merely consider the geographical influence as a *universalized* distance distribution for all users, we have proposed iGSLR to consider the geographical influence on a user's check-in behavior as a *personalized* distance distribution. iGSLR does not have any assumption about the form of the distance distribution required by previous work. Furthermore, we have integrated *user preference*, *social influence* and *the personalized geographical influence* into a unified location recommendation framework through two steps, namely, Input-Fusion and Output-Fusion. Finally, we have conducted extensive experiments to evaluate the performance of iGSLR using two large-scale real data sets collected from Foursquare and Gowalla. Experimental results show that iGSLR provides significantly superior location recommendation compared to all other recommendation techniques evaluated in our experiments.

We have three directions for future study: (1) how to recommend a trip of a series of points of interest (POIs), (2) how to incorporate the category information of POIs into our unified geo-social location recommendation framework, and (3) how to take temporal influence into account to capture the change of users' preferences.

8. REFERENCES

[1] J. Bao, Y. Zheng, and M. F. Mokbe. Location-based and preference-aware recommendation using sparse geo-social networking data. In *SIGSPATIAL*, 2012.

[2] C. Cheng, H. Yang, I. King, and M. R. Lyu. Fused matrix factorization with geographical and social influence in location-based social networks. In *AAAI*, 2012.

[3] E. Cho, S. A. Myers, and J. Leskovec. Friendship and mobility: User movement in location-based social networks. In *SIGKDD*, 2011.

[4] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE TPAMI*, 24(5):603–619, 2002.

[5] P. Cui, F. Wang, S. Liu, M. Ou, S. Yang, and L. Sun. Who should share what?: Item-level social influence prediction for users and posts ranking. In *SIGIR*, 2011.

[6] H. Gao, J. Tang, and H. Liu. gscorr: Modeling geo-social correlations for new check-ins on location-based social networks. In *CIKM*, 2012.

[7] R. Gemulla, E. Nijkamp, P. J. Haas, and Y. Sismanis. Large-scale matrix factorization with distributed stochastic gradient descent. In *SIGKDD*, 2011.

[8] A. Goyal and L. V. S. Lakshmanan. Recmax: Exploiting recommender systems for fun and profit. In *SIGKDD*, 2012.

[9] I. Guy, N. Zwerdling, D. Carmel, I. Ronen, E. Uziel, S. Yogev, and S. Ofek-Koifman. Personalized recommendation of social software items based on social relations. In *RecSys*, 2009.

[10] J. L. Herlocker, J. A. Konstan, A. Borchers, and J. Riedl. An algorithmic framework for performing collaborative filtering. In *SIGIR*, 1999.

[11] M. Jamali and M. Ester. *TrustWalker*: A random walk model for combining trust-based and item-based recommendation. In *SIGKDD*, 2009.

[12] I. Konstas, V. Stathopoulos, and J. M. Jose. On social networks and collaborative recommendation. In *SIGIR*, 2009.

[13] K. W.-T. Leung, D. L. Lee, and W.-C. Lee. Clr: A collaborative location recommendation framework based on co-clustering. In *SIGIR*, 2011.

[14] J. J. Levandoski, M. Sarwat, A. Eldawy, and M. F. Mokbel. Lars: A location-aware recommender system. In *ICDE*, 2012.

[15] H. Ma, I. King, and M. R. Lyu. Learning to recommend with social trust ensemble. In *SIGIR*, 2009.

[16] H. Ma, T. C. Zhou, M. R. Lyu, and I. King. Improving recommender systems by incorporating social contextual information. *ACM TOIS*, 29(2):9:1–9:23, 2011.

[17] B. Miller, J. Konstan, L. Terveen, and J. Riedl. Pocketlens: Toward a personal recommender system. *ACM TOIS*, 22(3):437–476, 2004.

[18] J. Riedl. Personalization and privacy. *IEEE Internet Computing*, 5(6):29–31, 2001.

[19] B. Sarwar, G. Kapypis, J. Konstan, and J. Riedl. Item-based collaborative filtering recommendation algorithms. In *WWW*, 2001.

[20] Y. Shen and R. Jin. Learning personal + social latent factor model for social recommendation. In *SIGKDD*, 2012.

[21] B. W. Silverman. *Density estimation for statistics and data analysis*. Chapman and Hall, London, 1986.

[22] S.-H. Yang, B. Long, A. Smola, N. Sadagopan, Z. Zheng, and H. Zha. Like like alike: Joint friendship and interest propagation in social networks. In *WWW*, 2011.

[23] X. Yang, H. Steck, and Y. Liu. Circle-based recommendation in online social networks. In *SIGKDD*, 2012.

[24] M. Ye, P. Yin, and W.-C. Lee. Location recommendation for location-based social networks. In *SIGSPATIAL*, 2010.

[25] M. Ye, P. Yin, W.-C. Lee, and D.-L. Lee. Exploiting geographical influence for collaborative point-of-interest recommendation. In *SIGIR*, 2011.

[26] J. J.-C. Ying, E. H.-C. Lu, W.-N. Kuo, and V. S. Tseng. Urban point-of-interest recommendation by mining user check-in behaviors. In *UrbComp*, 2012.

[27] V. W. Zheng, B. Cao, Y. Zheng, X. Xie, and Q. Yang. Collaborative filtering meets mobile recommendation: A user-centered approach. In *AAAI*, 2010.

[28] V. W. Zheng, Y. Zheng, X. Xie, and Q. Yang. Collaborative location and activity recommendations with gps history data. In *WWW*, 2010.

[29] V. W. Zheng, Y. Zheng, X. Xie, and Q. Yang. Towards mobile intelligence: Learning from gps history data for collaborative recommendation. *Artificial Intelligence*, 184-185(0):17–37, 2012.

[30] Y. Zheng, L. Zhang, Z. Ma, X. Xie, and W.-Y. Ma. Recommending friends and locations based on individual location history. *ACM TWEB*, 5(1):5:1–5:44, 2011.