Exploiting Geographical Influence for Collaborative Point-of-Interest Recommendation

Mao Ye[1†], Peifeng Yin[1†], Wang-Chien Lee[1†], and Dik-Lun Lee[2‡]

[†]Department of Computer Science and Engineering, The Pennsylvania State University, PA, USA.
[‡]Department of Computer Science and Engineering, HKUST, Hong Kong
[1]{mxy177,pzy102,wlee}@cse.psu.edu   [2]dlee@cse.ust.hk

ABSTRACT

In this paper, we aim to provide a point-of-interests (POI) recommendation service for the rapid growing location-based social networks (LBSNs), e.g., Foursquare, Whrrl, etc. Our idea is to explore user preference, social influence and geographical influence for POI recommendations. In addition to deriving user preference based on user-based collaborative filtering and exploring social influence from friends, we put a special emphasis on geographical influence due to the spatial clustering phenomenon exhibited in user check-in activities of LBSNs. We argue that the geographical influence among POIs plays an important role in user check-in behaviors and model it by power law distribution. Accordingly, we develop a collaborative recommendation algorithm based on geographical influence based on naive Bayesian. Furthermore, we propose a unified POI recommendation framework, which fuses user preference to a POI with social influence and geographical influence. Finally, we conduct a comprehensive performance evaluation over two large-scale datasets collected from Foursquare and Whrrl. Experimental results with these real datasets show that the unified collaborative recommendation approach significantly outperforms a wide spectrum of alternative recommendation approaches.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: Information Filtering; J.4 [**Computer Applications**]: Social and Behavior Sciences

General Terms

Algorithms, Experimentation.

Keywords

Collaborative Recommendation, Location-based Social Networks, Geographical Influence.

1. INTRODUCTION

With the rapid development of mobile devices, wireless networks and Web 2.0 technology, a number of location-based social networking services, e.g., Loopt[1], Brightkite[2], Foursquare[3] and Whrrl[4], have emerged in recent years.[5] These LBSNs allow users to establish cyber links to their friends or other users, and share tips and experiences of their visits to plentiful point-of-interests (POIs), e.g., restaurants, stores, cinema theaters, etc. In LBSNs, a *POI recommendation* service, aiming at recommending new POIs to users in order to help them explore new places and know their cities better, is an essential function that has received a lot of research momentum recently [25, 26].
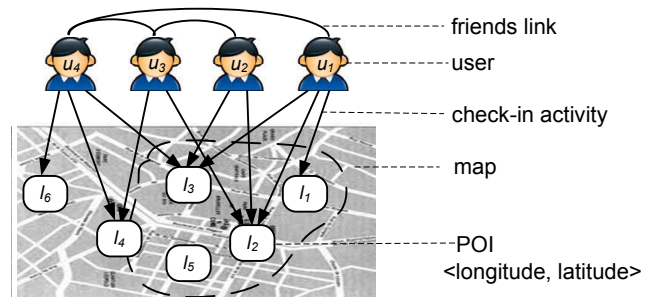

Figure 1: Graph representation of user-user friendship and user-location check-in activity in a LBSN

Indeed, facilitating POI recommendations in LBSNs is a promising and interesting research problem because valuable information such as the "cyber" connections among users as well as the "physical" interactions between users and locations have been captured in the systems. Nevertheless, these information have not been fully explored in prior research studies relevant to POI recommendations. For example, Zheng et. al. have extracted visited locations from GPS trajectory logs of mobile users for location recommendations [25, 26]. However, their studies consider neither the social links between users nor the interactions between users and locations in the recommendation process. In this paper, we aim to exploit the unique geographical implications em-

---

[1]www.loopt.com
[2]www.brightkite.com
[3]www.foursquare.com
[4]www.whrrl.com
[5]These services are often referred as location based social networks and thus abbreviated as LBSNs in the paper.

bedded in users' interactions with locations, in addition to applying the social influence from users' friends, for POI recommendations in LBSNs.

Users and POIs are two essential types of entities in LBSNs. As illustrated in Figure 1, users in an LBSN, denoted as u_1, u_2, u_3, u_4, are interconnected via social links to form a user social network. Moreover, POIs, denoted as $l_1, l_2, ..., l_6$, are connected with users via their "check-in" activities, which generally reflects the users' tastes on various POIs. Finally, as also logically illustrated in the figure, the POIs, geocoded by ⟨longitude, latitude⟩, are constrained geographically. To make recommendations of POIs to users, obviously the records of previous user check-in activities are very useful. With the availability of such information in LBSNs, an intuitive idea for supporting POI recommendations is to employ the conventional *collaborative filtering (CF)* techniques by treating POIs as the "items" in many successful CF-based recommender systems. The basic argument for this idea is that users' tastes can be deduced by other users who exhibit similar visiting behaviors to POIs in previous check-in activities. Thus, user-based or item-based collaborative filtering techniques may be applicable to POI recommendations. Additionally, the social network of users, which is handily available in the LBSN, can be explored to enhance performance of POI recommendations. Recent studies have argued that social friends tend to share common interests and thus can be used in the process of collaborative filtering for making recommendations [11, 12, 2, 13, 24].

While the ideas above aim to explore the essential information available in LBSNs, i.e., the user-location interactivities and user-user social links, we argue that the geographical influence naturally existing in the activities of users and their geographical proximities cannot be ignored. According to Tobler's First Law of Geography "Everything is related to everything else, but near things are more related than distant things" [19]. Thus, a user intuitively tends to visit nearby POIs. There are two major implications that can be derived from this intuition for POI recommendations: (1) people tend to visit POIs close to their homes or offices; and (2) people may be interested in exploring nearby POIs of a POI that they are in favor of, even if it is far away from home, e.g., a user may explore some restaurants and shops around Time Square when she goes there for a broadway show. Due to the geographical nature of the LBSNs, we believe the geographical influence between users and POIs as well as that amongst POIs are as important as the social influence amongst users, which as indicated earlier may play a positive role for supporting POI recommendations in LBSNs. In short, we are interested in studying the impact of geographical influence and social influence on POI recommendations in LBSNs.

Our approach for supporting POI recommendation service in LBSNs is to develop effective collaborative recommendation techniques that discover POIs of users' interests by incorporating the three complementary factors: i) user preference of POIs; ii) social influence; and iii) geographical influence. Notice that users' implicit preferences of POIs can be derived from their check-in activities on POIs. By considering two users who have checked into a lot of common POIs as similar users, we may discover the implicit preference of a user through the previous check-in activities of her similar users. Recall the example in Figure 1. Since u_1 and u_2 shares many commonly visited POIs, they may be con-

sidered as similar users who are assumed to share similar check-in behaviors, i.e., preference of POIs. As a result, l_1 is a good candidate for recommendation to user u_2 since u_1 has visited this POI before. On the other hand, social influence of friends can be incorporated in the recommendation process. For example, when considering l_4 as a recommendations candidate for u_1, the social influence of u_4 on u_1 may contribute to the decision making. Finally, the geographical influence of POIs on nearby POIs can be considered. As shown in the example, since u_2 has visited l_2 and l_3 before, their nearby POIs l_1 and l_5 may be considered positively due to the geographical influence.

As discussed earlier, the conventional item recommendation techniques based on user preference [9, 18, 14, 15, 13] and social influence [12, 13] seem to be applicable for POI recommendation. Nevertheless, their effectiveness on POI recommendations in LBSNs have not been studied. Most importantly, the idea of incorporating the geographical influence between POIs, which is refreshing and promising for POI recommendation, has not been investigated previously. In this paper, we examine the "geographical clustering phenomenon" of user check-in activities in LBSNs and propose a power-law probabilistic model to capture geographical influence among POIs. Accordingly, we realize the targeted collaborative POI recommendation service for LBSNs by incorporating the geographical influence of POIs via Bayesian theory. Finally, we propose a unified location recommendation framework to fuse user preference to POIs along with the social influence among users and the geographical influence among POIs.

In summary, the contributions we made in this research work are four-fold.

- We study the problem of supporting POI recommendation in location-based social networking systems (LBSNs), where POIs are uniquely different from other recommended items in conventional recommender systems because of the fact that "physical" interactions are required between users and POIs. Hence, we investigate the geographical influence between POIs and propose to incorporate geographical influence along with user preference and social influence in the collaborative recommendation techniques we develop for POI recommendations in LBSNs.

- We develop a novel idea to capture the geographical influence by investigating the geographical clustering phenomenon of user check-in activities in LBSNs. We propose to employ a power-law probabilistic model to capture the geographical influence among POIs, and realize our collaborative POI recommendations based on geographical influence via naive Bayesian method.

- We propose a unified recommendation framework for POI recommendations by fusing user preference, social influence and geographical influence to devise a check-in probability prediction model for a given user to visit a POI.

- Finally, we evaluate the proposed recommendation technique over large-scale datasets we collected from two well known LBSNs, i.e., Foursquare and Whrrl. Experimental results show that our proposed collaborative recommendation technique exhibits superior POI recommendation performance against other approaches. Important findings of our evaluation are summarized below.

– Geographical influence shows a more significant impact on the effectiveness of POI recommendations in LBSNs than social influence, improving the recommendation performance by at least 13.8% against social influence.

– Random Walk and Restart [12] may not be suitable for POI recommendation in LBSNs. Based on our analysis over the real data and the experimental results, we find that friends still reflect significantly different preferences and social tie cannot reflect the similarity of check-in behavior among users.

– Item-based collaborative filtering do not provide results comparable to user-based collaborative filtering, because many POIs, in the current state of LBSNs, only show a few user check-ins. Thus item similarity is not as accurate as user similarity.

The rest of this paper is organized as follows. In Section 2, we provide some background on conventional recommendation techniques according to user's own preference and social influence and review related works in the literature. In Section 3, we describe the location recommendation process according to geographical influence. In Section 4, we propose a location recommendation framework, which unifies all three factors together. In Section 5, we perform an empirical study on the different location recommendation algorithms upon two large scale datasets crawled from Foursquare and Whrrl, respectively. Finally, in Section 6, we conclude the paper.

2. PRELIMINARIES

In this section, we first provide background on user-based collaborative filtering and friend-based recommendation, which serve as the building blocks in our fusion approach to exploit user preference and social influence. Next we review some relevant studies in recommender systems.

2.1 User-based Collaborative Filtering

Based on collaborative filtering, users' implicit preference can be discovered by aggregating the behaviors of similar users. Let U and L denote the user set and the POI set in an LBSN, which keeps track of check-in activities in the system. The check-in activity a user $u_i \in U$ has at a POI $l_j \in L$ is denoted as $c_{i,j}$ where $c_{i,j} = 1$ represents u_i has a check-in at l_j before and $c_{i,j} = 0$ means there is no record of u_i visiting l_j. These recorded user check-in activities at POIs are thus used to discover a user's implicit preference of a POI, which can be represented as a probability to predict how likely the user would like to have a check-in at an unvisited POI. We denote this prediction by $\widehat{c}_{i,j}$ and obtain this predicted check-in probability of u_i to l_j as follows.

$$\widehat{c}_{i,j} = \frac{\sum_{u_k} w_{i,k} \cdot c_{k,j}}{\sum_{u_k} w_{i,k}} \qquad (1)$$

where $w_{i,k}$ is the similarity weight between users u_i and u_k.

To compute the similarity weights $w_{i,k}$ between users u_i and u_k, several similarity measures can be adopted, e.g., cosine similarity and Pearson correlation. In our study, we choose cosine similarity due to its simplicity. The cosine similarity weight between users u_i and u_k, denoted as $w_{i,k}^U$,

is defined as follows.

$$w_{i,k} = \frac{\sum_{l_j \in L} c_{i,j} c_{k,j}}{\sqrt{\sum_{l_j \in L} c_{i,j}^2} \sqrt{\sum_{l_j \in L} c_{k,j}^2}} \qquad (2)$$

2.2 Friend-based Collaborative Filtering

Friends tend to have similar behavior because they are friends and might share a lot of common interests, thus leading to correlated check-in behaviors [15, 13]. For example, two friends may hang out to see a movie together sometimes, or a user may go to a restaurant highly recommended by her friends. All those possible reasons suggest that friends might provide good recommendation for a given user due to their potential correlated check-in behavior. In other words, we can turn to user's friends for recommendation, and we call it recommendation based social influence from friends.

POI recommendations based on social influence can be realized by the friend-based collaborative filtering approach as described in [13].

$$\widehat{c}_{i,j} = \frac{\sum_{u_k \in F_i} SI_{k,i} \cdot c_{k,j}}{\sum_{u_k \in F_i} SI_{k,i}} \qquad (3)$$

where $\widehat{c}_{i,j}$ is the predicted check-in probability of u_i at l_j, F_i is the friends set of u_i, and $SI_{k,i}$ is *directional* social influence weight u_k has on u_i [14, 15, 13].

On the one hand, we think friends who have closer social tie may have better trust in terms of their recommendation; on the other hand, friends who show more similar check-in behavior should have more similar tastes with the active user, thus suggestions from those friends are more worthy. Thus, in the following, we introduce how to derive the social influence weight by combining the above two aspects.

One way to derive the social influence weight between two friends is based on both of their social connections and similarity of their check-in activities [12].

$$SI_{k,i} = \eta \cdot \frac{|F_k \cap F_i|}{|F_k \cup F_i|} + (1 - \eta) \cdot \frac{|L_k \cap L_i|}{|L_k \cup L_i|} \qquad (4)$$

where η is a tuning parameter ranging within $[0, 1]$, and F_k and L_k denote the *friend set* and *POI set* of user u_k, respectively.[6]

Another way of measurement is via the Random Walk with Restart (RWR) technique [20] over the graph that captures both the social connections among users as well as the check-in activities between users and POIs [24]. Starting from a node k, an RWR is performed by randomly following a link to another node at each step. Notice that there is a probability a in every step to restart at node k. By iterating RWR repeatedly until the whole process converges, a stationary (or steady-state) probability for each node can be obtained. The stationary probabilities of nodes give us a long-term visit rate for each user node (e.g., user u_i) given a bias towards a particular starting node (e.g. user u_k). This can be interpreted as the social influence weight user u_k have on u_i, i.e., $SI_{k,i}$.

2.3 Related Work

Content-based and collaborative filtering techniques are two widely adopted approaches for recommender systems [1]. A content-based system selects items for recommendation based on the similarity between item content (e.g., keywords/tags describing the items) and user profile [3, 8, 17].

---

[6]The friend set of a user refers to the socially connected friends of the user in the LBSN, while her POI set refers to the set of POIs she has check-in activities.

Since it mainly relies on dictionary-bound relations between the terms used in user profiles and item content, implicit associations between users are not considered.

The collaborative filtering systems are divided into two categories, i.e., memory-based and model-based. Memory-based systems can be further classified into *user-based* and *item-based* systems. For user-based systems [9], the similarity between all pairs of users is computed based on their ratings on associated items using some selected similarity measurement such as cosine similarity or Pearson correlation. Based on the user similarity, missing rating corresponding to a given user-item pair can be derived by computing a weighted combination of the ratings upon the same item from similar users. For item-based systems [18], instead of using similarity between users to predict missing rating, predications are made by finding similarly rated items first in order to compute a weighted combination of user ratings upon similar items. On the other hand, the model-based collaborative filtering systems assume that users may form clusters based on their similar behavior in rating items. A model can be learned based on patterns recognized in the rating behaviors of users using machine learning techniques such as clustering algorithms or Bayesian networks [5, 23].

Under the context of social networking systems, social friendship is shown to be beneficial for collaborative filtering based recommendation systems, e.g., memory-based [11, 12] and random walk based [2, 11, 12]. These works argue that social friends tend to share common interests and thus their relationships should be considered in the process of collaborative filtering. Random walk captures a social network as a graph with probabilistic weighted links to represent social relations and thus is able to accurately predict user preferences to items [12] and social influence to other users [24]. On the other hand, social friendship has also been explored in the model-based systems [14, 13]. These work mostly focus on conventional recommendation systems for recommending items such as movies.

Recently, location recommendation and mining has attracted a lot of attentions from the research community [27, 26, 7, 25, 22]. Among them, [27, 26, 7, 25] are mainly focused on GPS datasets which do not consider social relationships among users. In these works, unfortunately, the geographical influence among POIs are not explored [27, 26, 25]. Recently, the correlation of locations in GPS trajectories are explored [7]. In this work, however, locations are still treated as conventional items. As such, the correlations between locations are established through users' activities instead of their geographical influence. [22] is the first research to provide location recommendations services in LBSNs, but with the goal of improving efficiency of location recommendation.

Our study differentiates itself from all these prior works in four aspects: i) the application domain of location-based social networking systems, embracing both social and geographical features in the captured data, is new and unique; ii) the study of social influence and geographical influence in recommender systems for LBSNs is unexplored previously; iii) the proposal of unified collaborative recommendation approach, which incorporates geographical influence along with user preference and social influence, is new and innovative; iv) two large-scale real dataset collected from well known LBSNs, namely, Foursquare and Whrrl datasets, are adopted for performance evaluation.
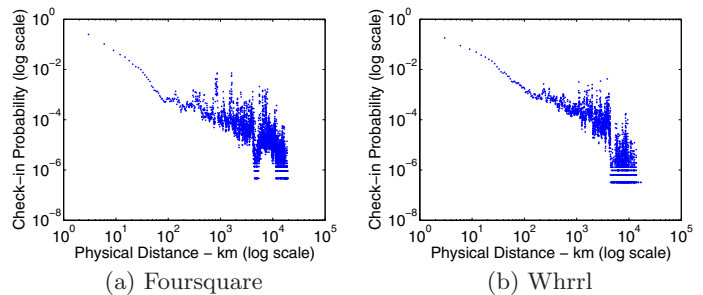

(a) Foursquare   (b) Whrrl

Figure 2: Geographical influence probability distribution

3. RECOMMENDATION VIA GEOGRAPHICAL INFLUENCE

As mentioned earlier, the check-in activities of users in LBSNs record their physical interactions (i.e., visits) at POIs, Thus, we argue that the geographical proximities of POIs have a significant influence on users' check-in behavior. To better understand this geographical influence on users, we perform a spatial analysis on real datasets of user check-in activities collected from two well known LBSNs, i.e., Foursquare and Whrrl. Specifically, we aim to study the implication of *distance* on user check-in behavior by measuring how likely two of a user's check-in POIs are within a given distance. To obtain this measurement, we calculate the distances between all pairs of POIs that a user has checked in and plot a histogram (actually probability density function) over the distance of POIs checked in by the same user. As shown in Figure 2, a significant percentage of POIs pairs checked in by the same user appears to be within short distance, indicating a *geographical clustering phenomenon* in user check-in activities.[7] This phenomenon may be attributed to the geographical influence which may be intuitively explained by the following tendencies: (1) people tend to visit POIs close to their homes or offices; and (2) people may be interested in exploring nearby POIs of a POI that they are in favor of, even if it is far away from home. As a result, the POIs visited by the same user tend to be clustered geographically. We believe that this geographical clustering phenomenon in user check-in activities can be exploited for POI recommendations in LBSNs. Thus, in the following, we study and model this geographical influence on user check-in behavior at POIs, aiming to utilize it in POI recommendations.

To achieve our goal, we would like to compute the likelihood that a user u_i would check in both POI l_j and l_k. Based on Figure 2, we intuitively think the check-in probability may follow the power-law distribution. Nevertheless, we observe that the check-in probability of POI pairs visited by the same person over distance is not a standard power-law distribution. Even though the left part of the figure decreases linearly (i.e., decreases exponentially in regular scale) and thus fits power-law distribution very well, the right part may sometimes deviate irregularly (i.e., the probability is high at some points). A reasonable explanation is that users may travel to different places and thus create multiple check-in spatial clusters. Generally speaking, the fact that a user's check-in POIs tend to be in a short distance is confirmed in our data analysis. As mentioned earlier, nearby POIs are more related to each other, which exhibits strong geographical influence. Moreover, the linear portion of the

---

[7]Note that the figure has been shown in log-log scale.

plot in Figure 2 covers the majority (90%) of the POI pairs. Thus, we propose to use power law distribution to model the check-in probability to the distance between two POIs visited by the same user as follows.

$$y = a \times x^b \tag{5}$$

where a and b are parameters of a power-law distribution, and x and y refer to the distance between two POIs visited by the same user and its check-in probability, respectively.

Equation (5) can transformed into Equation (6) in "log-log" scale to fit a linear model.

$$\log y = w_0 + w_1 \log x \tag{6}$$

Thus, the original power-law distribution can be recovered via the following equation.

$$a = 2^{w_0} \qquad b = w_1 \tag{7}$$

Hence, we can simply apply a linear curve fitting method to realize regression as follows. More specifically, let $y' = \log y$ and $x' = \log x$. We shall fit data as follows

$$y'(x', \mathbf{w}) = w_0 + w_1 \cdot x' \tag{8}$$

where w_0 and w_1 are the linear coefficients, collectively denoted by $\mathbf{w}$. In other words, the model can be learned in form of $\mathbf{w}$. In order to avoid over-fitting, we approach the weight coefficients by least square error method and add a penalty term (i.e., regularization term) to discourage the coefficients from reaching large values as below [6].

$$E(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^{N} \{y'(x'_n, \mathbf{w}) - t_n\}^2 + \frac{\lambda}{2} ||\mathbf{w}||^2 \tag{9}$$

where $E(\mathbf{w})$ denotes the loss function, N presents the cardinality of input dataset, t_n is the ground truth corresponding to x'_n, and λ is the regularization term.

Accordingly, the optimal values of a and b form the setting which minimizes the loss function $E(\mathbf{w})$ as below.

$$opt\{a, b\} = \arg\min_{a,b} E(\mathbf{w}) \tag{10}$$

In the following, we introduce a collaborative recommendation method based on the naive Bayesian method to realize POI recommendation in LBSNs. For a given user u_i and its visited POI set L_i, we define the probability that u_i has check-in activities at all locations in L_i by considering the pair-wise distances of POIs in L_i as follows.

$$Pr[L_i] = \prod_{l_m, l_n \in L_i \wedge m \neq n} Pr[d(l_m, l_n)] \tag{11}$$

where $d(l_m, l_n)$ denotes the distance between POIs l_m and l_n, and $Pr[d(l_m, l_n)] = a \times d(l_m, l_n)^b$ which follows the pow-law distribution model we obtained above. Note that here we assume the distances of POI pairs are independent.

Thus, for a given POI l_j (i.e., the recommendation candidate), user u_i, and her visited POI set L_i, we have the likelihood probability for u_i to check in l_j as follows.

$$\begin{aligned} Pr[l_j|L_i] &= \frac{Pr[l_j \cup L_i]}{Pr[L_i]} \\ &= \frac{Pr[L_i] \times \prod_{l_y \in L_i} Pr[d(l_j, l_y)]}{Pr[L_i]} \\ &= \prod_{l_y \in L_i} Pr[d(l_j, l_y)] \end{aligned} \tag{12}$$

To make a POI recommendation, we sort all the POIs in $L - L_i$ in accordance with their $Pr[l_j|L_i]$ ($l_j \in L - L_i$) to return the POI with the highest $Pr[l_j|L_i]$ to the user.

4. UNIFIED COLLABORATIVE POI RECOMMENDATION

In this section, we propose a unified framework to perform collaborative recommendation, which fuses ideas factors of user preference, social influence and geographical influence in POI recommendation. Notice that, different from predicting a POI's rating, we aim to return a ranked list of candidate POIs, which is very similar to conventional information retrieval [4].

4.1 Fusion Framework

As discussed, each factor, i.e., user preference, social influence or geographical influence, can be utilized to realize POI recommendation. Thus, we intuitively can implement three different recommender systems. We propose to use a linear fusion framework to integrate ranked lists provided by the three above-mentioned recommenders into the final ranked list [4, 21]. By integrating multiple recommenders, top-ranked POIs from each of the recommendation algorithms could increase both recall (due to the different highly ranked POIs) and precision (giving that the recommender systems have a high density of user-preferred POIs on top of the results lists.

Let $S_{i,j}$ denote the check-in probability score of user u_i at POI l_j, i.e., the more likely u_i has a check-in activity at l_j, the larger $S_{i,j}$ is. Let $S_{i,j}^u$, $S_{i,j}^s$ and $S_{i,j}^g$ denote the check-in probability scores of user u_i at POI l_j, corresponding to recommender systems based on user preference, social influence and geographical influence, respectively. We have $S_{i,j}$ as follows.

$$S_{i,j} = (1 - \alpha - \beta)S_{i,j}^u + \alpha S_{i,j}^s + \beta S_{i,j}^g \tag{13}$$

where the two weighting parameters α and β ($0 \leq \alpha + \beta \leq 1$) denote the relative importance of social influence and geographical influence comparing to user preference. Here $\alpha = 1$ states that $S_{i,j}$ depends completely on the prediction based on social influence; $\beta = 1$ states that $S_{i,j}$ depends completely on the predication based on geographical influence; while $\alpha = \beta = 0$ states that $S_{i,j}$ counts only on user preference.

4.2 Check-in Probability Score Estimation

According to the above fusion framework, in order to estimate the check-in probability score $S_{i,j}$, we need to predict the check-in probability score of $S_{i,j}^u$, $S_{i,j}^s$ and $S_{i,j}^g$ corresponding to user preference, social influence and geographical influence, respectively. Accordingly, we estimate the check-in probability $p_{i,j}^u$, $p_{i,j}^s$ and $p_{i,j}^g$ for a user u_i to visit a POI l_j in order to obtain $S_{i,j}^u$, $S_{i,j}^s$ and $S_{i,j}^g$, respectively.

First, the predication of $p_{i,j}^u$ can be estimated based on the idea of user-based collaborative filtering as discussed before. More specifically, we utilize the behavior of similar users to realize the predication as Equation (1). Thus we have

$$p_{i,j}^u = \frac{\sum_{u_k} w_{i,k} \cdot c_{k,j}}{\sum_{u_k} w_{i,k}} \tag{14}$$

where $w_{i,k}$ is the similarity weight between users u_i and u_k.

Similarly, the prediction of $p_{i,j}^s$ can be estimated based on based on the idea of friend-based collaborative filtering. Thus, according to Equation (3), we have

$$p_{i,j}^s = \frac{\sum_{u_k \in F_i} SI_{k,i} \cdot c_{k,j}}{\sum_{u_k \in F_i} SI_{k,i}} \tag{15}$$

where F_i is the friends set of u_i, $SI_{k,i}$ is the weight measuring social influence from u_k to u_i.

Finally, $p_{i,j}^g$ can be directly obtained from Equation (12)

$$p_{i,j}^g = Pr[l_j|L_i] = \prod_{l_y \in L_i} Pr[d(l_j, l_y)] \qquad (16)$$

where L_i is the visited POI set of u_i, and $d(l_j, l_y)$ denotes the distance between POIs l_j and l_y.

After we get the check-in probability estimation, we obtain the corresponding scores as follows.

$$S_{i,j}^u = \frac{p_{i,j}^u}{Z_i^u}, \text{ where } Z_i^u = \max_{l_j \in L - L_i}\{p_{i,j}^u\}$$

$$S_{i,j}^s = \frac{p_{i,j}^s}{Z_i^s}, \text{ where } Z_i^s = \max_{l_j \in L - L_i}\{p_{i,j}^s\} \qquad (17)$$

$$S_{i,j}^g = \frac{p_{i,j}^g}{Z_i^g}, \text{ where } Z_i^g = \max_{l_j \in L - L_i}\{p_{i,j}^g\}$$

where Z_i^u, Z_i^s and Z_i^g are normalization terms.

5. EMPIRICAL EVALUATION

In this section, we design and conduct several experiments to compare the recommendation qualities of the proposed collaborative recommendation algorithms with some state-of-the-art recommendation techniques, including collaborative filtering and random walk with restart, and to investigate several interesting questions. Specifically, the design of the experiments aims to achieve the following goals. (1) As our proposed method factors in user preference, social influence from friends and geographical influence from nearby location, we intent to study parameters α and β to understand the roles/weights of the above-mentioned factors in obtaining optimal recommendations. (2) We intend to validate our ideas by comparing the effectiveness of the proposed approach with other state-of-the-art techniques. (3) Due to the growing research interests in social influence from friends, we intend to further study the similarity of check-in behaviors in terms of the strength of "social ties" between two friends. (4) In our proposal, user-based collaborative filtering approach has been employed to discover user preference. We intend to explore the feasibility and necessity of integrating item-based collaborative filtering approach to further enhance the recommendation quality. (5) We would like to understand how data sparsity may affect POI recommendations in LBSNs. (6) How well our techniques deal with cold start users, who do not have many check-in records for discovery of their interests [10].

5.1 Dataset Description

We crawled the websites of Foursquare and Whrrl, two of the most representative LBSNs, for a month to collect two datasets consisting of 153,577 users and 96,229 POIs in Foursquare, and 5,892 users and 53,432 POIs in Whrrl, respectively. Our performance evaluation is conducted on these two large-scale real datasets. After summarizing the check-in records, we get the user-POI check-in matrix densities as 4.24×10^{-5} for Foursquare dataset and 2.72×10^{-4} for Whrrl datasets, respectively. Note that, the effectiveness of recommendation service for sparse dataset (i.e., low density user-POI check-in matrix) is usually not high due to the limited information provided by the dataset. For example, the reported precision in [12] is 0.17 over a pre-prossed dataset with 7.8×10^{-4} density. Thus, in our experiments, we focus on observing the relative performance of algorithms instead of their absolute effectiveness measures, which we expect to improve as the number of LBSN users continues to grow and

more check-in activities are logged. To facilitate our evaluation, for each individual user in the datasets, we randomly mark off $x\%$ ($x = 10, 30, 50$ (with 30 as the default value) of all POIs visited by the user. In the experiments, the evaluated POI recommendation algorithms are used to recover the missing user-POI pairs that have been marked off.

5.2 Performance Metrics

A POI recommendation algorithm under evaluation computes a ranking score for each candidate POI (i.e., POI that user has not visited) and returns the top- N highest ranked POIs as recommendations to a targeted user. To evaluate the prediction accuracy, we are interested in finding out how many POIs previously marked off in the preprocessing step recovered in the returned POI recommendations. More specifically, we examine two metrics: (1) the ratio of recovered POIs to the N recommended POIs, and (2) the ratio of recovered POIs to the set of POIs deleted in preprocessing. The former is *precision@N* while the latter is *recall@N*, and collectively referred as *performance@N*. In our experiment, we test the performance when $N = 5, 10, 20$ with 5 as the default value.

5.3 Evaluated Recommendation Approaches

Three factors, namely user preference (U), social influence from friends (S) and geographical influence from POIs (G), are incorporated in our unified collaborative recommendation algorithm, denoted by `USG` in our evaluation. A number of state-of-the-art and new collaborative filtering approaches, some of which can be configured by controlling the weight parameters, $0 < \alpha, \beta < 1$, in `USG`, are also evaluated for comparison. In addition of `USG`, the recommendation approaches under evaluation are listed below.

- `user-based CF (denoted by U)` - this is a special case of `USG` by setting both α and β as zeros. In other words, only user preference is considered for recommendation.
- `friend-based CF (denoted by S)` - this is also a special case of `USG`, where $\alpha = 1$. Here, only friends of the targeted user are used in making a specific recommendation. As introduced before, there are two alternative methods to derive the social influence weight between friends. One is to compute the social influence weight based on friends based on Equation(4) [12] and the other is to derive social influence weight between friends using Random Walk and Restart technique [24]. To differentiate these two approaches, we denote them as `S` and `S`$_{rwr}$, respectively.
- `GI-based recommendation (denoted by G)` - this approach, considering only the factor of geographical influence, is a special case of `USG` where $\beta = 1$.
- `Random Walk with Restart (denoted by RWR)` - this is a state-of-the-art algorithm recently developed for collaborative item recommendation based on social networks [12]. Users' preferences to items are predicted by Random Walk and Restart over a graph capturing social graph and user-item matrix.
- `User preference/social influence based recommendation (denoted by US)` - this method, considering both user preference and social influence from friends, is a special case of `USG`, where $0 < \alpha < 1$ and $\beta = 0$.
- `User preference/geographical influence based recommendation (denoted by UG)` - this approach, considering both user preference and geographical influ-
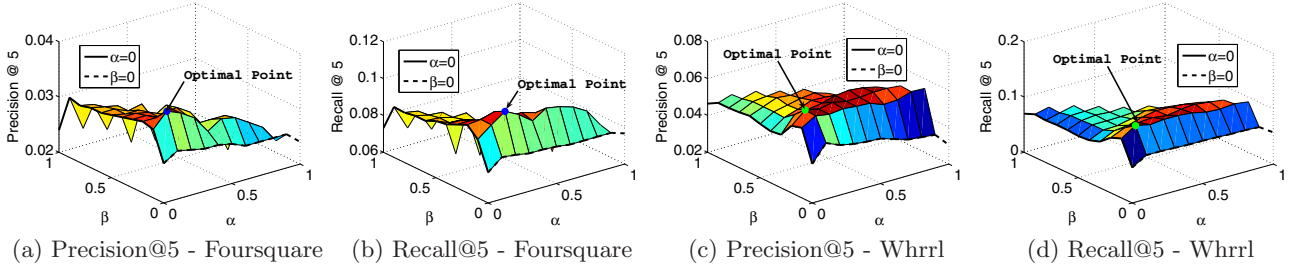
(a) Precision@5 - Foursquare (b) Recall@5 - Foursquare (c) Precision@5 - Whrrl (d) Recall@5 - Whrrl

Figure 3: Tuning parameters

(a) Precision@N - Foursquare (b) Recall@N - Foursquare (c) Precision@N - Whrrl (d) Recall@N - Whrrl
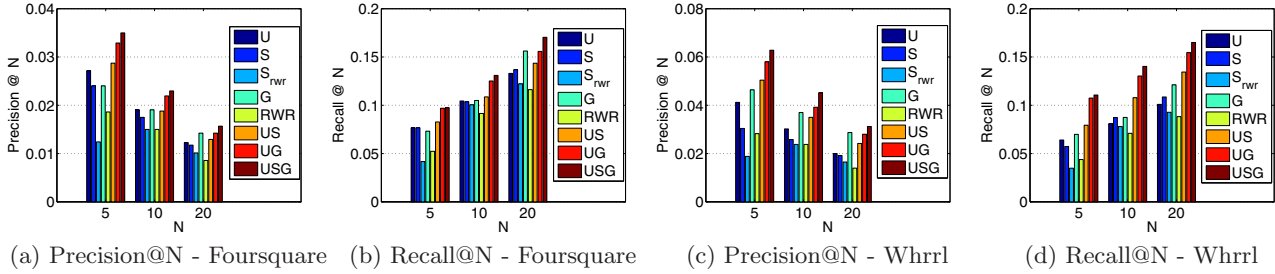
Figure 4: Performance comparison

ence, is a special case of USG, where $0 < \beta < 1$ and $\alpha = 0$.

5.4   Tuning Parameters

As mentioned, two parameters α (for social influence factor) and β (for geographical influence factor) can be controlled to tune the performance of USG and to configure it into other recommendation approaches for evaluation. Here we vary them in USG to understand the roles of user preference, social influence from friends and geographical influence from POIs played in achieving the optimal USG performance. Similarly, we tune α in US and β in UG to findout their optimal settings as well. Figure 3 shows the performance@5 results of USG under different α and β settings, where the best parameter settings are indicated in the figures. The optimal settings for US and UG can also be observed in the figures, i.e., dashed line for US and solid line for UG. Those optimal parameter settings are also summarized in Table 1.

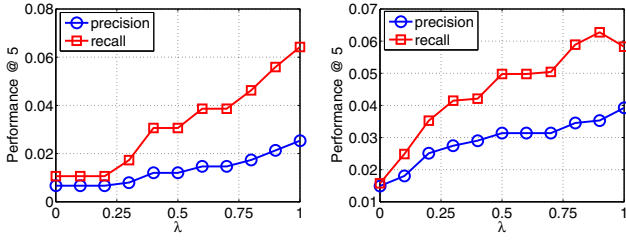|		Precision@5		Recall@5	
		α	β	α	β
Foursquare	US	0.1	—	0.1	—
	UG	—	0.2	—	0.2
	USG	0.1	0.1	0.2	0.1
Whrrl	US	0.1	—	0.1	—
	UG	—	0.1	—	0.1
	USG	0.1	0.2	0.1	0.1

Table 1: Optimal parameter settings

Through this study, we can easily observe that user preference plays a dominate role in contributing to the optimal recommendation, while both social influence and geographical influence are innegligible. More specifically, as shown in Table 1, the factor of user preference contributes at least 70% in making the best recommendation, while both social influence and geographical influence contribute at least 10%.

5.5   Performance Comparison

Next, we compare the effectiveness of the recommendation approaches under evaluation. Figure 4 shows the performance@N ($N = 5, 10, 20$) of all approaches in terms of their best performance (i.e., the performance under the optimal parameter settings). The experiments used both Foursquare and Whrrl datasets. The precision and recall for them are plotted in Figure 4(a) and Figure 4(b), and Figure 4(c) and Figure 4(d), respectively. In these figures, USG always exhibits the best performance in terms of precision and recall under all values of Ns, showing the strength of combines all three factors of user preference, social influence and geographical influence. Notice that both of our real datasets (i.e., Foursquare and Whrrl) have low density. According to the empirical study in [12], the reported precision is about 0.17 over a pre-processed dataset with 7.8×10^{-4} density of user-item matrix. Thus, the measured low precision over our datasets (which are not preprocessed) is reasonable. Most importantly, USG outperforms the baseline approach U (i.e., user-based CF) by about 50% percentage of performance improvement in both datasets.

Between the two alternative social influence measurement methods (i.e., S and S_{rwr}) for friend-based CF, we find S to have much better performance than S_{rwr}. Moreover, RWR shows poor performance for POI recommendation in these experiments. This raises a very interesting question of whether Random Walk and Restart technique is suitable for POI recommendations. In a later section, we shall answer this question by analyzing the correlation between (i) the similarity of check-in behavior among friends and (ii) social ties among friends. For the rest of the experiments, we use S as the component of social influence from friends in US and USG.

Figure 4 also indicates that both social influence and geographical influence can be utilized to perform POI recommendation. As shown, both S and G provide comparable results against U. Notice that, in LBSNs, since the check-
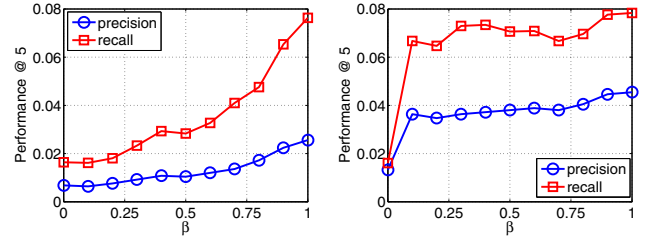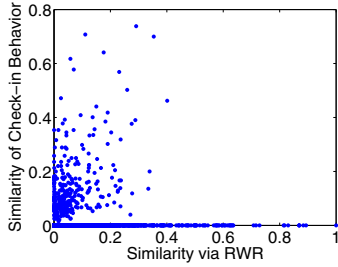
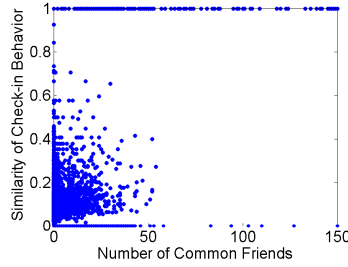(a) Performance@5 - Foursquare (b) Performance@5 - Whrrl

Figure 5: Fusion of U and L

(a) Performance@5 - Foursquare (b) Performance@5 - Whrrl

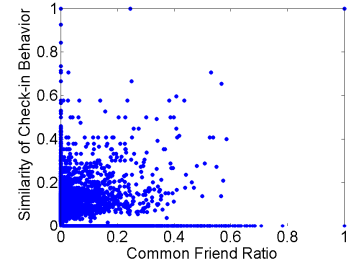**Figure 6: Fusion of L and G**
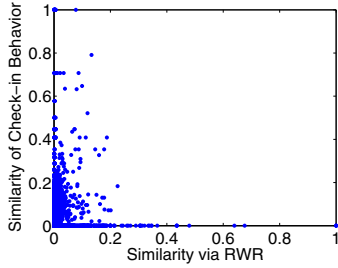

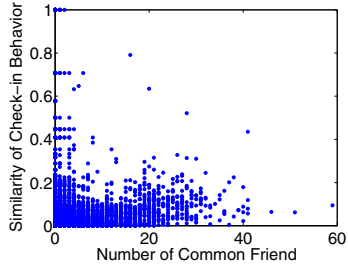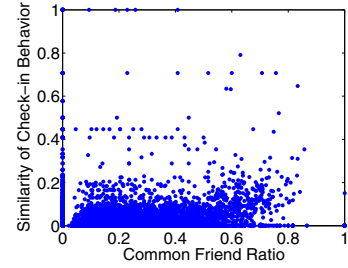(a) Similarity computed via RWR - (b) Number of common friend - (c) Common friend ratio - Foursquare
Foursquare   Foursquare

(d) Similarity computed via RWR - (e) Number of common friend - (f) Common friend ratio - Whrrl
Whrrl   Whrrl

Figure 7: Social tie and its influence implication

in activities involve physical interaction between users and POIs, geographical influence matters a lot, which is confirmed in the study. As shown, G usually outperforms S and sometimes even performs better than U, e.g., when $N = 20$. Also, UG always show better performance than US. This is due to the spatial clustering phenomenon appearing in user check-in activities. Thus, when N is relatively large, there is very good chance to discover most of user's check-in activities based on social influence.

In both Foursquare and Whrrl datasets, we find that when more factors are considered the performance turns out to be better. For example, US is better than U and S, UG is better than U and G, and USG shows the best performance.

5.6 Study on Item-based CF

In addition to user-based CF, item-based CF can also estimate a user's preference to an item, by exploring the similarity between items instead of users [18]. In [21], a CF technique has been proposed to fuse both user-based and item-based similarity to overcome the data sparsity problem [10]. Thus, a potential idea for POI recommendations is to employ the item-based CF (denoted by L). Additionally, geographical influence, which models the influence among POIs, may be seemingly similar to "item similarity" in item-based CF. However, we would like to point out that they are conceptually different and thus should not be mistaken.

In this section, we explore the idea of further incorporating L into our framework by examining whether fusing L with U and G respectively into new approaches denoted by UL and GL would outperform U and G alone.

Similar to [21], we introduce a weighting parameter λ in UL. When $\lambda = 1$, UL is reduced to U; and when $\lambda = 0$, L is obtained. Similarly, we introduce a weighting parameter γ in GL. Figure 5 and Figure 6 show the performance of UL and GL on Foursquare and Whrrl datasets under various settings of λ and γ. Surprisingly, these figures show that L brings no advantage at all in enhancing U or L in POI recommendations, indicating item-based CF is not an effective approach in our application. Our explanation is that, at the current stage, POIs in LBSNs may not have been visited by sufficient many users to make item-based CF work well. In other words, the computed similarity between two POIs may not provide a good clue to decide whether a user likes a POI or not. Since U or G alone show much better performance than L, we don't integrate L in our recommendation framework.

5.7 Study on Social Influence

As shown earlier, Random Walk with Restart [12] does not perform well for POI recommendations. To obtain a comprehensive understanding of the reasons behind, we analyze the correlation between the similarity of user check-in behaviors and the user similarity calculated based on Random
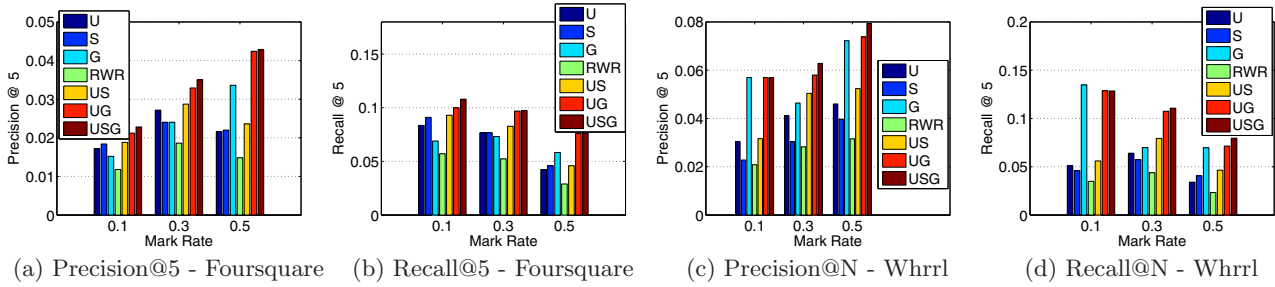
(a) Precision@5 - Foursquare   (b) Recall@5 - Foursquare   (c) Precision@N - Whrrl   (d) Recall@N - Whrrl

Figure 8: Impact of Data Sparsity

(a) Precision@5 - Foursquare   (b) Recall@5 - Foursquare   (c) Precision@5 - Whrrl   (d) Recall@5 - Whrrl
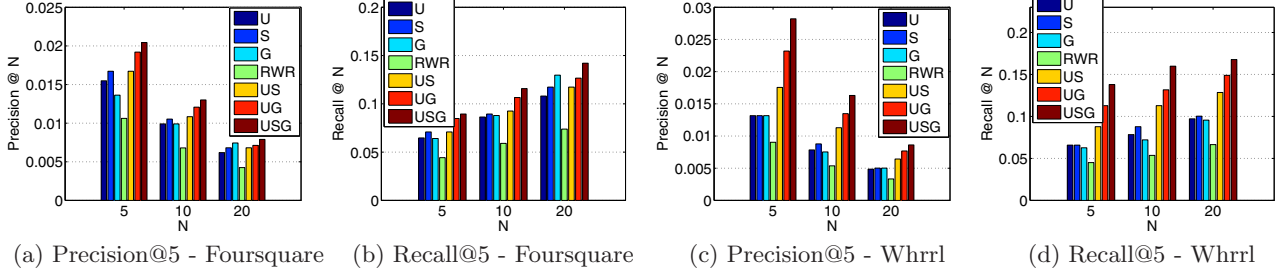
Figure 9: Performance for cold start users

Walk and Restart. Note that, based on [12], user similarity can be derived from the social graph matrix and user-POI check-in matrix. Figure 7(a) and Figure 7(d) show the plots on Foursquare and Whrrl datasets under the best RWR settings. Both figures show that similar users do not necessarily have high similarity in their check-in behaviors. For example, user pairs with similarity larger than 0.1 usually share nothing in their check-in behavior in both Foursquare and Whrrl datasets. The results indicate that the tastes of a user's friends may actually vary significantly, which has also be discussed in [16] recently. To further verify this finding, we also examine the correlation between the similarity of check-in behaviors between two friends and the strength of their social ties. In our tests, the social tie is defined in two forms: 1) number of common friends (see Figure 7(b) and Figure 7(e) for experimental results) and common friend ratio (see Figure 7(c) and Figure 7(f) for experimental results), where common friend ratio is measured by Jaccard coefficient. For friends who have very strong social tie (i.e., larger number of common friends or larger common friend ratio), we again find their check-in behaviors are not necessarily similar as shown in the figures.

From the above observations, we conclude that friends have different tastes. The similarity in friends' check-in behaviors may not necessarily be reflected in terms of the strength of their social ties. As a matter of fact, in measuring the social influence between friends, we find the factor of check-in behavior to be more important than the factor of social tie. Through our experiments on the S algorithm, we find the optimal setting for η in Equation (4) to be smaller than 0.05 in both Foursquare and Whrrl datasets, which indicates the factor of check-in behavior weighs more than the factor of social tie. While RWR treats both factors equally, thus degrading the performance in our Foursquare and Whrrl datasets.

5.8 Impact of Data Sparsity

Here, we study how USG deals with the data sparsity problem. In order to produce user-POI check-in matrix with different sparsity, we mark off x% = 10%, 30% and 50%

of user's check-in activity records from the original check-in datasets for three groups of tests as shown in Figure 8. The larger the mark-off ratio x is, the sparser the user-POI check-in matrix is. As shown, USG always exhibits the best performance@5 under all mark-off ratios. Particularly, when the data is very sparse, e.g., x% = 50%, geographical influence plays an extremely important role in recommending interesting POIs to users. The reason is that both users and their social friends have relatively small check-in logs. Thus, the similarity weight or social influence score derived from such sparse data may be misleading. On the other hand, geographical influence, reflecting a global behavior affected by geography, fits the behaviors of most users in LBSNs. Thus, the approaches incorporating geographical influence factor, i.e., G, UG and USG, show great strengths under various data sparsity scenarios.

5.9 Test for Cold Start Users

Finally, we test the performance of POI recommendations for cold start users. Here, we consider those users who have less than 5 check-in activities in the user-POI check-in matrix after removing 30% check-ins as cold start users. As shown in Figure 9, in all cases we tested, USG always shows the best performance. Note that in POI recommendations for cold start users, user preference is hard to capture as POIs visited by this user are few. Consequently, U shows the worst performance as it only considers user preference. G, which explores the spatial clusters of user check-in activities, is also affected. On the other hand, S overcomes the lack of user's check-ins as social friends may supply many useful check-ins, potentially useful for POI recommendations. Thus, in this experiment, we find that the recommendation performance of S usually works better than U and G do. Notice that, we find the performance of G to be better than S in extremely sparse scenario in Figure 8 because in that scenario, social friends' check-in records are very limited as well. Thus, geographical influence prevails due to its applicability to most of the people. However, it is noteworthy that all three factors are very important for the POI recommendations to cold start users, as USG is always the best.

6. CONCLUSIONS AND FUTURE WORK

This research attempts to facilitate a POI recommendation service in location-based social networks. Our idea is to incorporate user preference, social influence and geographical influence in the recommendation. In addition to deriving user preference by user-based collaborative filtering and capturing social influence from friends, we model the geographical influence among POIs by employing power law distribution to uncover the spatial clustering phenomenon in user check-in activities. Furthermore, we propose a unified POI recommendation framework, which fuses user preference to a POI with social influence and geographical influence. We conduct a comprehensive performance evaluation over two large-scale real datasets collected from Foursquare and Whrrl. Experimental results show that the unified collaborative recommendation technique is superior to all other recommendation approaches evaluated. Additional findings have been uncovered through analysis of the experimental results, including 1) geographical influence shows a more significant impact on the effectiveness of POI recommendations than social influence; 2) Random Walk and Restart may not be suitable for POI recommendation in LBSNs, because friends exhibit significantly different preferences (i.e., the strength of social ties do not reflect the similarity of check-in behavior among users in LBSNs; 3) Item-base CF is not an effective approach in our application due to insufficient number of visitors to many locations at the current state of LBSNs.

The semantic tags of POIs contain very rich information brought in by LBSN users. As for the next step, we plan to incorporate the semantic tags of POIs to further improve the unified POI recommendation framework we proposed in this paper.

7. REFERENCES

[1] G. Adomavicius and A. Tuzhilin. Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. *IEEE TKDE*, 17(6):734–749, 2005.

[2] R. Andersen, C. Borgs, J. T. Chayes, U. Feige, A. D. Flaxman, A. Kalai, V. S. Mirrokni, and M. Tennenholtz. Trust-based recommendation systems: an axiomatic approach. In *WWW*, pages 199–208, 2008.

[3] M. Balabanovic and Y. Shoham. Content-based collaborative recommendation. *CACM*, 40(3):66–72, 1997.

[4] B. T. Bartell, G. W. Cottrell, and R. K. Belew. Automatic combination of multiple ranked retrieval systems. In *SIGIR*, pages 173–181, 1994.

[5] D. Billsus and M. J. Pazzani. Learning collaborative information filters. In *ICML*, pages 46–54, 1998.

[6] C. M. Bishop. *Pattern Recognition and Machine Learning*. 2006.

[7] X. Cao, G. Cong, and C. S. Jensen. Mining significant semantic locations from gps data. *PVLDB*, 3(1):1009–1020, 2010.

[8] A. M. Ferman, J. H. Errico, P. van Beek, and M. I. Sezan. Content-based filtering and personalization using structured metadata. In *JCDL*, page 393, 2002.

[9] J. L. Herlocker, J. A. Konstan, A. Borchers, and J. Riedl. An algorithmic framework for performing collaborative filtering. In *SIGIR*, pages 230–237, 1999.

[10] Z. Huang, H. Chen, and D. D. Zeng. Applying associative retrieval techniques to alleviate the sparsity problem in collaborative filtering. *ACM TOIS*, 22(1):116–142, 2004.

[11] M. Jamali and M. Ester. *TrustWalker*: a random walk model for combining trust-based and item-based recommendation. In *KDD*, pages 397–406, 2009.

[12] I. Konstas, V. Stathopoulos, and J. M. Jose. On social networks and collaborative recommendation. In *SIGIR*, pages 195–202, 2009.

[13] H. Ma, I. King, and M. R. Lyu. Learning to recommend with social trust ensemble. In *SIGIR*, pages 203–210, 2009.

[14] H. Ma, M. R. Lyu, and I. King. Learning to recommend with trust and distrust relationships. In *RecSys*, pages 189–196, 2009.

[15] H. Ma, H. Yang, M. R. Lyu, and I. King. SoRec: social recommendation using probabilistic matrix factorization. In *CIKM*, pages 931–940, 2008.

[16] H. Ma, D. Zhou, C. Liu, M. R. Lyu, and I. King. Recommender Systems with Social Regularization. In *WSDM*, 2011.

[17] P. Melville, R. J. Mooney, and R. Nagarajan. Content-Boosted Collaborative Filtering for Improved Recommendations. In *AAAI/IAAI*, pages 187–192, 2002.

[18] B. M. Sarwar, G. Karypis, J. A. Konstan, and J. Riedl. Item-based collaborative filtering recommendation algorithms. In *WWW*, pages 285–295, 2001.

[19] W. R. Tobler. A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, 46:234–240, 1970.

[20] H. Tong, C. Faloutsos, and J.-Y. Pan. Fast Random Walk with Restart and Its Applications. In *ICDM*, pages 613–622, 2006.

[21] J. Wang, A. P. de Vries, and M. J. T. Reinders. Unifying user-based and item-based collaborative filtering approaches by similarity fusion. In *SIGIR*, pages 501–508, 2006.

[22] M. Ye, P. Yin, and W.-C. Lee. Location Recommendation in Location-based Social Networks. In *GIS*, pages 458–461, 2010.

[23] H. Yildirim and M. S. Krishnamoorthy. A random walk method for alleviating the sparsity problem in collaborative filtering. In *RecSys*, pages 131–138, 2008.

[24] Q. Yuan, S. Zhao, L. Chen, S. Ding, X. Zhang, and W. Zheng. Augmenting collaborative recommender by fusing explicit social relationships. In *ACM RecSys-Workshop*, pages 49–56, 2009.

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

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

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