Location Recommendation for Location-based Social Networks

Mao Ye Peifeng Yin Wang-Chien Lee
Department of Computer Science and Engineering
The Pennsylvania State University
University Park, PA 16802, USA
{mxy177,pzy102,wlee}@cse.psu.edu

ABSTRACT

In this paper, we study the research issues in realizing location recommendation services for large-scale location-based social networks, by exploiting the *social* and *geographical* characteristics of users and locations/places. Through our analysis on a dataset collected from Foursquare, a popular location-based social networking system, we observe that there exists strong social and geospatial ties among users and their favorite locations/places in the system. Accordingly, we develop a *friend-based collaborative filtering* (FCF) approach for location recommendation based on collaborative ratings of places made by social friends. Moreover, we propose a variant of FCF technique, namely *Geo-Measured FCF* (GM-FCF), based on heuristics derived from observed geospatial characteristics in the Foursquare dataset. Finally, the evaluation results show that the proposed family of FCF techniques holds comparable recommendation effectiveness against the state-of-the-art recommendation algorithms, while incurring significantly lower computational overhead. Meanwhile, the GM-FCF provides additional flexibility in tradeoff between recommendation effectiveness and computational overhead.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Recommendation system

General Terms

Application

Keywords

Location recommendation, social networks, location-based systems

1. INTRODUCTION

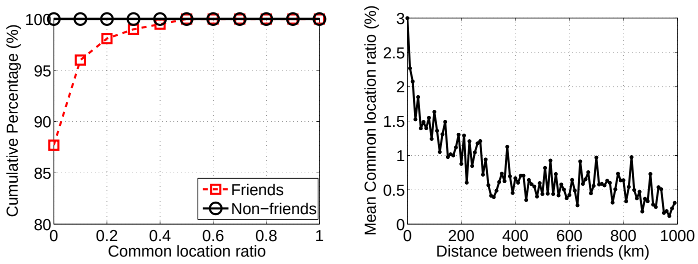
With the rapid development of mobile devices, wireless networks and Web 2.0 technology, a number of location-based social networking services, e.g., Loopt[1] and Foursquare[2], have emerged in recent years. These location-based social networking services allow users to connect with friends, explore places (e.g., restaurants, stores, cinema theaters, etc), share their locations, and upload photos, video, and blogs. As city and neighborhood exploration is one of the main themes in many location-based social networking services, it is highly desirable for such services to provide *location recommendations* to their users. Moreover, as the users and locations in location-based social networking services are rapidly growing, it is essential to adopt efficient techniques for realizing location recommendation.

Fueled by a broad application base in on-line shopping, recommendation systems have received a lot of attention from both the industry and academia. Recently, a number of studies have explored the use of social relationship amongst users in recommendation systems to enhance the effectiveness of recommendation techniques [1, 3, 4]. The main ideas behind these studies are to employ the trust and interest similarity carried along in social relationships amongst friends to enhance personalized search and recommendations. However, these prior works mainly focus on conventional recommendations in social networks. In this paper, we study the research issues in realizing location recommendation for large-scale location-based social networks. Particularly, we aim at exploiting both *social* and *geographical* characteristics of relationships amongst users and places to support the location recommendation services.

While existing recommendation techniques for conventional social networking systems may be applicable to the location recommendations, these techniques [2, 3] usually incur very high computational overhead. Owing to the rapidly growing number of users and locations in location-based social networking services, these existing techniques are not efficient for on-line recommendation services. By exploring the strong social tie between friends, in this paper, we propose the *friend-based collaborative filtering* (FCF) approach for location recommendation based on collaborative ratings of commonly visited places made by social friends. Additionally, we propose a variant of FCF technique, namely *Geo-Measured FCF* (GM-FCF) based on heuristics derived from observed geospatial characteristics in the Foursquare dataset. We conduct a comprehensive performance evaluation to validate our proposal and make comparison with the state-of-the-art recommendation algorithms, including the conventional collaborative filtering (CF), social collaborative filtering (SCF) and random walk and restart (RWR). The evaluation results show that the family of FCF techniques holds comparable recommendation effectiveness against the compared algorithms, while incurring significantly lower computational overhead. Meanwhile, the GM-FCF provides additional flexibility in tradeoff between recommendation effectiveness and computational overhead.

[1]www.loopt.com
[2]www.foursquare.com

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage, and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. ACM GIS '10 , November 2-5, 2010. San Jose, CA, USA. (c) 2010 ACM ISBN 978-1-4503-0428-3/10/11...$10.00.
.



(a) Friends vs. Non-friends (b) Local vs. Distant friends

Figure 1: Spatio-Social analysis

2. SPATIO-SOCIAL ANALYSIS

Location-based social networking systems are different from conventional social networking systems in its geospatial emphasis. To identify unique social and spatial characteristics in such systems, we perform a spatio-social analysis upon the collected Foursquare data.

2.1 Data Collection

Due to the lack of publicly available location-based social networking data for analysis and experimentation, we crawl live data from Foursquare, one of the most popular location-based social networking services. We aim the data crawling task at collecting user and location data. We select a well-connected user (who has hundreds of friends) as a seed to start the crawling. When a user is visited, we examine his/her profiles to extract i) the address information; ii) the list of friends; and iii) the locations where he/she is the mayor.[3] Through a breath-fist traversal of the social network, we obtain the user dataset. Afterwards, by crawling the locations collected in the user data collection process, we extracts the location addresses and the list of visitors. After aggregating the user and location data, we obtain a dataset of 58,659 users and 96,219 locations. Note that each user and location in the dataset is associated with an address which is converted into geographical point (i.e., latitude and longitude) via Google map service.

2.2 Data Analysis

Upon the collected data, we perform an analysis to better understand how the users and locations in location-based social networks are associated with each other.

First, we would like to see whether socially connected users (i.e., friends) on location-based social networks are different from an arbitrary pair of users in terms of commonly visited locations. To proceed with this analysis, we first define the *common location ratio* between a given pair of users u_i and u_j and denote it as $\alpha(u_i, u_j)$. Let L_i and L_j denote the sets of locations visited by u_i and u_j, respectively. Then, the common location ratio $\alpha(u_i, u_j)$, measuring the degree of overlap in visited locations, is defined as $\alpha(u_i, u_j) = \frac{|L_i \cap L_j|}{|L_i \cup L_j|}$

We use common location ratio to examine two different sets of user pairs — one consists of friends and the other non-friends. Figure 1(a) plots the cumulative percentage of these two types of user pairs corresponding to common location ratios. As shown in the figure, the common location ratios between user pairs is generally very small. Nevertheless, friends obviously have a higher chance to share commonly visited locations than non-friends. For example, as indicated in Figure 1(a), about 4% of friends have a common location ratio greater than 10%. We consider visited locations as an indicator for user interests. Thus, the common location ratio may represent the common location interests between two users. Indeed, the finding shown in Figure 1(a) is consistent with our life experience, i.e., friends share much more common interests than two arbitrary people do. Because the basic idea behind some recommendation techniques such as collaborative filtering is to make use of common interests between similar users to predict missing ratings, we believe friends in location-based social networks may play a positive role in collaborative recommendation.

Besides the difference in common location ratios between friends and non-friends, nevertheless, Figure 1(a) also shows that about 96% of friends share less than 10% commonly visited locations and about 87.7% of friends share nothing in common. This reveals some important information — since there is a great number of friends sharing nothing in common, not all social friends will contribute to location recommendation. Thus, we next investigate what kind of friends may share more commonly visited locations than others. To find the correlation between common location ratio and distance of friend pairs, we group friends based on their distance in steps of 10 km, i.e., friends located in $0 \sim 10$ km form group 1, those located in $10 \sim 20$ km form group 2, and so on. Figure 1(b) shows the mean common location ratio corresponding to these groups. As shown, the mean common location ratio follows a decreasing trend as the distance between friends increases. This verifies our expectation that nearby friends have a much higher probability to share common locations since it is easier for them to participate in activities at the same locations.

In summary, based on the above observations, a natural hypothesis is that social activity (visiting locations) of a user in location-based social network is largely affected by geographical proximity.

3. COLLABORATIVE LOCATION RECOMMENDATION

In location-based social networks, users update their visited locations anytime and anywhere. To achieve highly accurate location recommendation, it's essential to realize the service efficiently. As discussed earlier, through large-scale matrix multiplications, many existing collaborative filtering techniques require a full scan of all users and locations in the process of location recommendation for a given user. Hence, scalability is a critical issue since large-scale matrix multiplication incurs tremendous computational overhead for collaborative recommendation [5, 6, 7].

3.1 Friend-based Collaborative Filtering (FCF)

According to our data analysis, we observe that social friends share more common locations than non-friends. Thus, an idea is to consider only friends when processing collaborative filtering for a targeted user, since non-friend users do not have much value for reference in recommendation. As such, we only need to compute the similarity weight between friends, instead of all users, and the given user. We expect great savings in matrix computation because the matrix size is significantly reduced. There is a tradeoff in relying only on friends in collaborative recommendation. On the one hand, since non-friends are not considered, much noise is reduced and thus good for *precision*. On the other hand, there are indeed some cases where non-friends share common locations. Those persons also contribute to recommendation, especially when there are few similar users to the targeted person. Eliminating these non-friends may hurt *recall* as a result, i.e., some potentially preferred locations are not recommended due to the lack of non-friends with similar location interests. We will study the impact on precision and recall later in the performance evaluation.

In FCF, only friends are used as references in collaborative filtering. In other words, we only need to calculate the simi-

[3]Foursquare does not show visited locations in user profiles.

larity weight between friends, instead of every pair of users. Notice that we do not need to introduce social friendship explicitly to adjust similarity weight as in SCF. Since only friends are included to calculate the similarity weight $w_{i,k}$, the social friendship has already been taken into consideration implicitly. For simplicity, we adopt cosine similarity measurement here. In the following, we redefine the rate predication function to consider only friends. Let U and L denote the user set and the location set in a location-based social networking system. Besides, let U_i denote the friend set for a given user u_i. We assume that the system keeps track of the rating a user $u_i \in U$ put on a visited location $l_j \in L$ and denote it as $r_{i,j}$. These recorded user ratings on locations are thus used to predict possible ratings of the user on unvisited locations. We denote this predication as $\widehat{r}_{i,j}$ and obtain this predicted rating of u_i on l_j as follows.

$$\widehat{r}_{i,j} = \frac{\sum_{u_k \in U_i'} r_{k,j} w_{i,k}}{\sum_{u_k \in U_i'} w_{i,k}} \tag{1}$$

where U_i' ($\subseteq U_i$) is the set of friends with top-m similarity weight. Notice that the number of friends ($|U_i|$) is usually much smaller than $|U|$, i.e.$|U_i| \ll |U|$.

In FCF, for a given user, only his friends are involved in the computation of similarity weight and contribute to the missing rating prediction. Thus, we estimate the computation cost for collaborative filtering as $C(FCF) = |U_i| \times |L| + m \times |L| < C(CF)$ because of $|U_i| \ll |U|$. Furthermore, we conclude that $C(FCF) < C(SCF)$.

3.2 Geo-measured friend-based collaborative filtering (GM-FCF)

According to the spatio-social analysis of Foursquare data, we observe that nearby friends tend to share more commonly visited locations (see Figure 1(b)). Thus, instead of scanning friends' visited locations to calculate their similarity weights, an idea is to model the similarity weight between friends by their distance. Accordingly, we propose *geo-measured friend-based collaborative filtering (GM-FCF)* which uses linear regression method upon power-law distribution of distances between friends to learn a friend similarity model. The power-law distribution is formulated as $y = \alpha x^{\beta}$, where x denotes the variable of distance between friends and y here denotes the variable of common location ratio. Both y and x are often transformed into "log-log" scale, where a linear model can fit as below.

$$\log_{10} y = w_0 + w_1 \log_{10} x \tag{2}$$

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

$$\alpha = 10^{w_0} \qquad \beta = w_1 \tag{3}$$

Hence, we can simply apply a linear curve fitting method to realize regression as follows. More specifically, let $y' = \log_{10} y$ and $x' = \log_{10} x$. We shall fit data as follows

$$y'(x', W) = w_0 + w_1 \cdot x' \tag{4}$$

where w_0 and w_1 are the linear coefficients, collectively denoted by W. In order to avoid over-fitting when we approach the weight coefficients by least square error method, we add a penalty term (i.e., regularization term) to discourage the coefficients from reaching large values as below.

$$E(W) = \frac{1}{2}\sum_{n=1}^{N}\{y'(x_n', W) - t_n\}^2 + \frac{\lambda}{2}||W||^2 \tag{5}$$

where N presents the cardinality of input dataset, t_n is the ground truth corresponding to x_n', and λ is the regularization term.

As to be discussed further later in performance evaluation, we experimentally apply the above linear regression upon the Foursquare dataset and obtain $\alpha = 0.0414$ and $\beta = -0.508$, which will be used in experiments. Then, by utilizing $y = \alpha x^{\beta}$, we have the similarity weight between two friends u_i and u_k

$$w_{i,k} = \frac{y(x = d(u_i, u_k), \alpha, \beta)}{\sum_{u_k \in U_i} y(x = d(u_i, u_k), \alpha, \beta)} \tag{6}$$

where U_i is the friend set of u_i.

In GM-FCF, instead of scanning all locations to calculate the similarity weight between friends, we only access the latitude and longitude (a 2-dimension geo-point) for similarity weight estimation. Thus, the computation cost for similarity weight calculation is estimated as $2 \times |U_i|$. The total computation cost is $C(GM - FCF) = 2 \times |U_i| + m \times |L|$, where $m \times |L|$ accounts for the cost of missing rating predication.

4. EXPERIMENTS

In this section we evaluate performance of the proposed FCF techniques and compare them against *social collaborative filtering* (SCF) and *random walk with restart* (RWR), two state-of-the-art collaborative recommendation techniques.

We crawled the Foursquare website, one of the most representative location-based social networks, for a month to collect a data set consisting of 58,659 users and 96,219 locations in total. Due to the lack of location ratings in Foursquare, we use for each user/location pair a binary rating, where "1" means the user visited the location and "0" otherwise. Also, to avoid sparsity problems, we only include users who have visited more than 2 locations in our experimental dataset, similar as in [3]. After preprocessing, we have a User-User (UU) matrix with density of 5.9×10^{-3} and a User-Location (UL) matrix with density of 2.2×10^{-3}. For each individual user in the dataset, we randomly remove 20% of all locations that he/she has visited, and use location recommendation algorithms to recover the missing user-location pairs we remove.

4.1 Metrics

In location-based social networking systems where users are able to instantly share their visited locations, the performance of provided recommendation services are required to be highly effective and efficient.

4.1.1 Effectiveness

A location recommendation algorithm computes a ranking score for each candidate location (i.e., one user has not visited) and returns the top-N highest ranked locations as recommendation to a targeted user. To evaluate the prediction accuracy, we are interested in how many locations previously removed in the preprocessing step re-appear in the recommended results. More specifically, we examine (1) the ratio of recovered locations to the N recommended locations, and (2) the ratio of recovered locations to the set of locations deleted in preprocessing. The former is called *precision@N* ($P(N)$) while the latter is called *recall@N* ($R(N)$), collectively denoted as $P@N$ (where P represents "Performance"). Moreover, we take another further step to consider the precision and recall relative to random recommendations [8]. More specifically, the relative precision and recall are defined as follows.

$$P_r(N) = \frac{P(N)}{P_{rand}(N)} = P(N) \times \frac{|L||U|}{|D|} \tag{7}$$

$$R_r(N) = \frac{R(N)}{R_{rand}(N)} = R(N) \times \frac{|L|}{N} \tag{8}$$

where $|D|$ is the total number of deleted locations among all users.

4.1.2 *Efficiency*

Collaborative filtering consists of two major parts, similarity calculation and rating predication, both of which essentially incur matrix multiplication. More specifically, the cosine similarity between users can be projected as multiplication of a $|L|$-dimension vector $\overrightarrow{V_i}$, whose values are either 0 or 1, indicating whether u_i visited the location, and a location-user matrix $M_{|L|\times|U|}$. The result is a vector $\overrightarrow{V_i'}$ of length $|U|$ which includes all similarity between any user u_k and the given targeted user u_i. For rating predication, computation is a multiplication between $\overrightarrow{V_i'}$ and a user-location matrix $M_{|U|\times|L|}$, and the result is a vector of predicated ratings for the targeted user u_i to all locations in L. As for RWR, the main computation is about the "random walk", which can be treated as rating predication [2, 3]. Each iteration of random walk is a rating predication based on current status vector and the transitive matrix. Different from collaborative filtering, there are multiple iterations (I) for RWR till the result converges. Let $C_i(\cdot)$ denote the computation cost incurred for user u_i, then we have the average computation cost as $C(\cdot) = \frac{\sum_{u_i \in U_A} C_i(\cdot)}{|U_A|}$ where U_A denotes the set of targeted users. We compare the efficiency by their computation costs summarized below.

$$C(CF) = |U| \times |L| + m \times |L|$$

$$C(FCF) = \frac{\sum_{i=1}^{|U_A|} |U_i|}{|U_A|} \times |L| + m \times |L|$$

$$C(GM\text{-}FCF) = 2 \times |L| + m \times |L|$$

$$C(SCF) = |U| \times |L| + \frac{\sum_{i=1}^{|U_A|} U_i}{|U_A|} + m \times |L|$$

$$C(RWR) = |I| \times (|U| + |L|)^2$$

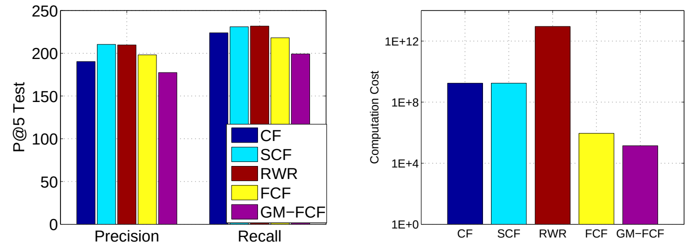
where U_i denotes u_i's friend set.

4.2 Experiments and Results

Here we compare the overall performance of CF, SCF, FCF, GM-FCF and RWR[4]. Figure 2(a) plots both relative precision and recall for P@5. As shown, social friendship is beneficial for location recommendation in social networks, since all algorithms that take social relation into account outperform CF in terms of precision. However, in contrast to the significant improvement reported in [3], the social factor only brings minor improvement over the CF algorithm. As for the recall, SCF and RWR outperforms CF. FCF considers only friends in its recommendation, suffering missing knowledge of non-friend users and thus hurting its recall a little bit. However, the differences in recall amongst all compared algorithms are insignificant. The experimental result shows that the proposed FCF technique is very competitive in comparison with other collaborative recommendation techniques.

On the other hand, the strength of FCF lies in its computational efficiency which is a mandate for on-line location recommendation. Figure 2(b) shows the computational cost (in *log* scale) among the compared techniques. As shown, FCF significantly outperforms all other techniques in several order of magnitude. In summary, FCF prevails because it only considers friends (a small number compared with the whole user set) when processing location recommendation for a given user, leading to a much lower computation overhead. On the other hand, the effectiveness of FCF remains competitive because nearby friends provide a high-quality pool of references for location recommendations.

[4]With extra experiments not shown due to space constraint, the probability of "restart" for RWR is set to 0.9 for RWR to achieve optimal performance.



(a) Effectiveness (b) Efficiency - log scale

Figure 2: Performance comparison among CF, SCF, RWR, FCF and GM-FCF

Also in Figure 2, we evaluate the performance of GM-FCF. As shown, the effectiveness of GM-FCF is competitive to FCF, while shows great advantages in efficiency. To sum up, the GM-FCF location recommendation techniques provide flexible tradeoff between recommendation effectiveness and computational overhead.

5. CONCLUSIONS

In this paper, we study the research issues in location recommendation services for location-based social networking systems. Through the spatio-social analysis of Foursquare dataset, we observe that there exist strong social and geospatial ties among users and their visited locations in the system. In light of friends displaying similar behavior, we develop *friend-based collaborative filtering* (FCF) approach for location recommendations based on collaborative filtering of locations among social friends. Moreover, we propose a variant of the FCF technique, namely, *GM-FCF*, based on heuristics derived from observed geospatial characteristics in the Foursquare dataset. Finally, we validate the proposed ideas and evaluate the FCF family of techniques via comprehensive experimentation. The evaluation results show that the family of FCF techniques holds comparable recommendation effectiveness against the state-of-the-art recommendation algorithms, while incurring significantly lower computational overhead.

6. REFERENCES

[1] J. Golbeck. Tutorial on using social trust for recommender systems. In *RecSys*, pages 425–426, 2009.

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

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

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

[5] G. Takács, I. Pilászy, B. Németh, and D. Tikk. Scalable collaborative filtering approaches for large recommender systems. *Journal of Machine Learning Research*, 10:623–656, 2009.

[6] H. Tong, C. Faloutsos, and J.-Y. Pan. Fast random walk with restart and its applications. In *ICDM*, pages 613–622, 2006.

[7] K. Yu, S. Zhu, J. D. Lafferty, and Y. Gong. Fast nonparametric matrix factorization for large-scale collaborative filtering. In *SIGIR*, pages 211–218, 2009.

[8] T. Zhou, Z. Kuscsik, J.-G. Liu, M. Medo, J. R. Wakeling, and Y.-C. Zhang. Solving the apparent diversity-accuracy dilemma of recommender systems. *PNAS*, 2010.