공학석사 학위논문

# 다양한 영향을 고려한 장소 추천

POI Recommender System with Various Effects

2017년 2월


서울대학교 대학원

컴퓨터공학부

한 다 현

# 다양한 영향을 고려한 장소 추천

POI Recommender System with Various Effects

지도교수 김 종 권

이 논문을 공학석사 학위논문으로 제출함

2016 년 10 월


서울대학교 대학원

컴퓨터 공학부

한 다 현


한다현의 공학석사 학위논문을 인준함

2016년 12월

| | | |
|---|---|---|
| 위 원 장 : | 전 화 숙 | (인) |
| 부위원장 : | 김 종 권 | (인) |
| 위　　원 : | 권 태 경 | (인) |

# Abstract

# POI Recommender System with Various Effects

DaHyeon Han

Department of Computer Science & Engineering

The Graduate School

Seoul National University

More and more Location-based Social Networks(LBSNs) becomes significantly popular due to generalization of smartphones and tablets. With the high accessibility for LBSNs dataset, Point-of-Interest (POI) recommendation gets lots of attention not only in academic area but also in industry recently. As normal recommendation (move and product), it has data sparsity problem. In addition, as opposed to normal recommendation, it has to deal with physical distance. Hence, make the problem harder. To address the aforementioned problems, many researchers try to utilize assorted effects such as user's preference, temporal effect, geographical effect and social effect. In this paper, we proposed UFPCF-G which utilize User's preference, different types of Friends' preference, POI similarity and User's activity based Collaborative Filtering with Global Influence. We also explore extended meaning of "friends". Unlike other research, we do not regard "social relationship" as the solely "friends". In addition, we propose structure based and activity based global influence. We execute extensive experiment on two real LBSNs dataset and our experiment results outperform other baseline methods in terms of precision and recall.

Keyword : POI Recommender System, Collaborative Filtering, Location Based Social Networks. Social Network.

Student Number : 2015-21275

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

With the generalization of smartphone and growth of smartphone markets, services who provide mobile services grow rapidly. Among variety of services, recommendation services is judged as a one on most promising services. From traditional POI (Point-Of-Interest) recommendation services providers like Foursquare and Yelp to Social Network Services like Facebook and Twitter, A great number of companies jump into recommendation system. The service users can share to have record of how they feel, whom they enjoy with, physical location and check-in time as well as can obtain recommendation of service contents. In MovieLens which is one of the most famous movie rating services, users can share comments on a specific movie. In foursquare which is one of the most popular POI recommender services, users can share the comments and put check-in to leave record. Besides typical recommender system, one of the most familiar recommendation services is the one in YouTube, which recommends some video contents that are considered to be liked by users. When users experience convenience and comfort to the service they uses, their loyalty to the service provider increases. In order to maximize user's affection and satisfaction to LBSNs (Location-Based Social Networks), utilizing check-in data thoroughly to recommend proper POIs for target user is necessary. The quality of recommendation is also able to attract more users to join LBSNs which makes the company more renowned.

Recommendation system is attractive not only to service providers but also to researchers. There are extensive studies on movie recommendation and POI recommendation. Movie recommender system is one of the most famous type, and POI recommender system is getting lots of attentions because of easy access to large size dataset from LBSNs. Despite researchers' efforts to find attractive locations, researchers have to face two major challenges. The former one is data sparsity. User visits only few locations, while we have more than thousands to millions locations. The latter one is that it has to deal with physical distance which makes the problem harder.

There are similar challenges also exist in movie recommender system. Movie recommender system also have data sparsity problems. To overcome mentioned difficulty, movie recommender system uses variety of additional information. Of those additional information like social information, similar users and similar items. Social information provided from social networks become one of the key factor to solution of data sparsity. So, POI recommender system also adopts social information as a solution to performance improvement.

The usage of social information in POI recommendation generates new challenges. That is because of different characteristics between POI and movie (1) User's POI selection characteristics is affected by geographical influence greatly (2) if friends in online social networks have different activity area, there's exceedingly low probability of sharing visited POI. (3) User shares POI with other users who live in the same area. Therefore, proper techniques to employ social information is needed.

Our contribution of paper is as follows

- We define new types of friends for users. We define social friend, neighbor friend and core friend to enrich friend information. We describe details in section 3.2

- We also utilize POI-based CF with regard to both similarity and distance to reflect POI' similarity and user's activity range. We explain details in section 3.3

- We propose global influence for both users and POIs. We propose pagerank and check-in frequency to catch user's global influence (structure and activity based approach). We also employ POI's popularity to catch global influence of POI. Details are explained in section 3.4

- Our proposed model outperforms six baseline on two real world large datasets significantly.(section 4.3)

# Chapter 2

# Related Work

## 2.1. Recommendation Technique

Content-based and Collaborative Filtering techniques are two widely adopted techniques in recommender system. Content-based technique recommends items by similarity score between users' profile contents (gender, age, region etc.) [1][2][3]and item content (main actor, genre etc.). Since it employs only profile contents, it fail to utilize interactions and associations between users. Collaborative Filtering. (CF) can be further categorized into model-based CF and memory-based CF. Model-based CF technique proposes a model using data mining knowledge such as clustering (cluster item/location by Euclidean distance). Then get a score for a specific item/location based on proposed model. The model-based CF needs extensive time for learning compact model based on observed user-item ratings. Memory-based CF can be divided into user-based CF and item-based CF. User-based CF aims to find users who has similar taste first, then calculate the score of items by weighting the similar users' historical ratings. Cosine Similarity and Pearson Correlation Coefficient is well known similarity measure. Unlike User-based CF, Item-based CF focuses on finding similar items that user liked before and recommends these items to users.

Furthermore, recently, to cope with large datasets, there are many research to approximate user-item rating matrix by low dimensional matrix[4][5]. Traditionally, the Singular Value Decomposition (SVD) method is used to approximate user-item rating matrix by minimizing related Frobenius norm. However, user-item matrix is always vey sparse, hence researchers only factorize observe ratings in user-item rating matrix. In order to avoid overfitting, some regularization terms are used. Then we are able to solve this optimization problem by Gradient Descent or Alternating Least Square method. Unlike movie or product recommender system POI recommender has new challenges. That is because of different characteristics between POI and movie, which is spatial characteristics. (1) User's POI selection characteristics is affected by geographical influence greatly (2) if friends in online social networks has different activity area

## 2.2. Recommendation

Before POI recommender systems get attention, almost recommender systems have focused on movie and product recommender system. In addition, most of them are contents-based and CF-based algorithms[6]. Due to data sparsity problem, many researchers employ additional information. The most widely adopted additional information is social information. Social information is well known for similar interests of friends and many research[5][7][8] show their improvement with social information. However, [8] insists that particularly for POI recommender system social information might not that beneficial due to physical distance. Many research shows that [9][10] users' willingness to POI follows the power law distribution with distance. Even [8][11] prove that similarity between user and social relation in LBSN like Foursquare and Gowalla is much less than movie or product website like MovieLens and Youtube. Recently, with the substantial growths of mobile device markets and associated LBSNs, POI recommender have been extensively researched[9][10][12][13][14]. [7] proposed social topic modeling that social friend shares topics of POI rather than actual POI itself. [13] uses similar user and social relation simultaneously only. [15] introduce global influence based matrix factorization, however they regard global influence as pagerank value only. [10] proposed matrix factorization with three types of friend set.[9] suggest CF algorithm with temporal effect and spatial effect, but they do not utilize social information. [16][17] recommend POIs which are closer to users. [18] [19] have propose geographical influence to catch local popularity. However, [9][14][17][18][19] have not consider social influence. And none of above researches employs POI-based CF, “core friend” and correct usage of Global Influence.

# Chapter 3

# Proposed Scheme

In this section we describe our proposed CF step by step. We introduce user-based CF in section 3.1, new definition of friends and friend-based CF in section 3.2, location-based CF in 3.3. Then finally we will introduce global context for users and locations respectively in 3.4

## 3.1. User−based CF

User-based CF is used for recommending item/POI by aggregating the history of similar users. Let $u_i$ denote the $i^{th}$user in user set $U$ and $l_j$ denote the $j^{th}$POI in POI set $V$ in LBSN. Let $c_{i,j}$ denote the check-in activity where $c_{i,j}$ =1 indicates a user $u_i$ has a check-in record at POI j and $c_{i,j}$=0 means that user $u_i$ has never visited POI $l_j$ before. Given a user $u$ the preference score of $u$

check-in at unvisited POI *l* is calculated as following equation.

$$p_{u,l}^{U} = \frac{\sum_{v\in u'} w_{u,v}}{\sum_{v\in u'} w_{u,v}} \quad (1)$$

Where *U*’ denote most similar top k users, $w_{u,v}$ is a similarity between $u_i$ and $u_v$. To compute similarity between two users, cosine similarity and Pearson correlation coefficient are two popular similarity measures. Owing to simplicity, we employ cosine similarity measure to calculate similarity between users. The cosine similarity equation is as follow.

$$w_{u,v} = \frac{\sum_l c_{u,l}\, c_{v,l}}{\sqrt{\sum_l c_{u,l}^2}\sqrt{\sum_l c_{v,l}^2}} \qquad (2)$$

Let us assume that, $u_i$ and $u_v$ have quite different interest, but both of them have visited the same POI once. Then either of them would be included in equation (2), in order to avoid regarding dissimilar users as similar users, we choose top k (30) users as similar users.

## 3.2. Friend−based CF

In real world, we share lots of common interests with our friends and tend to have similar behavior. For example, when we look for an Italian restaurant or picnic place, it is normal to ask our friends. Furthermore, we often hang out with friends for shopping, theater and gym etc. Hence “social relationship” in dataset has been widely adopted for recommender system.[14][15]. However we assume that "social relationship" on LBSNs cannot reflect all real world friends. In order to efficiently catch "real world friends" we have define new types of friends with specific reasons.

- Social Friend: Social Friend is commonly used social relation. As previous work[5][7], we regard user *i* and user *v* as friend if there exist social relationship in dataset. Since social friend is widely adopted concept, we also employ social friend as one of proposed friend types
- Neighbor Friend: Particularly, in POI recommender system, user's willingness to visit a POI follows a power law distribution with distance. It means distance is very crucial factor for selecting POI. It is also common that we often talk to our neighbor. For example "Where did you go during last weekend?" or "I visited newly opened restaurant in front of the post office, it was great!" Thus, we choose *k*(30) users as user i's neighbor friends whose home address is closest to user i's home address. Unfortunately, due to privacy problem, users' home addresses are not accessible. To defer the home address, we adopt the method introduced by [20][21]. This method divide the world into 25km by 25km cells then average the position of check-ins

in the cell with most of his/her check-ins. With this method we are able to defer user's home address (latitude, longitude) then, use Harversine formula to calculate physical distance.

$$d = 2r \arcsin\left(\sqrt{\operatorname{hav}(\varphi_2 - \varphi_1) + \cos(\varphi_1)\cos(\varphi_2)\operatorname{hav}(\lambda_2 - \lambda_1)}\right) \\ = 2r \arcsin\left(\sqrt{\sin^2\left(\frac{\varphi_2 - \varphi_1}{2}\right) + \cos(\varphi_1)\cos(\varphi_2)\sin^2\left(\frac{\lambda_2 - \lambda_1}{2}\right)}\right) \quad (3)$$

where *r* is the radius of the sphere, $\varphi_1$,$\varphi_2$ indicate latitude of point 1 and latitude of point 2 in radians, $\lambda_1$,$\lambda_2$ denote the longitude of point 1 and longitude of point 2 in radians and hav(θ) is Harversine function defined as below

$$\operatorname{hav}(\theta) = sin^2\left(\frac{\theta}{2}\right) = \frac{1-\cos(\theta)}{2} \quad (4)$$

- Core Friend: Social relation is widely used information. However, we assume that existing social relation in online social networks cannot reflect real world friend with two specific reasons. First and foremost, some of online friends are not familiar friends. For example, Tony has joined the summer camp of Boy Scout for two days as a volunteer, then make a social relation with other volunteers during the short camp. After the camp, there would exist social relation even though they have talked to each other one or two days. After one or two years later can we still assert that they share similar interest due to online social relation? Probably not. In addition, some users are not that active online, hence do not subjectively add friend online. Due to these two reasons, we have consider “core friend” which is structurally close friend. Our assumption is that users who share similar friends have similar interest. Hence we utilize Jaccard coefficient, which is widely adopted mechanism in link prediction. The basic concept of jaccard coefficient is that an edge<x,y> is more likely to form if edge<x,z> and <z,y> are already present[22][23]. The pure jaccard coefficient is calculated as below

$$\mathrm{J(A,B)} = \frac{|\Gamma(A) \cap \Gamma(B)|}{|\Gamma(A) \cup \Gamma(B)|} \quad and\ A \notin \Gamma(B), B \notin \Gamma(A) \qquad (5)$$

Where $\Gamma(\cdot)$ indicates friend set of user. The difference between pure jaccard coefficient our mechanism is that we allow $A \in \Gamma(B)$ and $B \in \Gamma(A)$. Hence we can filter the just normal friend and structurally closer friends. In order to catch really strong relation, we choose top k (10) users as core friend set

With above mentioned three types of friends. The preference of user based on three types of friend is calculated as follow

$$p_{u,l}^{F} = \frac{1}{3} * \frac{\sum_{v \in s'} w_{u,v}\, c_{v,l}}{\sum_{v \in s'} w_{u,v}} + \frac{1}{3} * \frac{\sum_{v \in s'} w_{u,v}\, c_{v,l}}{\sum_{v \in n'} w_{u,v}} + \frac{1}{3} * \frac{\sum_{v \in s'} w_{u,v}\, c_{v,l}}{\sum_{v \in c'} w_{u,v}} \quad (6)$$

Where *s'*, *n'* and *c'* indicate social friend set, neighbor friend set and core friend set respectively. We simply average the preference of three types of friends.

Some readers might assume that, there would be lots of overlapped users. Table 1 and Table 2 show the overlapped users' ratio in Foursquare and Yelp respectively. "Similar User" indicate the user set which is introduced in section 3.1, similar user set of user-based CF. According to Table 1 and Table2, the overlapped ratio of users is insignificant. Interestingly, overlapped ratio between core friend and similar user is higher than core friend and social friend. This result indicates that influence of core friend is more crucial and influence of commonly used social friend. We double count the overlapped users with following reason. Let's assume that Tony and Sam are neighbors, they are classmates and they are friends on social network but they do not have similar interest. Ray is also Tony's neighbor and he has more similar interest with Tony compare to Sam. Even though Ray and Tony have similar interest, Tony has more opportunity to be influenced by Ray, because Tony can meet Ray not only in the apartment but also online and in the school.

**Table 1 Overlapped Ratio on Foursquare**

| | Similar User | Neighbor Friend | Core Friend | Social Friend |
|---|---|---|---|---|
| Similar User | | 0.026 | 0.053 | 0.018 |
| Neighbor Friend | 0.026 | | 0.019 | 0.014 |
| Core Friend | 0.053 | 0.019 | | 0.044 |
| Social Friend | 0.018 | 0.014 | 0.044 | |

**Table 2 Overlapped Ratio on Yelp**

| | Similar User | Neighbor Friend | Core Friend | Social Friend |
|---|---|---|---|---|
| Similar User | | 0.033 | 0.003 | 0.001 |
| Neighbor Friend | 0.033 | | 0.002 | 0.001 |
| Core Friend | 0.003 | 0.002 | | 0.024 |
| Social Friend | 0.001 | 0.001 | 0.024 | |

In order to verify effects of different types of friends, we borrow Soreg[5] model to check the accuracy with different types of friends set as social relation set. We choose SoReg for testing with following two reasons, first and foremost, it is very well known model for purely using social relation and matrix factorization. Second, in our proposed model we do not employ matrix factorization, however we intentionally show that, “Core friend set” and “Neighbor friend set” outperform widely used “Social relation set” in dataset and “Similar user set” in user-based CF. The result is described at figure 1

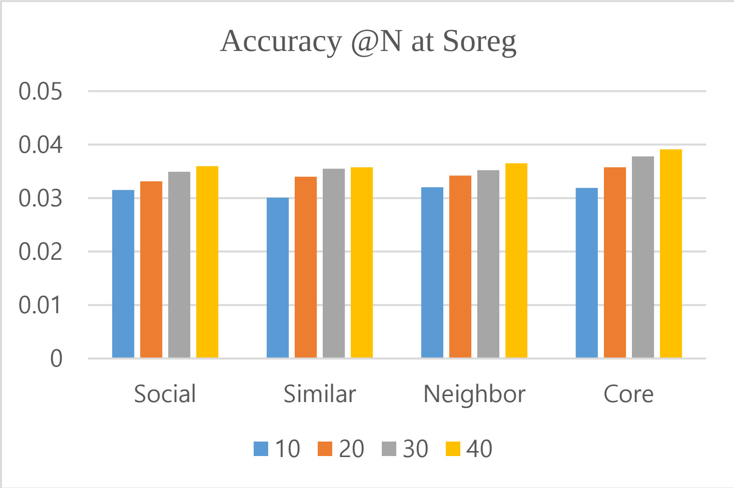


**Figure 1 Effect of Different Friend Set in SoReg on Foursquare**

Figure 1 shows Accuracy at different *N* where *N* indicates recommendation list size and accuracy denotes how many POI are visited by users. Surprisingly, Core friend set out performs any other friend sets and Neighbor friend set also exceeds Social friend set and Similar friend set.

## 3.3. POI−based CF

Location-based CF has similar properties with User-based CF. However, it focus on finding similar locations that user visited or liked before. Unlike user, location does not have a friend. We define two kinds of location as location set

- Similar POI: Similar POI is measured by equation (2), it indicates that how many users have visited the two POIs simultaneously. It can reflect the implicit characteristics of POI. For example, there are three restaurants, two of them sell steak and one of them sell sushi. Then the similarity between two steak houses would be higher than similarity between one sushi house and one steak house.

- Closest POI: In POI recommender system, physical distance plays very important role. Hence with collect top k closest POIs as closest POI set. Closest POI set can reflect user's activity range. Let us assume that Tony tends to visit down town, even though down town is very far away from his home. Then it is plausible to recommend POI around down town for Tony. Distance between two POIs is measured by Harversine formula equation (3)

For POI-based CF, the preference of user $u_i$ at POI $l_j$ is calculated as follow.

$$p_{u,l}^{L} = 0.5 * \frac{\sum_{k \in l'} w_{l,k}\, c_{u,k}}{\sum_{k \in l'} w_{l,k}} + 0.5 * \frac{\sum_{k \in c'} w_{l,k}\, c_{u,k}}{\sum_{k \in c'} w_{l,k}} \quad (7)$$

Where *l'* denote similar POI set for POI *l*, it reflect whether user has visited similar POI before and c' denote closest POI set for POI *l*. It mirrors that, whether this POI is in users' activity range.

## 3.4. Global Influence

### 3.4.1. Definition

In previous sections 3.1-3.3, we proposed user-based CF, friend-based CF and location based-CF. However, none of them employs global influence. The global influence reveals the reputation of a user in the whole social network [24]. User reputation can gives additional power for persuade someone. Hence there are lots of research to utilize global context[24][25]. There are many scheme to calculate global influence of node [26][27]. Our proposed global influence for user is divided by structure based influence and activity based influence. For structure based influence, we adopt PageRank [27]which is one of the most popular method for measuring the influence of nodes by edges. PageRank is computed as follow

$$\mathrm{PageRank(u)} = \frac{\delta}{N} + (1 - \delta) * \sum_{i=1}^{N} T[i,u] * PageRank(i) \quad (8)$$

Where *N* is total number of nodes, δ indicates probability of node jumps to any other node. T[i,u]=1/$N_i$ where $N_i$ denote number of links out of node *i*. Hence *PageRank(u)* indicates user u's global influence in terms of PageRank. For activity based influence, we define as follow.

$$\mathrm{Check-in(u)} = \frac{\#\,of\,check\text{-}in\,by\,user\,u}{Total\,\#\,of\,check\text{-}in} \quad (9)$$

Hence *Check-in(u)* indicates user's activity score. We can view a user with higher activity score as expert, since active users tend to visit more POI than non-active users. Finally, we combine structure based influence and activity based influence as follow

$$G_u = \alpha * \frac{PageRank(u)}{PageRank_{max}} + (1-\alpha) * \frac{Check-in(u)}{Check-in_{max}} \quad (10)$$

Where $PageRank_{max}$ and $Check-in_{max}$ indicates the maximum value of PageRank and Check-in respectively. These values are used for normalization. Depending on reader's taste, reader can pick $\alpha \in [0,1]$ to control the preference between structure influence and activity influence. In our experiment we pick $\alpha$ as 0.85 for Foursquare and 0.9 for Yelp.

In terms of global influence of POI, we are not able to employ structure based influence since POI does not have "social relationship" in LBSNs. Hence we only utilize activity based influence and defined as

$$GH_l = \frac{Check-in\ at(l)}{Check-in_{max}} \quad (11)$$

As we mentioned above, for a user with higher activity score means that user is an expert. Interestingly, the interpretation for a POI with higher activity score is different from user. We can regard a POI with higher activity score as popular POI, since the higher activity score for POI indicates that there are lots of check-ins at that POI.

## 3.4.2. Usage of Global Influence

To combine global influence with user-based CF, friend-based CF and location-based CF. We combine it as follow

$$p_{u,l}^{GU} = \frac{\sum_{v \in u'} w_{u,v} G_v \, c_{v,l}}{\sum_{v \in u'} w_{u,v} G_v} \qquad (12)$$

$$p_{u,l}^{GF} = \frac{1}{3} * \frac{\sum_{v \in s'} w_{u,v} G_v \, c_{v,l}}{\sum_{v \in s'} w_{u,v} G_v} + \frac{1}{3} * \frac{\sum_{v \in s'} w_{u,v} \, G_v c_{v,l}}{\sum_{v \in n'} w_{u,v} G_v} + \frac{1}{3} * \frac{\sum_{v \in s'} w_{u,v} G_v \, c_{v,l}}{\sum_{v \in c'} w_{u,v} G_v} \quad (13)$$

$$p_{u,l}^{GL} = \beta * (0.5 * \frac{\sum_{k \in l'} w_{l,k} \, GH_k c_{u,k}}{\sum_{k \in l'} w_{l,k} GH_k} + 0.5 * \frac{\sum_{k \in c'} w_{l,k} \, GH_v c_{u,k}}{\sum_{k \in c'} w_{l,k} GH_k}) + (1 - \beta) * GH_l \quad (14)$$

For equation (12)(13), only global influence term is added from equation(1) and(6) respectively, however for POI-based CF we add not only global influence of similar POIs' but also global influence of its own $GH_l$ we choose β as 0.9 in both dataset . Hence we reflect the own popularity of POI *l*.

Finally, we rank POI *l* for user *u* as following score.

$$p_{u,l} = \alpha \, \frac{p_{u,l}^{U}}{p_{u,max}^{U}} + \beta \, \frac{p_{u,l}^{F}}{p_{u,max}^{F}} + (1 - \alpha - \beta) \, \frac{p_{u,l}^{L}}{p_{u,max}^{L}} \qquad (15)$$

$$p_{u,l}^{G} = = \alpha \, \frac{p_{u,l}^{GU}}{p_{u,max}^{GU}} + \beta \, \frac{p_{u,l}^{GF}}{p_{u,max}^{GF}} + (1 - \alpha - \beta) \, \frac{p_{u,l}^{GL}}{p_{u,max}^{GL}} \qquad (16)$$

Where equation (15) means UCPCF (User-based, Friend-based and POI-based Collaborative Filtering) equation(15) means UCPCF-G (User-based, Friend-based and POI-based Collaborative Filtering with Global Influence) and $p_{u,max}^{\cdot}$ is used for normalization. We set α=1/3, $\beta = 1/3$ for Foursquare and $\alpha = 0.7$ $\beta = 0.1$ for Yelp.

# Chapter 4

# Experiment

In this section, we describe our dataset in section 4.1, describe comparison method in section 4.2. Furthermore we describe evaluation metrics in section 4.3, experiment on real dataset in section 4.4 and finally we show impact of each effect in section 4.5.

## 4.1. Data Description

We perform experiments on real world dataset (Foursquare and Yelp ).

**Table 3 Data Statistics**

| | Foursquare | Yelp |
|---|---|---|
| # of users | 4,163 | 96,193 |
| # of POIs | 119,896 | 74,195 |
| # of check-ins | 481,420 | 1,494,819 |
| # of user-user relationship | 32,512 | 349,615 |
| Time span | Dec 2009-Jul 2013 | Oct 2004-Dec 2015 |

**Foursquare.** A widely used LBSNs. The Foursquare dataset we used in our experiment is obtained from[20]. It consists of 481,420 check-in data at 119,896 POIs of 4163 users. Check-in data of this dataset spread worldwide. It contains social network information as well as other basic information such as check-in POIs, POI-address, POI-content, and check-in time. The time span of this dataset is between Dec. 2009 and Jul. 2013. We separate the train set and test set by time sequence and we divide it as 80% for train, 20% for test.

**Yelp.** One of the most widely used LBSNs with Foursquare. This dataset is available from Yelp website (www.yelp.com). It consists of 1,494,819 check-in data at 74,195 POIs by 96,193 users. It contains social relation, POI-address, POI-contents. In order to utilize rich information, we remove users whose check-in frequency is less than ten. Hence total users are reduced to 26,635. We separate the train set and test set by time sequence and we divide it as 80% for train, 20% for test.

Figure 2, represents the check-in distribution of both Foursquare and Yelp. Foursquare check-in record is spread worldwide, but Yelp record is spread in only few parts of U.S. Since dataset size of Yelp is bigger than Foursquare, it indicates that Yelp dataset is much denser than Foursquare dataset

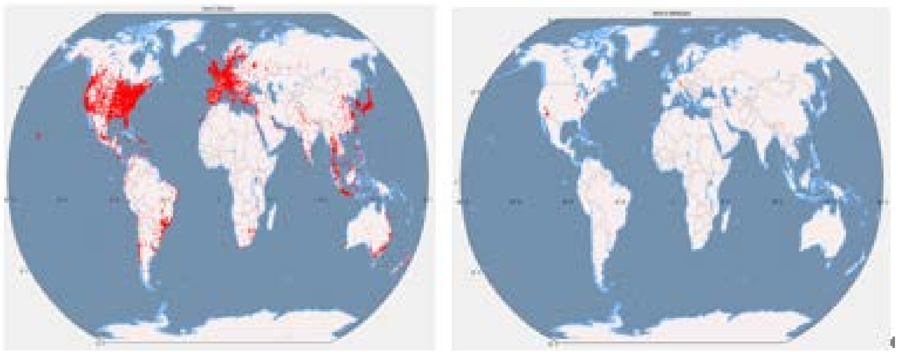

**Figure 2 Check-in Distribution of Foursquare(Left) and Yelp(Right)**

## 4.2. Comparison Method

To comparatively demonstrate the accuracy of our proposed models, we compare our UFPCF and UFPCF-G with following six models

- **PMF[4]:** Basic matrix factorization model, it factorize user-location rating matrix R to two lower dimensional user-specific matrix and location-specific matrix. Then using Gradient Descent to predict estimated user-location matrix.

- **SoReg[5]:** SoReg is very well known model. SoReg use matrix factorization, however it additionally uses influence of social network by adding a social regularization constrain during predicting user-location rating matrix.

- **ASMF[10]:** ASMF uses neighbor friend, similar friend and social relationship as friend set. In addition it introduces potential location which is visited by above mentioned three types of friends. It proposes the augmented square error based matrix factorization model and uses distance information when recommend a POI at last.

- **TACF[9]:** TACF is a collaborative filtering POI recommendation model based on temporal and spatial information combined. TACF use temporal effects as split hour-based slots, and compute temporal preference of a given user in time slots depending on his or her visited POIs in a time slot. For spatial information TACF utilizes distance between visited POIs and recommending POI. The longer distance is, the lower spatial preference a user has.

- **USG[13]:**USG is a collaborative filtering POI recommendation model based on user's similarity, social information and distance information. The difference of social similarity with other models is that, it combine POI similarity and intersection ratio of two users' friend sets to measure similarity between two users in social dataset. Then also uses user's visited POIs to current POI distance to measure preference

- **UPS-CF[14]:**UPS-CF is also a collaborative filtering POI recommendation model like TACF. UPS-CF uses user similarity and social information to improve the accuracy of recommendation by computing the preference of users based on check-in history of user and user's social friends.

During experiment we test two proposed models UFPCF and UFPCF-G models. The model without global influence is defined as UFPCF and with global influence is defined as UFPCF-G

## 4.3. Evaluation Metric

We adopt two typical metrics named Precision@N and Recall@N to compare our proposed models with other baseline models. Where N is the number of recommendations. We set N as 10 to 100. Given a specific user we calculate Precisioin@N and Recall@N as follow

$$Preision\ @\ N = \frac{|\{top\ N\ recommendations\} \cap \{true\ items\}|}{|\{top\ N\ recommendations\}|}$$

$$Recall\ @\ N = \frac{|\{top\ N\ recommendations\} \cap \{true\ items\}|}{|true\ items|}$$

For both of *precision* and *recall* the higher value means it makes the better recommendation.

## 4.4. Experiment Result

In this section, we compare our proposed models with comparison method. Precision on Foursquare and Yelp is presented at figure 3 and figure 4 respectively. According to figure 3(Foursquare), our proposed models significantly outperform other baseline models. PMF shows the worst performance this is due to simplicity. PMF only focuses on observed user-POI matrix without considering any other effects. SoReg seems a little bit better than PMF because social information enrich the performance. However using solely social information is not enough to make a good recommendation. ASMF outperforms PMF and SoReg this is because, it further uses distance information and potential POIs. TACF uses spatial effect and 24-hours slot based CF, however compare to UPSCF and USG, 24-hours slot based similarity makes the model overfitting. USG seems exceeds other models except UFPCF and UFPCF-G, this is because it corporates users' preference, social influence and spatial effect well. In addition, UFPCF-G makes better performance than UFPCF, this result explains global influence is essential. In figure 4 we remove TACF since Yelp dataset does not provide specific check-in time. The performance shape is similar however, interestingly UPSCF outperforms USG and UFPCF. This is due to characteristics of Yelp dataset. Unlike Foursquare, it spread in only specific regions of U.S. hence similar user plays significant role in Yelp Dataset. UPSCF only considers similar users and social relations, so it seems better than other models except UFPCF-G.

In terms of recall figure 5 and figure 6, the shape is like precision and UPFCF-G outperforms other models significantly. Interesting point is that, in figure 5, matrix based approaches TACF and UPFCF outperforms UPFCF-G at recall @ 30.

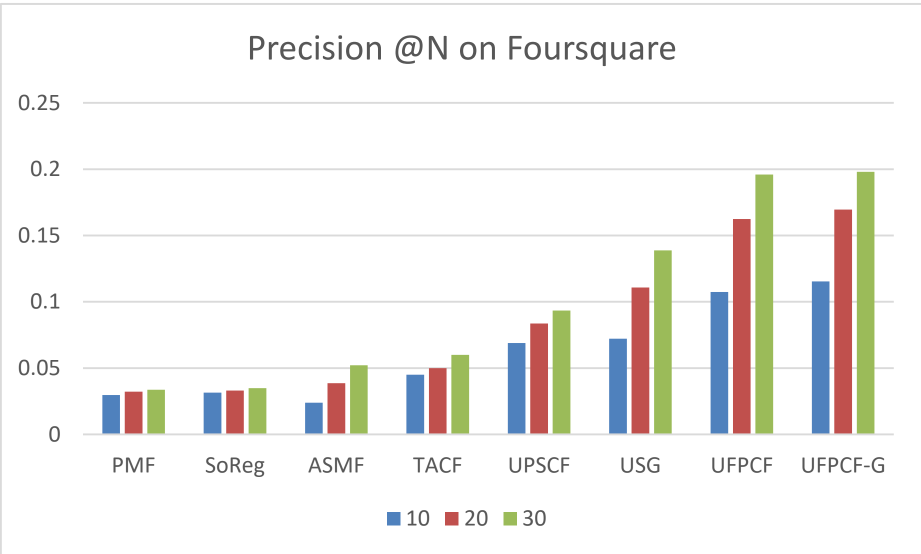


**Figure 3 Precision on Foursquare**

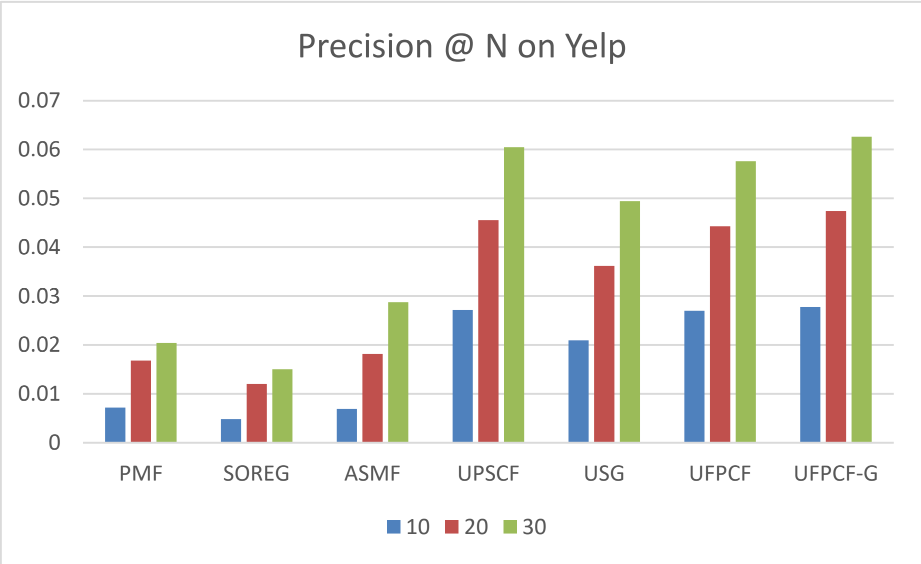


**Figure 4 Precision on Yelp**

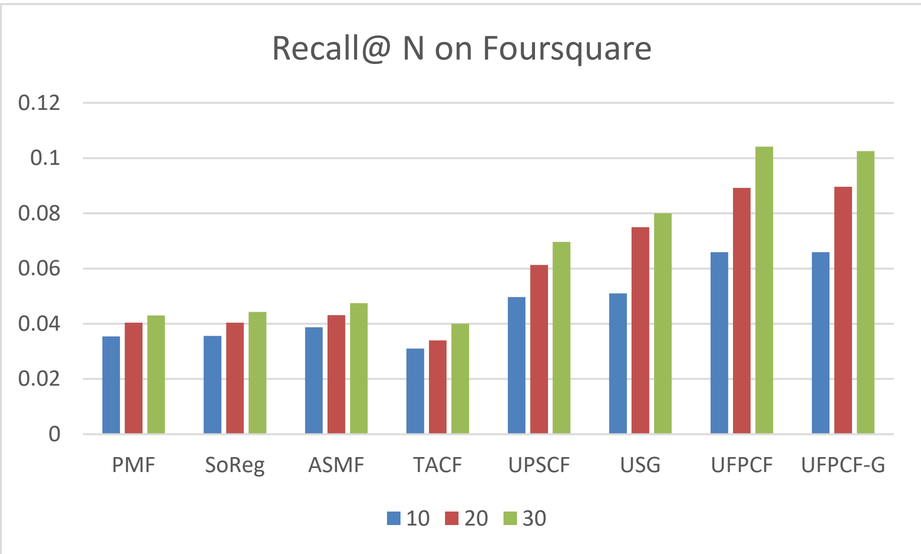


**Figure 5 Recall on Foursquare**

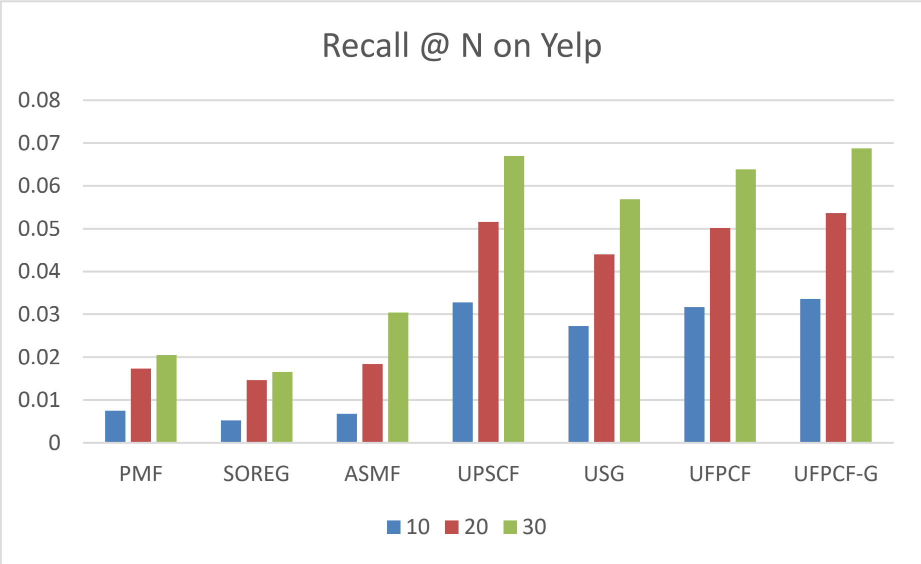


**Figure 6 Recall on Yelp**

## 4.5. Impact of Effect

In this section, in order to catch the improvement gap of each effect, we will represent the result without one effect. Since the tendency in UFPCF and UPFCF-G similar, we show the performance of UFCF, UPCF, FPCF, which loss POI-based CF, Friend-based CF and User-based CF separately. According to figure 6, UFCF perform worst among them, it indicates POI-based CF plays most significant role in Foursquare dataset. In addition User-based CF and Friend-based CF seems have similar effect of improvement. Unlike figure 7, figure 8 shows FPCF performs worst, it indicates user-based CF is most essential in Yelp dataset. Furthermore, POI-based CF also plays significant role for performance improvement.

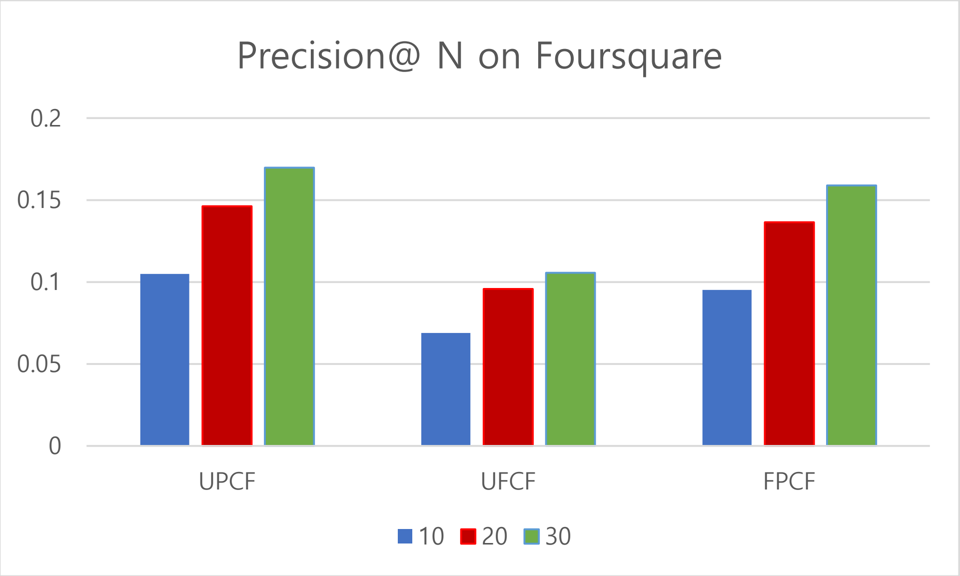


**Figure 7 Without one Effect on Foursquare**

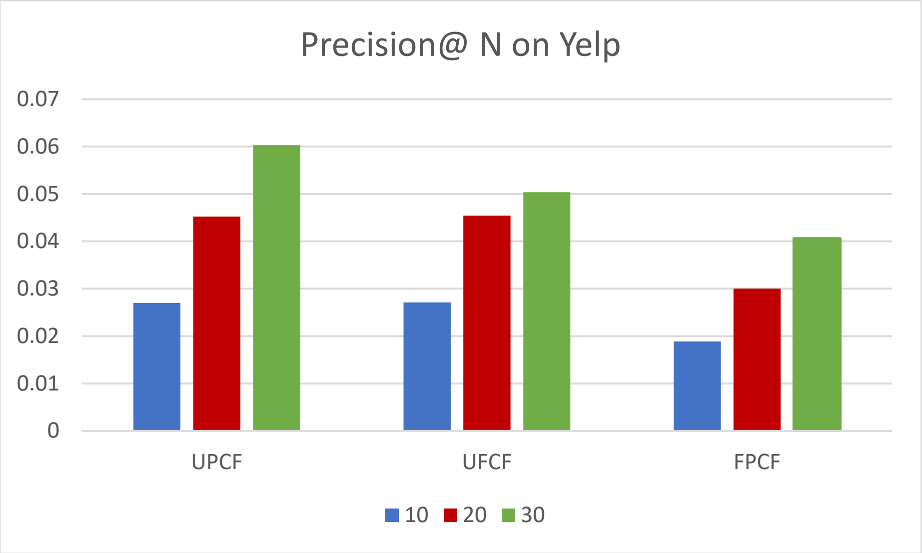


**Figure 8 Without one Effect on Yelp**

# Chapter 5

# Conclusion

In this paper, we introduce a UFPCF-G which efficiently employs user's preference, preference of new types of friends, and POI's similarity with global effect. As our experiment shows, for POI recommender system, similarity between POI plays a significant role. We also executed various experiments to evaluate the performance of our model in real world datasets. We also show that "Core friend" and "Neighbor friend" outperforms "Similar friend" and" Social friend" with regard to matrix factorization.

For future work, it would be really interesting to execute extensive experiments on other dataset. It would be also great to find another new type of friends with plausible reasons.

# Bibliography

[1] Cao, Xin, Gao Cong, and Christian S. Jensen. "Mining significant semantic locations from GPS data." *Proceedings of the VLDB Endowment* 3.1-2 (2010): 1009-1020..

[2] Herlocker, Jonathan L., et al. "An algorithmic framework for performing collaborative filtering." *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 1999..

[3] Lee, Kyunghan, et al. "Slaw: A new mobility model for human walks." *INFOCOM 2009, IEEE*. IEEE, 2009.

[4] R. Salakhutdinov and A. Mnih. Probabilistic matrix factorization. In NIPS, pages 1257–1264, 2007.

[5] Ma, Hao, et al. "Recommender systems with social regularization." *Proceedings of the fourth ACM international conference on Web search and data mining*. ACM, 2011.

[6] Adomavicius, Gediminas, and Alexander Tuzhilin. "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions." *IEEE transactions on knowledge and data engineering* 17.6 (2005): 734-749.

[7] Ma, Hao, et al. "Sorec: social recommendation using probabilistic matrix factorization." *Proceedings of the 17th ACM conference on Information and knowledge management*. ACM, 2008.

[8] Hu, Bo, and Martin Ester. "Social topic modeling for point-of-interest recommendation in location-based social networks." *2014 IEEE International Conference on Data Mining*. IEEE, 2014.

[9] Yuan, Quan, et al. "Time-aware point-of-interest recommendation." *Proceedings of the 36th international ACM SIGIR*

*conference on Research and development in information retrieval*. ACM, 2013..

[10] Li, Huayu, Yong Ge, and Hengshu Zhu. "Point-of-Interest Recommendations: Learning Potential Check-ins from Friends." KDD 2016.

[11] Yang, Xiwang, et al. "On top-k recommendation using social networks." *Proceedings of the sixth ACM conference on Recommender systems*. ACM, 2012..

[12] Ye, Mao, Peifeng Yin, and Wang-Chien Lee. "Location recommendation for location-based social networks." *Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems*. ACM, 2010.

[13] Ye, Mao, et al. "Exploiting geographical influence for collaborative point-of-interest recommendation." *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*. ACM, 2011.

[14] Ference, Gregory, Mao Ye, and Wang-Chien Lee. "Location recommendation for out-of-town users in location-based social networks." *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*. ACM, 2013.

[15] Tang, Jiliang, et al. "Exploiting Local and Global Social Context for Recommendation." *IJCAI*. 2013..

[16] Chow, Chi-Yin, Jie Bao, and Mohamed F. Mokbel. "Towards location-based social networking services." *Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks*. ACM, 2010..

[17] Horozov, Tzvetan, Nitya Narasimhan, and Venu Vasudevan. "Using location for personalized POI recommendations in mobile environments." *International Symposium on Applications and the*

*Internet (SAINT'06)*. IEEE, 2006.

[18] Liu, Bin, et al. "Learning geographical preferences for point-of-interest recommendation." *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2013.

[19] Yin, Hongzhi, et al. "LCARS: a location-content-aware recommender system." *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2013.

[20] Yin, Hongzhi, et al. "Joint modeling of user check-in behaviors for point-of-interest recommendation." *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*. ACM, 2015..

[21] Scellato, Salvatore, et al. "Socio-Spatial Properties of Online Location-Based Social Networks." *ICWSM* 11 (2011): 329-336.

[22] Davidsen, Jörn, Holger Ebel, and Stefan Bornholdt. "Emergence of a small world from local interactions: Modeling acquaintance networks." *Physical Review Letters* 88.12 (2002): 128701.

[23] Jin, Emily M., Michelle Girvan, and Mark EJ Newman. "Structure of growing social networks." *Physical review E* 64.4 (2001): 046132.

[24] Massa, Paolo. "A survey of trust use and modeling in real online systems." *Trust E-services: Technologies, practices and challenges. Idea Group Inc*(2007): 51-83.

[25] Weng, Jianshu, et al. "Twitterrank: finding topic-sensitive influential twitterers." *Proceedings of the third ACM international conference on Web search and data mining*. ACM, 2010.

[26] Kleinberg, Jon M. "Authoritative sources in a hyperlinked environment." *Journal of the ACM (JACM)* 46.5 (1999): 604-632.

[27] Page, Lawrence, et al. "The PageRank citation ranking: bringing order to the web." (1999).

# 요 약

스마트폰과 태블릿의 보편화로 점점 많은 위치 기반 사회망 네트워크들이 인기를 끌고 있다. 이러한 위치기반 사회망 네트워크 데이터의 접근성이 쉬워지면서, 장소추천 시스템은 학계와 상업에서 많은 주목과 관심을 받고 있다. 장소추천 시스템은 기존의 추천 시스템(영화 혹은 상업 아이템) 들과 같이 데이터가 매우 성기다. 또한, 기존의 추천 시스템과는 달리, 실질 적인 거리 문제를 직면 하게 되는데 이는 장소추천 시스템의 문제를 더 어렵게 만든다. 앞에서 언급한 문제들을 해결 하기 위해 많은 연구자들은 사용자의 취향, 시간 적인 영향, 거리적인 영향, 소셜 정보망 네트워크의 영향 등 다양한 영향을 고려하고 있다. 이번 논문에서 우리는 사용자의 취향, 친구의 취향 장소간의 연관성 및 사용자의 이동 거리와 전역적 영향을 고려한 UFPCF-G 모델을 제안한다. 또한 우리는 기존의 연구에서 사용 하였던 "사회망 네트워크" 에서의 관계만이 친구가 아닌 확장된 의미의 친구를 사용하고 구조와 활동력을 기반으로 하는 전역적 영향 방식을 제안한다. 우린 실제 데이터 셋에 대하여 현존하고 있는 5개의 모델들과 성능을 비교하고 우리가 제안한 알고리즘이 다른 모델들의 여러 가지 성능 평가 방법에서 더 좋은 성능을 보이는 것을 증명한다.

주요어 : 장소 추천 시스템, 협업 필터링, 위치 기반 소셜 네트워크, 소셜 네트워크

학번 : 2015-21275