**Exploiting check-in sequences for collaborative point-of-interest recommendation.**

Shokirkhon Oppokhonov, Seyoung Park

School of Computer Science and Engineering

Kyungpook National University

Daegu, S.Korea

shokirxon@mail.ru, seyoung@knu.ac.kr

**Abstract**

**Combination of mobile and location aware services, such as Yelp, Gowalla and Foursquare have been making changes in the way people interact with physical location of businesses, and the availability of community contri-
buted large volume of check-in data from very active location-based-social-networks(LBSN) makes it easier for location-aware services to serve sufficiently. One of those services is Point-of-interest(POI) recommendation, which recommends places where user has not visited before and may find it interesting to visit. So far several techniques have been implemented for the POI recommendation. Although people movement exhibits similar sequential patterns, so far no existing studies on POI recommendation considered the sequential check-in patterns similarity between users on LBSNs. In this paper we propose a new approach which exploits similar check-in sequential patterns on location recommendations. We propose a new collaborative recommendation model that makes a use of similar sequential patterns for POI recommendation. And conduct an experiment on real-world data set. The result shows that our proposed approach performs better than state-of-the-art user-based collaborative POI recommend-ation methods.**

Keywords: Point-of-interest, Location- based Social Networks,

Sequential check-ins, Directed graph.

**I. Introduction**

Information Technology, advanced of mobile devices and location acquisition technologies are the most rapidly growing industry in the world. With these growings LBSNs (e.g., Foursquare and Gowalla) have attracted millions of users and location-aware services [1]. User can post his/her visited location in a form of ‘check-in’(See “Fig.1”) on LBSNs, and even they can leave comments or tips for point-of-interests(POIs) such as coffee shop, restaurant, museum. By the end of 2015 Foursquare had 7 billion check-ins by 55 million monthly active users. And the availability of large volumes of community contributed check-in data on LBSNs enables us to improve the location aware services, like POI recommendation. Objective of POI recommendation is to discover and recommend new places where user has never checked-in before. Several techniques have been implemented to improve the POI recommendation. Among the implemented techniques colaborative filtering has always performed well for POI recommendation according to [2,3].



Fig. 1: An example of check-in on LBSN

User-based collaborative filtering method relies on similar users, and recommend POI visited by similar users but has not been visited by current user. Since user-based collaborative filtering relies on similar users when recommending new POI to the user, our task is to discover the most similar users. In our work we intend to improve the user-based collaborative filtering by utilizing an important factor that has not been considered in previous works. Even though human check-in sequence patterns have a high degree of freedom, they also expose almost the same patterns [4]. As people move back and forth between their homes and workspaces they exhibit periodic behavior in their daily movements [5, 6]. Different people have different daily routines, but it doesn’t mean that theres no same or similar routines. Example of eraly bird(EB)(who gets up early and goes to bed early) and night owl(NO)(who gets up late and goes to bed late). EB and NO may visit the same POI but in different sequence. EB check-ins at the Restaurant at the lunch time, then coffee shop and ends up at the park or cinema. On the other hand, NO may skip the lunch and drop by coffee shop on his way to work, then restaurant in the evening and ends up at the bar or club. Therefore to find the most similar users, the sequence of check-ins play an important role. In order to find the most similar users, we depict the users’ check-in sequences as a directed graph and refer to the graph for similarity between users.The contributions of our work are as follow:

●Define a new directed graph based POI recommendation problem.
●We sort users’ check-ins according to date and time.

●Depict each user’s check-in sequences made within a day separately and merge them in order to get each user’s complete directed graph for user similarity computation.

●Conduct experiment on a real-world LBSN dataset and demonstrate that users’ check-in sequence has significant influence and proposed model performs better than previous works.

**2. Related work**

In this section, we introduce related works on POI recommend-

ation. So far many approaches have been proposed to improve the POI recommendation system. Namely Social influence, Geographical influence, Collaborative filtering tech- nique and Sequential influence.

**Social influence.** Friends on LBSNs have been widely exploited [4, 2, 7] to improve the POI recommendation system. Most of friends on LBSNs may live far away from each other in different part of the country or city and may do not check-in at the same POI. In earlier works [8,9] they proposed social network-based methods for item recommendation. These works assume that friends have similar taste and show similar behavior patterns in social networks. And according to [7] only 20.6% of pairs of LBSN friends co-check-in at POIs but 80.3% of their visited POIs are semantically similar.

**Geographical influence**. As the Tobler’s First Law of Geography, “ Everything is related to everything else, but near things are more related than distant things”. In LBSNs, POI recommendation is distinct from other item recommendations such as movies on Netflix, books on Amazon. Because POIs require physical interaction to visit POIs. In [10] to derive the probability of a user visiting a new POI they utilized the geographical information(latitude and longitude) of adjacent check-ins made by the user. In [2, 3] they model the distance between user’s adjacent check-ins by power-law distribution to compute the check-in probability of candidate POIs.

**Collaborative filtering(CF)**. The premise key point of CF is that if two users show similar actios on the same items then they are more likely to act similarly on the other items, too. CF is widely adopted technique for recommendation system and has always performed well for POIs recommendation. So far many CF POI recommendation methods ([2, 3, 11]) have been proposed. They proposed a CF model to incorporate temporal informat- on in work [3] and friend-based CF in [2].

**Sequential Influence**. As discovered in earlier studies[12, 13] human movement exhibits sequential patterns and they made a use of it for POI recommendation, and also different sequential mining techniques [14] have been proposed for next POI prediction. Users’ check-in sequences are derived from their history to compute the probability of visiting next candidate POIs in [10], which is closer to our work but only in term of check-in sequence mining. We employ the users’ check-in sequences to discover the most similar users. In work [10] they poved that the probability of visiting to the new POI is derived not only from the latest visited POI but also all the visited POIs.

**3. Sequential of user check-ins**

In section 3.1, we provide a brief introduction to the baseline user-based CF method which will be our proposed method’s fundament. We depict users’ check-in sequences as a directed graph and calculate the user similarity weight between users in section 3.2. Then we calculate the score that a user will check-in a given POIs by user-based collaborative filtering in section 3.3.

All the notations that we have used in this paper are listed in Table 1.

Table 1: Notations

|  |
| --- |
| Symbol Definition. |
| $U, L$ user set, POI(location) set$u,v,l$ users $u, v∊U. $POI,$l$ ∊ $L.$$w\_{v,u}$ similarity weight between users $v and u.$$w\_{v,u}^{(l\_{i}l\_{j})}$ check-in sequence similarity.$c\_{u,l}$ the frequency check-in vector of user $u$ over $L$. $c\_{u,l\_{i}l\_{j}}$ frequency of occurring of path in check-in history. |

**3.1 User-based Collaborative filtering**

Given a user and his check-in history, user-based collaborative filtering utilize the similar users who have co-checked-in at same POIs with our user, and recommend the POIs which are exist in similar users check-in history but have not been visited by our user yet. Let $L $and $U $denote the POI set and the user set in an LBSN. If the user $u$ checked-in $l$ before, we set it as $C\_{v,l}=1$ otherwise $C\_{v,l}=0$. This user contributed implicit data can be used to calculate the similarity between users and the probability of user $v$ visiting POI $l$. We denote this predication by $C\_{u,l}$ and the probability score computed by following equation.

$C\_{u,l}=\frac{\sum\_{v}^{}w\_{u,v}C\_{v,l}}{\sum\_{v}^{}w\_{u,v}}$ (1)

where $w\_{u,v}$ is the similarity weight between users $v$ and $u$. The similarity weight $w\_{u,v}$ between users $v$ and $u$ can be computed by different measures. Such as:

 **Jaccard Similarity**: Its typically used when we don’t have numeric rating but just a Boolean value like a POI has been visited or not.

**Pearson Similarity**: is the pearson coefficient between the two vectors.
**Cosine Similarity**: is the cosine of the angle between the two vectors of user $u$’s and $v$’s check-in history. As the vectors become similar, smaller angles and larger the cosine.

Among these measures cosine similarity performs better than the others in our work, so we use cosine similarity(2) to measure the similarity weight between users. Superiority of the cosine similarity measure over the others is that it can take the advantage of check-ins frequency when calculating user similarity.

$w\_{u,v}=\frac{\sum\_{l}^{}c\_{u,l}c\_{v,l}}{\sqrt{\sum\_{l}^{}c\_{u,l}^{2}}\sqrt{\sum\_{l}^{}c\_{v,l}^{2}}}$ (2)

**3.2 Generation of directed graph discoverig similar users**

In this section, we sort users check-in history according to date and time of the day. All location sequence of user $u$ is denoted by

 $S\_{u,day\_{1}}$=($l\_{a}t\_{1}$), ( $l\_{b}t\_{2}$), ($l\_{c}t\_{3}$), ($l\_{d}t\_{4}$), …...( $l\_{n}t\_{n}$)

$S\_{u,day\_{2}}$=($l\_{a}t\_{1}$), ( $l\_{c}t\_{2}$), ($l\_{b}t\_{3}$), ($l\_{d}t\_{4}$), …...( $l\_{n}t\_{n}$)

:

$S\_{u,day\_{n}}$=($l\_{a}t\_{1}$), ( $l\_{b}t\_{2}$), ($l\_{c}t\_{3}$), ($l\_{b}t\_{4}$)…...( $l\_{n}t\_{n}$)

where $t\_{1}$ ≤ $t\_{2}$ ≤ $t\_{3}$ within the same day.

When we depict the user $u$’s check-in sequences of each day separately as a location-location graph and merge them, it will look like in (Fig.2)



Fig 2: check-in sequence of user $u$.

 The directed graph $G=(L, E)$ consist of a set of nodes $L$ and edges $E⊆L×L$, each node $l\_{i}∊L$ denotes a location and each edge $(l\_{i},l\_{j})∊E$ denotes $l\_{i}⟶l\_{j}$ associated with a transition count.

 We assume that if two users check-in sequences are similar then their next visiting POIs also will be similar. But due to the data sparsity(users don’t check-in in LBSNs every time they visit at POIs) its hard to find the similar users by comparing the users’ whole day check-in sequences. So to tackle data sparsity problem, we make a use of edges($l\_{i}\rightarrow l\_{j}$) in users’ check-in sequences to compare the graphs(Fig.3) for user similarity weight. There are many directed graph similarity computition techniques have been provided in [15]. But we prefer using cosine similarity by taking edges($l\_{i}\rightarrow l\_{j}$) as unique ids, and if the user $u$ has gone through the path ($l\_{i}\rightarrow l\_{j}$),

$c\_{u,l\_{i}l\_{j}}=f$(frequency of occurring of path in user’s check-in history) instead of just giving 1, otherwise $c\_{u,l\_{i}l\_{j}} $= 0. By using the frequency we can derive usre’s preference, too.

$w\_{u,v}^{(l\_{i}l\_{j})}=\frac{\sum\_{l\_{i}l\_{j}}^{}c\_{u,l\_{i}l\_{j}}c\_{v,l\_{i}l\_{j}}}{\sqrt{\sum\_{l\_{i}l\_{j}}^{}c\_{u,l\_{i}l\_{j}}^{2}}\sqrt{\sum\_{l}^{}c\_{v,l\_{i}l\_{j}}^{2}}}$ (3)

**3.3 Recommendation**

Given a user $v$, and a candidate POI $l$, the recommendation score that user $v$ will visit the POI $l$ is calculated by following equation.

$C\_{v,l}=\frac{\sum\_{u}^{}w\_{v,u}^{(l\_{i}l\_{j})}C\_{v,l}}{\sum\_{u}^{}w\_{v,u}^{(l\_{i}l\_{j})}}$ (4)

The check-in score for each candidate POI is calculated by above equation and returned as a top ranked POIs as a recommendation result.



Fig 3: Comparision of check-in sequences of user $u$ and $v$.

Table 3: Methods for comparision

|  |
| --- |
| (U) User-based CF.(UT) U with temporal preference.(US) U with sequential check-ins. |

Table 2 : Statistics of dataset

|  |
| --- |
| Number of users 2,321Number of locations (POIs) 5,596Number of check-ins 194,108Density of training data 6.35×$10^{-3}$ |

**4.Experiment**

**4.1 Real-world data set**

We used the data set from [3] which is collected from Foursquare, made within Singapore between August, 2010 and July, 2011. (see Table 2).

**4.2 Evaluation of recommendation technique**

 **Recommendation accuracy:** The accuracy of recommend-

ation system is depends on how many recommended locations to a user are visited in testing data by the same user.

 We employ Precision(5) and Recall(6) metrics to evaluate the accuracy of our system. And F-measure(7) as well wich is the harmonic mean of precesion and recall.

$Precision=\frac{number of discovered locations in a test data}{number of recommended locations}$ (5)

$Recall=\frac{number of discovered locations in a test data}{number of visited locations in a test data}$ (6)

$ F=2×\frac{precision×recall}{precision+recall}$ (7)

**4.3 Performance of our method**

 We compare the accuracy of our proposed method with previously proposed user-based collaborative filtering methods. Methods are reported in (Table 3).

 In terms of Pre@5, Pre@10 , Pre@20 our method (US) outperforms the user-based CF baseline (U) and the work[3]’s first part (UT), wich makes a use of CF that incorporates the temporal information. The comparision result is depicted in

Fig 4.

In terms of Rec@5, Rec@10, Rec@20 (US) performs poorer than (UT) but better than (U). In Fig 5.

In terms of F-measure@5, F-measure@10, F-measure@20 we observe that (US) outperforms both (UT) and (U). In Fig 6.

**5. Conclusion**

The availability of large volumes of community contributed check-in data on LBSNs makes it possible to to improve the POI recommendation system. In light of this data, we proposed a new method by exploiting users’ check-in sequences to find the most similar users to refer when recommending new POIs.

Our propesd method achived the best result so far comparing to the other user-based CF based works. We observed that sequential check-ins play an important role in computing the user similarity weight.

We have three directions for future study.

● Incorporating similarities of seasons and weekdays to our method.

● Depicting users’ check-in sequences according to seasons and weekdays.

● Referring to the users’ check-in sequences when recommending POIs.



Fig. 4: Performance of methods in term of Precision.



Fig. 5: Performance of methods in term of Recall.



Fig. 6: Performance of methods in term of F-measure.

ACKNOWLEDGMENT

 This work was supported by BK21 Plus project (SW Human Resource Development Program for Supporting Smart Life) funded by the Ministry of Education, School of Computer Science and Engineering, Kyungpook National University,

Korea (21A20131600005).

**References**

[1] Y. Zheng Location-based social networks: Users. In computing with spatial Trajectories. Springer, 2011.

1. M. Ye, P. Yin, W. -C.Lee. Exploiting geographical influence for collaborative point-of-interest recommendation. In SIGIR, pages 325-334, 2011.
2. Q. Yuan, G. Cong, Z. Ma, A. Sun, and N. Magnenat-Thalmann.
Time-aware point-of-interest recommendation. In SIGIR, pages 363-372, 2013.
3. E. Cho, S. A. Myers, and J. Leskovec. Friendship and mobility: User movement in location-based social networks. In ACM KDD, 2011.
4. N. Eagle and A. Pentland. Eigenbehaviors: Identifying structure in routine. Behavioral Ecology and Soc., 2009.
5. Z. Li, Bi. Diang, J. Han, R. Kays, and P. Nye. Mining periodic behaviors for moving objects. KDD ’10, 2010.
6. B. Hu and M. Ester. Social Topic Modeling for Point-of-Interest Recommendation in Location-based Social Networks. In IEEE, 2014.
7. H. Ma, D. Zhou, C.Liu, M. R. Lyu, and I. King. Recommender systems with social regularization. WSDM, 2011
8. X.Yang, H. Steck, Y.Guo, and Y.Liu. On top-k recommendation using social networks. RecSys, 2012.
9. J. Zhang, C. Chow, Y. Li. LORE: Exploiting Sequential Influence for Location Recommendations. In ACM SIGSPATIAL, 2014.
10. H. Huang and G. Gartner. Using Context-Aware Collaborative Filtering for POI Recommendations in Mobile Guides. In Springer, 2012.
11. J.-D. Zhang, G. Ghinita, and C.-Y.Chow. Differentilly private location recommendations in geosocial networks. In IEEE MDM, 2014.
12. L. Yanchi, L. Chuanren, L. Bin. Unified Point-of-Interest Recommendation with Temporal Interval Assessment. In ACM KDD, 2016.
13. J. J.-C. Ying, W.-C. Lee, T.-C. Weng, and V. S. Tseng. Semantic trajectory mining for location prediction. In ACM SIGSPATIAL , 2011.
14. R. Sedgewick, K. Wayne. Algorithms. Addison-Wesley Professional; 4th edition, 2011.