POINT OF INTEREST RECOMMENDATION ENGINE

T. Sakthi Sree¹, S. Abirami², S. Gokul Raj³, S. Sivasankari⁴

¹Assistant Professor, Department of Computer Science and Engineering, Kathir College of Engineering, Coimbatore, Tamilnadu, India.
²Student, Department of Computer Science and Engineering, Kathir College of Engineering, Coimbatore, Tamilnadu, India.

Abstract—The popularity of location-based social networks (LBSNs) has led to an enormous amount of user-based check-in data. Recommended Points of Interest (POIs) plays a key role in meeting the needs of LBSN users. While recent work has explored the thought of adopting a collaborative ranking (CR) for recommendations, few attempts are made to include time-based information for POI recommendations using CR. In this article, we propose a two-phase CR algorithm that comes with the geographical influence of POIs and is regularized supported the variance of recognition of POIs and user activities over time. Time-sensitive regularized penalizes users and POIs that have been more time-sensitive in the past, helping the model to account for long-term behavioural patterns while learning from user-POI interactions. Moreover, in the first phase, it attempts to rank the visited POIs higher than the unvisited ones and, at the same time, to apply the geographical influence. In the second phase, our algorithm attempts to rank the preferred POI users higher on the recommendation list. Both phases use a collaborative learning strategy that enables the model to capture complex latent associations from two different perspectives. Real-world dataset experiments show that our proposed time-sensitive collaborative ranking model beats the state-of-the-art POI recommendation methods.

Keywords — Point-of-interest recommendation, time-aware recommendation, collaborative ranking, location-based social networks.

I. INTRODUCTION

With the launch of location-based social networks (LBSNs) such as Yelp, Trip Advisor and Foursquare, users can share check-in data on their mobile devices. LBSNs collect valuable information about mobile user records with check-in details. Generating Points of Interest (POIs) guidelines play a key role in addressing user needs, such as exploring a new POI or visiting a city. In reality, every city has multiple POIs, and a user may have visited just a few in her hometown as well as when out of town. POI Recommendation attempts to ensure the satisfaction of users by proposing the most interesting locations in their vicinity, taking into account their preferences and contextual constraints.

The accuracy of POI recommendation is constrained by many challenges. For example, data scarcity is a major challenge in Recommendation of the POI. Despite the fact that there are LBSNs with a large number of locations, in practice users visit a very limited number of locations, making the user-item matrix sparse. Moreover, as users spend most of their time in their hometown, the data scarcity problem is aggravated when a user visits a new city where has no history of visited locations. Some studies seek to address the question of data scarcity by integrating additional data into the model, such as geographical and temporal data. The data scarcity problem is actually even worse when suggesting POIs as opposed to other things like movies or songs. This is mainly due to the fact that the check-in data provides an implicit feedback whereas users usually express their
opinions on movies or songs with different ratings. Argued that check-ins offer only positive examples, whereas non-check-in POIs remain undiscovered and that could potentially be of interest to the user. They claim, however, that if a user has only visited a place once, they can not necessarily infer a positive feedback from, but we can also infer the types of locations that the user is interested in. For places with higher check-in frequencies, we assume that more than are preferred than those with fewer check-in frequencies.

Recommendation of products is often viewed as a ranking prediction or matrix completion function in specific literature. However, argued that the square loss as is not an accurate measure of the effectiveness of predictions. In other words, a user should be rewarded for being able to produce a more reliable rated list. Collaborative Ranking (CR) is based on this concept and focuses on recommendations being correct at the top of the recommendation list for users each. A lot of work has taken on CR using clear user feedback such as element ratings. For example, and optimized the loss of ranking to recommend movies to users has shown that ranking-based recommendation learning is also successful when dealing with implicit feedback from users. Exploring the CR recommendation for POI using implicit feedback, however, is difficult since aspects of the learning process require an effective strategy for sampling unvisited places based on user versatility and POI proximity.

Several studies have begun integrating temporal information in order to improve POI recommendations. For examples, temporal information was integrated hourly in, and sequentially in to suggest the "latest" POI. Further advanced models also found the time as well as the sequential order of check-ins. However, it is still necessary to analyze the long-term dynamics of the user’s check-in behaviour and popularity of venues. For example, an open-air bar is predominantly popular during the summer months of, or a student is supposed to be more active during holidays.

In this post, we propose a two-phase POI CR algorithm. Our model is inspired by positive CR findings with explicit user feedback in other domains and ranking approaches using implicit user feedback for POI recommendation. In our algorithm we use a two-phase implicit feedback inference. In fact, we conclude that one single check-in means that a user "likes" the POI and multiple visits means that the user loves the POI more. On the basis of this assumption, we push POIs with single or multiple check-ins to the top of the recommendation list in the first phase, taking into account the geographical influence of POIs in the same neighbourhood. In the second phase we force POIs over those with a single visit with multiple check-ins. As argued in the learning consideration of both visited and unvisited POIs, this alleviates the scarcity problem. Thus, the first step of our algorithm solves the scarcity issue, while the second phase improves our model's accuracy by moving more important POIs to top of the list. To take user dynamics into account, and to position check-in, we add a time-sensitive regularize in the loss of the ranking. The regularize models each user's and venue's behaviour pattern over time. Adding this regulating parameter to the objective function fuses the patterns of activity in a collaborative way into the ranking function.

For summary, we can summarize our contribution in this article as follows:

- In order to demonstrate the underlying trends of choice and popularity over time, we conduct an extensive analysis.
- They suggest a typically time-sensitive regularized, taking into account the variation in the prevalence over time of consumer behaviours and locations.
- We are proposing a new, two-phase CR-based POI recommendation algorithm that combines implicit check-in feedback from users with an emphasis at the top.
- They suggest a measure of geographic similarity, and apply its effect to the objective function of the model.

Two benchmark datasets experiments show that the proposed approach outperforms state-of-the-art recommendations for POIs and CR methods. In particular, we demonstrate that the joint learning
approach helps the model not only to address the problem of scarcity but also to rank higher specific POIs. The first section primarily addresses the problem of scarcity by introducing the geographical impact and considering both visited and unvisited training sites. The second phase increases the consistency of the model, placing the POIs that users prefer more at the top of the recommendation list. In addition, we show the time-sensitive secularizer’s efficient which is extended to both phases of the algorithm taking into account users' long-term behaviour and POI popularity.

The rest of the article is structured as follows, Section 2 briefly discusses the related work while an in-depth analysis is conducted on the datasets in Section 3. In Section 4 and in Section 5, we explain our process, comparing our model's output against competitive models. Finally, the article is concluded under section 6.

II. RELATED WORK

The related work can be divided into the following topics: recommendation for POI, collaborative ranking, and recommendation for time-conscious purposes.

TRADITIONAL RECOMMENDATION ALGORITHM

In the 1990s, research on personalized recommendation systems was proposed. In general, the recommendation methods are often basically divided into four types: collaborative filtering, content-based recommendation, Location-based social network recommendation and mixed recommendation.

Content based recommendation. Content-based recommendation is a recommendation mechanism widely used in recommendation engines. By analyzing metadata of project content, it recommends to users the things with similar metadata information and user preference. The usual content based recommendation process is: project keyword extraction, project to project similarity calculation, and recommendation using Item CF, although this recommendation algorithm can relatively well model the user’s behaviour preferences, it also has certain drawbacks. For example, common problems are: missing information, cold start and classification, and labels are difficult to control.

Location-based social network recommendation. Location-based social network recommendations (LBSNs) added the situation to the normal social network. The location-based social network in is introduced as follows: additionally, to the situation attribute added to the social network, the LBSNs lets the user to share the location. More importantly, the location information shows the location of the user at a specific time, and can also reveal User’s historical access record.

III. ALGORITHMS

1. AGGLOMERATIVE ALGORITHM

An agglomerative algorithm is one such hierarchical clustering algorithm where each individual element to be clustered is in its own cluster. These clusters are grouped iteratively until all the elements are merged to form one big cluster. It is bottom-up approach, where initially each data point is treated as a single cluster. These clusters are then joined by taking the two most similar clusters together and grouping them together. For each cluster, you further break it down to two, until you hit the desired number of clusters. This technique is also named Hierarchical cluster analysis or HCA is an unsupervised clustering algorithm. It involves creating clusters that have predominant ordering from top to bottom. For instance the files and folders in our hard disk are arranged in hierarchy.

2. COLLABORATIVE FILTERING.

The collaborative filtering recommendation method is the most widely used recommendation method. It was originally proposed to be applied to the mail filtering system and it can be divided into the following two categories: User-based collaborative filtering (User-based CF) and Item-based collaborative filtering (Item-based CF). But this approach does not take into account the impact to
fgeographic information and Social correlations for points of interest. Moreover, the method of collaborative filtering is only applicable to the case where the amount of data is small. If the amount of data is large, the cost of calculating matrix similarity is large.

3. BASIC SYMBOLS AND DEFINITIONS
This subsection will describe the data structure and basic definition of this paper. These symbols are extracted from the check-in dataset from LBSNs, including users’ historical access information, social relationships between users, and geographic locations in formation. Table 1 lists the key symbols used in this article and their relevant definitions are listed at the same time.

**DEFINITION 1:**
Sign-in record tuple. Once a sign-in record is expressed as \( u, l, t \) which means that the user \( u \) checks in at the point of interest \( l \) at time \( t \). The \( u \in U \), \( U \) is a set of all users in the LBSN and \( l \in L \), \( L \) is a collection of points of interest.

**DEFINITION 2:**
Sign-in record collection. The sign-in collection means that all users \( u_i \) access all POIs \( l_i \) at different time \( t_i \). It is a collection of all user check-in records.

**DEFINITION 3:**
Spatiotemporal sequence collection. The sign-in order of a user \( u \) is expressed as \( S_u = \{l_1, t_1, \ldots, l_n, t_n\} \) where the user \( u \) accesses the POI \( l_i \) at time \( t_i \). It is abbreviated as \( S_u = l_1, l_2, \ldots, l_n \).

**DEFINITION 4:**
Transition, Predecessor, and Successor. Two consecutive POI \( l_i \) and \( l_{i+1} \) and a certain time threshold \( \Delta T \) are given in the spatiotemporal sequence \( S_u = \{l_1, t_1, \ldots, l_n, t_n\} \). If \( t_{i+1} - t_i \leq \Delta T \), the \( l_i \) to \( l_{i+1} \) is a transition, and it is defined as \( l_i \rightarrow l_{i+1} \). The transition predecessor of \( l_{i+1} \) is \( l_i \) and the transition successor of \( l_i \) is \( l_{i+1} \).

**DEFINITION 5:**
Sign-in frequency matrix. Each element in the matrix represents the sign-in frequency of the user \( u (u \in U) \) at the POI \( l (l \in L) \). Therefore, \( R_{u,l} \) means the number of times a user \( u \) accesses a POI \( l \). Because user check-in locations often account for only a small fraction of allocations, most of the elements of the matrix \( R \) are 0.

**DEFINITION 6:**
Social relationship matrix. For two different users \( u (u \in U), u' (u' \in U) \), if \( u \) and \( u' \) are friends, then the friend relationship matrix \( F \) is defined as \( F_{u,u'} = 1 \), otherwise \( F_{u,u'} = 0 \).

**DEFINITION 7:**
Popularity matrix. We think that the number of times user signed in at the location can indicate the attraction of this location \( l (l \in L) \) to user \( u (u \in U) \). So, for a better understanding, although Popularity matrix and Check-in frequency matrix are numerically the same, the meanings are different. \( P_{u,l} \) represents the attraction of POI \( l \) to user \( u \). Notice that, we think the higher the number of the total frequency of all users check-in, the POI has a better popularity, so we define \( P_{u,l} \) as the prevalence of POI \( l \).
IV . POI RECOMMENDATION

A. POPULARITY AND SOCIAL CORRELATIONS ANALYSIS

In the recommendation service of LBSNs, the social relationship between users and the popularity of POI both will influence user’s habits about choice the point of interest to a large extent. Users more like to go to more popular places or some places where friends prefer to go. In this article, we introduced an effective modelling method to take advantage of the impact of popularity and social relationship factors. In order to establish the model of social influence and popularity influence, we first aggregate the social relations of user \( u \) and all users’ check-infrequencies into friend check-infrequencies. Then we model social influence and popularity influence as power-law distribution through friend check-infrequencies and POI popularity to calculate the influence of the mon recommendation. Notethat, as defined in Definition 7, we believe that the popularity of POIs with a higher frequency of sign-infrequencies is higher and if the user checks in frequently at a POI, the POI more attractive to this user.

In addition, we define the popularity influence and social relationship influence as \( F_{pop} \) and \( F_{fri} \) and calculate them by training our model with a large amount of real historical data.

**PowerLaw.** It means that the product of thenumber of connections a node has and the number of such nodes is a fixed value, that is, the geometric average is a fixed value. For example, there are 10 people with $1000, 100 people with $100, and 1000 people with $10. Drawing it in logarithmic Coordinates gives you an oblique downward line.

**Power-Law Distribution.** The form of power-law distribution is \( y = kx^{-r} \), where \( k \) is a constant and \( r \) is the law’s exponent and always greater than zero. Here, we take the logarithm of both sides of the formula, wegetthat \( \ln y = \ln k - r \ln x \). That is to say, on the logarithmic axes, the Power Law Distribution is represented by a straight line with a negative slope, which is also the basis for judging whether the random variable in a given event obeys the *Power-Law Distribution* (PD).

1) Popularity correlations Analysis

Here, we presuppose that the POI popularity \( p \) is subject to the PD. The probability density function is:

\[
f_{pop}(p) = (\beta - 1)(p+1)^{-\beta}
\]

where \( \beta \) can be obtained by applying the Maximum Likelihood Estimate (MLE) in the popularity matrix \( P \), and the calculation formula is:

\[
\beta = 1 + |L|(\ln \sum_{u(u \in U) Pu,l}P_{u,l} + 1))^{-1}
\]

where \( P_{u,l} \) is the popularity of POI \( l \) (DEFINITION 7).

To verify your hypothesis, we get the by training with real dataset Gowalla and substituted the \( \beta \) into Formula 1. After taking logarithm on both sides of the equation and presented it on the logarithmic graph, we get a straight line with a negative slope. After that, we analyzed the real public sign-in dataset Gowalla and obtained the result shown in the graph, which reflects that POI popularity (points in the graph) fits with the power-law distribution (lines in the figure) that we estimated before. This result verifies that the feasibility to model popularity as a PD.
As a result of the POI popularity influence increases as the popularity increases, the cumulative distribution function of $f_{pop}$ is used to obtain the influence of POI popularity ($F_{pop}$), defined as:

$$F_{pop}(P_{u,l}) = \int P_{u,l} f_{pop}(p) dp = 1 - (P_{u,l} + 1)^{1-\beta}$$

where $P_{u,l}$ means the POI $l$’s attraction to User $u$ (DEFINITION7).

### 2) Social correlations Analysis

Similar to the above, for the distribution of the check-in frequency of the friends, we presuppose that the friend sign-in-frequency $z_{0}$ be $\text{PD}$, and its probability density function is:

$$f_{fr}(z) = (\gamma - 1)(z + 1)^{-\gamma} z \geq 0, \gamma \geq 1$$

where $\gamma$ is estimated by applying the Maximum Likelihood Estimate (MLE) in the sign-in-frequency matrix $R$ and the friend relationship matrix $F$:

$$\gamma = 1 + |L||U| (\sum_{u \in U} \sum_{l \in L} \ln(F_{u,u} R_{u,l} + 1))^{-1},$$

where $F_{u,u} R_{u,l}$ is the total sign-in-frequency of user friend on the location $l$.

As before, we first compute the $\gamma$ by training with real data and bring it in to the Formula 4. After taking logarithms on both sides of the equation, we get a straight line with a negative slope. As shown in Figure 2, we compare this line with our analysis of the public sign-in-data set, and prove that the friends’ check-in frequency (points in the graph) also conforms to the power-law distribution (the line in the graph) we assumed before. The results show that our experiment is effective. Because the influence of the friends’ check-in frequency increases as the frequency of the friend’s check-in increases, it is similar to the popularity factor, we obtain the influence of social relations ($F_{fr}$) through the cumulative distribution function of $f_{fr}$, expressed as Formula:

$$F_{fr}(z_{u,l}) = \int z_{u,l} f_{fr}(z) dz = 1 - (z_{u,l} + 1)^{1-\gamma}$$

### B. GEOGRAPHIC CORRELATIONS ANALYSIS

In LBSNs, points to interests are different from other non-spatial items because the user needs to physically interact with the location. Therefore, geographic information (like position coordinates) has a great impact on users’ visit behavior. Some researchers turned to explore how to use geographic information to serve users. One way is that because the near by friends have more commoner places to check in than others, we can use the distance between the user’s social friends are residence places to adjust their similarity weights. However, users of then migrate from one location another, so their home address sometimes cannot reflect their real physical location. Not
only that, the improvement of location recommendation quality by incorporating user residence information is also very limited.

![Figure 2: Userperson check-indistribution](image)

**C. SEQUENTIAL CORRELATIONS ANALYSIS**

User movement activity is, in essence, a space-time sequence which is transferred from \((ln, tn)\) to another \((ln+1, tn+1)\). To take advantage of the user's access pattern knowledge, we first sorted each user's check-in records in chronological order to get a list of Spatiotemporal Series, i.e., \(Su\). Then, we try to mine all users' space-time sequence pattern information and uniformly model the information into a Location-Location Transition Graph \((L 2 T G)\). This model diagram effectively reflects the general pattern of access sequence of all users in the user range. Based on this graph model, we are able to get users' likelihood of position change among all locations. Specifically, this graph model can be used to obtain the likelihood that the user will access \(ln+1/L\) after accessing \(ln/L\). Even if \(ln+1\) is a POI that the user has never accessed, the cumulative access sequence pattern of all users and recently visited location \(ln\) always gives user \(u\) the likelihood of accessing \(ln+1\). Sequence of all users and location \(ln\) recently visited.

However, simply using the above method to obtain the transfer probability is not scientific, because it only considers the influence of the most recently visited spot on the next to be visited, ignoring the impact of the unique historical access sequence of each user. So, to solve one of the defects, we combine the user's own historical access order information with the location transfer graph using the Additive Markov Chain after obtaining the position transfer diagram, and finally obtained a more reasonable location transfer probability. Location-Transition Map. Location-Location Transition Graph \((L 2 T G)\) is a kind of graph model that is generated by combining the graph concept with the access pattern for the user. As shown in Figure, \(L 2 T G\) is composed of a set of vertices and edges directed. The node (circle) represents a POI \(l_i(l_i)\), and the figures in the circle reflect the out-degree of the POI, meaning the number of POI \(l_i\) as predecessor of the transformation, denoted as \(OCount(l_i)\). - guided edge(arrow) is a transition from POI \(l_i\) to POI \(l_i+1(l_i+1 + 1 + 1)\), referred to as \(l_i l_i l_i+1\). The number on each edge represents the transition frequency from POI \(l_i\) to POI \(l_i+1\), denoted as \(T Count(l_i, l_i+1)\). For example, transferring \(l_1\) to \(l_2\) three times as the \(l_2\) transition predecessor and transferring it to \(l_4\) two times as the \(l_4\) transition predecessor, so the \(l_1\) out-degree is 5, that is, \(OCount(l_1) = 5\).
V. EXPERIMENTS AND ANALYSIS

A. REAL DATA SETS
The dataset we used to test and train our framework is openly large-scale real check-in dataset which is crawled from Gowalla between February 2009 and October 2010. The details of the dataset after data cleaning.

B. EXPERIMENTAL ENVIRONMENT
In the experiment, since we must use the past check-in data to predict future check-in events, so the data set is divided into training set and test set according to check-in time, instead of using random partition method. The train set is half of the earlier data, and the other half is the test set (see Table for more information about the dataset). We set the time interval (\( \Delta T \)) in Definition 4 to one day and set the attenuation parameter of Formula 22 to 0.05. The specific reasons will be analyzed later. In the test, we evaluated the Precision and Recall of recommendation techniques, covering the top-k only range from 2 to 20, because we do not think it makes sense to recommend too many locations to users.

C. PERFORMANCE METRICS
There commendation algorithm usually calculates a target user’s preference score for each unvisited location, and then recommends to the user the top-k location candidates by this return score. To evaluate the quality of the commendation algorithm, the most important thing is the ratio of the number of POI actually accessed by the target user to the number of POI recommended by the recommendation algorithm.

D. COMPARISON OF EXPERIMENTAL RESULTS
After considering a series of factors such as sequence factors, geographic factors, friend relationships and popularity of interest points, we proposed the GFP-LORE POI recommendation framework. In order to verify the quality of our proposed framework, we compared our framework with some other advanced recommendation algorithms.

VI. CONCLUSION AND FUTURE WORK
In this article, we propose a novel bridge recommendation algorithm framework called GFP-LORE. We integrate social influence, popularity influence, geographic influence and sequential influence into a unified framework, and prove that this method effectively improves the accuracy of recommendations. Firstly, we prove that the user’s social correlations and POI’s popularity are subject to power-law distribution, and modelled them according to this phenomenon, implemented through social factor and popularity factor recommend a new POI. Then, we analyzed the user’s individual check-in distribution through the user’s check-in history, explore the personalized geographic information in the user’s check-in behaviour, and calculate the probability of the user’s arrival to the new location based on the method of Kernel Density Estimation (KDE). After that, our system minimize these sequence pattern from the check-in data of all users in the form of dynamic \( L^2 TG \) which can...
reflect over all transfer sequence pattern, and derive the probability of the user accessing the new POI based on the Additive Markov Chain (AMC), realize recommend to the user through the user’s history sequence pattern. Finally, we integrate the above four influence factors into a unified recommendation framework, get a unified relevant probability score and recommend new locations to users based on this probability score.

We train and test our algorithms by using the publicly check-in datasets. The final result proves that the commendation accuracy of our GFP-LORE algorithm is better than other recommendation algorithms in the experiment. In future work, we expect to combine the existing recommendation algorithm framework with mainstream Deep Learning algorithms. Extract the keywords in the user’s evaluation of the POI, analyze the emotions contained in them, and add them as an impact factor to the framework to improve the accuracy of the recommendation.

REFERENCES

IV. J. He, X. Li, L. Liao, Category-aware next point-of-interest recommendation via wisdom Bayesian personalized ranking (2017) 1837–1843.
IX. L. Hu, A. Sun, Y. Liu, Your neighbours affect your ratings: on geographical neighbourhood influence in rating prediction (2014) 345–354.