Mobile Networks and Applications (2021) 26:2434–2444
https://doi.org/10.1007/s11036-021-01782-w

PR-RCUC: A POI Recommendation Model Using Region-Based Collaborative Filtering and User-Based Mobile Context

Jun Zeng[1] · Haoran Tang[1] · Yizhu Zhao[1] · Min Gao[1] · Junhao Wen[1]

Accepted: 10 May 2021 / Published online: 8 June 2021

Abstract
Recent years have witnessed the rapid prevalence of big data and it is necessary for mobile application to filter out information for users. As a significant means of information retrieval, recommendation system that recommends a ranked list of items to users according to their preferences has become a key functionality in Location-Based Social Networks (LBSNs). Point of interest (POI) recommendation that aims to recommend satisfactory locations that users may be interested in plays an important role in LBSNs. However, the traditional POI recommendation uses the original user-POI matrix, which faces a huge challenge of data sparsity because most users just check in a few POIs in their phones. Moreover, it is hard for POI recommendation to give reasonable explanations on why user will visit these locations that we recommend. Therefore, in terms of the challenges mentioned above, we propose a new POI recommendation model called PR-RCUC that uses region-based collaborative filtering and user-based mobile context. Firstly, we cluster locations into different regions and enhance the traditional collaborative filtering with region factor. Secondly, we capture the preferences of users on mobile context such as geographical distance and location category. Thirdly, by combing the two parts we present, we finish the final computation of prediction score and recommend Top-K locations to users. The results of experiments on two real-world datasets collected from Foursquare demonstrate the PR-RCUC model outperforms some popular recommendation algorithms and achieves our expected goal.

Keywords Recommendation system · Point of interest · Region · Collaborative filtering · Mobile context

1 Introduction

With the emergence of Location-Based Social Networks (LBSNs) such as Facebook, Twitter and Foursquare, users can search for interesting locations (e.g. restaurants and museums) to visit, share their location to their friends by making a check-in record on the locations they have visited [22]. These check-in data embed abundant hints of implicit preferences of users on locations, which give us chance to provide better recommendation diversity [2]. The application of mobile devices has grown fast in recent years and user preferences have shifted gradually from traditional desktop and laptop computers to portable devices [8]. Therefore, it has been an important and urgent problem to provide personalized and accurate recommendation [5]. Point-of-interest (POI) recommendation that aims to recommend nearby interesting locations facilitates the outdoor activities of users greatly [14]. Therefore, the accurate and personalized POI recommendation is a crucial demand in mobile services since there are massive locations and it is difficult for users to find their preferred ones in an efficient way [15]. Fortunately, utilizing the check-in data users share helps us to explore new satisfactory locations and also bring more benefits to the third-parties like

✉ Jun Zeng
zengjun@cqu.edu.cn

Haoran Tang
tanghaoran@cqu.edu.cn

Yizhu Zhao
zhaoyizhu@cqu.edu.cn

Min Gao
gaomin@cqu.edu.cn

Junhao Wen
jhwen@cqu.edu.cn

[1] School of Big Data, Software Engineering, Chongqing University, Chongqing, China

advertisers [1]. As one of the personalization technologies, POI recommendation has attracted more and more attention from academic and industrial researchers [19]. Hence, it is still worth studying POI recommendation system in terms of its value [10].

There are many popular algorithms or strategies for recommendation system. As we know, deep neural network has been applied in many fields such as mobile and web services [3]. Although deep learning technology has gradually become the mainstream with the wake of developments in artificial intelligence, the traditional technology will not be replaced from the perspective of cost and efficiency. Collaborative filtering (CF) is not only studied in recommendation system extensively but also used in industry widely [26]. It is based on a simple intuition that if users have similar rating records in the past, they are likely to rate new items similarly in the future [24]. The basic CF adopt cosine similarity or Pearson correlation coefficient to find the relationship between two users [12]. However, there is a huge number of users and locations and most users just check in a few locations, which causes the challenge of data sparsity [31]. That means most of the elements in the user-POI matrix are zeros. If recommendation model calculates scores of unknown items by a sparse matrix, the computational overhead will be useless. Matrix factorization seems to be an effective way to solve data sparseness [25]. The weakness of matrix factorization is that it produces results only in terms of mathematical nature and it cannot give reasonable explanations. Compared with a large number of locations, the number of regions which is clustered by locations will be greatly reduced. Moreover, it is possible to mine preferences of users on regions since users always visit regular regions in their daily life.

The key of reasonable explanation is to make full use of user-based mobile context, as it greatly influences the decision of user on visiting a location [18]. It is observed that users prefer to visit locations that are geographically adjacent to their checked-in locations and this geographical characteristic is benefit for advancing the preference ranking model according to geographical distance information [34]. Geographical distance is extremely important in user-based mobile context since users are more willing to visit closer locations [13]. For example, if a user wants to watch a new movie, he may choose a nearby cinema instead of going to one 20 kilometers away from home. For moving objects, distance shows obvious regularity and can be used in various service computing [6]. Moreover, since only modeling the distance distribution may ignore the multi-center characteristics of individual visited locations [17], we could adopt geo-clustering techniques to divide locations into different regions. In addition to the geographic distance, user daily activities usually present category-level transition patterns [9]. For instance, user who loves stage plays may always go to locations that are classified as theater. Content information of locations could be related to a check-in actions of users and provide a unique opportunity for POI recommendation [7]. However, compared with geographical information, the mobile context of location category is still not effectively utilized.

In this paper, we propose a POI recommendation model using region-based collaborative filtering and user-based context, which is called PR-RCUC. On one hand, most users just visit a few locations and that causes data sparsity which has negative effect on collaborative filtering. On other hand, there are many user-based mobile contexts collected from check-ins of users, such as geographical distance and location category. Therefore, we decide to extract region factor since the number of regions is small and it is possible to capture preferences of users on different regions. Meanwhile, we adopt two kinds of user-based mobile contexts to advance the ranking results of POI recommendation. Our biggest innovation is that we capture the preferences of users on regions and explore the combination between region-based collaborative filtering and context-based model. Besides, we find the optimal integration of the geographical distance based a classical probability model and the location category based on a new transition model.In this paper, we also call POI location.

The main contributions of this paper can be summarized as follows.

- First, for the sake of improving performance of CF and alleviating the negative impact caused by data sparseness, we cluster locations into different regions and add region factor into CF to get the balance of individual region and location.
- Second, considering the user-based mobile context is benefit for mining preferences of users effectively, we combine geographical distance with location category to enhance the ranking results of POI recommendation. The computation of geographical distance and location category is based on a probability model and transition matrix respectively.
- Third, we choose two datasets collected from two cities on Foursquare and conduct our experiments. The results demonstrate that under optimal parameters our PR-RCUC outperforms some popular recommendation models, which achieves our expected goal.

The rest of the paper is organized as follows. Related works are reviewed in Section 2. Section 3 describes in detail the design of PR-RCUC. Experiments are presented in Section 4 to analyze PR-RCUC and demonstrate its effectiveness compared to other recommendation algorithms. Finally, the Section 5 gives conclusion of this paper.

2 Related works

There are lots of studies on POI recommendation no matter based on traditional machine learning or deep neural network. Before this paper, we also have done other works and they all achieve ideal results from different perspectives. We first fuse CF with geographical and social influence for the same purpose that is to alleviate data sparseness [30], but clustering locations into regions will be a more suitable method since users always visit some regular regions. Nowadays, neural network has become a popular technology for predicting tasks including recommendation system. The neural network uses its special non-linear way to realize computing, such as cloud [4]. Xue proposes a novel matrix factorization model based on neural network to learn a common low dimensional space for both users and locations [26]. Yang proposes a general and principled framework to solve the sparsity problem by smoothing among users and locations [27]. So we also apply neural network into POI recommendation and present a deep model based on restricted Boltzmann machine and non-negative matrix factorization [31]. Then, in terms of the sequence of check-ins, we present a next location predicting approach based on recurrent neural network and self-attention [29]. Zhao proposes adaptive sequence partitioner with power-law attention to automatically identify each semantic subsequence of POIs and discover their sequential patterns [33]. Deep neural network dose have its own advantages but there is an obvious weakness, that is it cannot give reasonable explanations on results. Considering of that, we argue that traditional technologies will not be replaced, which encourages us to complete the work in this paper. Different from analyzing positive preferences of users, Tran constructs a joint model which combines user embedding, user positive preference embedding and user negative preference embedding [25].

In addition to original user-POI matrix, the user-based mobile contexts always indicate the preferences of users directly and could make reasonable explanations effectively. An excellent POI recommendation system, no matter what technology is used, should fully consider the contexts, as it greatly influences the decisions of users on visiting a new location [18]. For POI recommendation, geographic distance is the most intuitive and important context. The correlation between distance and number of visits is shown in Fig. 1, which is based on the datasets we used in Section 5. In Fig. 1, based on the check-in sequences of all users, we count the distance of two adjacent locations and the times of each distance. Then, we calculate the probability of happening of each distance, The left and right parts are based on the two datasets we used. The correlation we find proves the same conclusion with Ye [28]. He argues that influence of geographical distance among two POIs is able to decide user check-in behaviors and model it by power law distribution. Many geo-based POI recommendation models refer to his work so far. In contrast, Liu models the geographical influence from the perspective of location instead of the preferences of users and then adopts geographical neighborhoods of locations [21]. Li proposes a ranking based geographical factorization method that puts both checked and unchecked POIs into learning the ranking and incorporates the geographical context [16]. Liu proposes an adversarial learning model based on geographical information to dig deeper geographical representations [20]. While most existing works discover the spatial, temporal and social patterns of user behavior, the information of mobile context itself is often ignored [7]. Category that represents the functional attribute of a location indicates the explicit information of context. He uses a third-rank tensor to predict the next preferred category that user may visit and then fuse distance to filter out POI candidates [9]. Zhang applies the bias of a user on a POI category to weigh the popularity of a POI in the corresponding category and models the weighed popularity as a power-law distribution to leverage the categorical correlations between POIs [32].

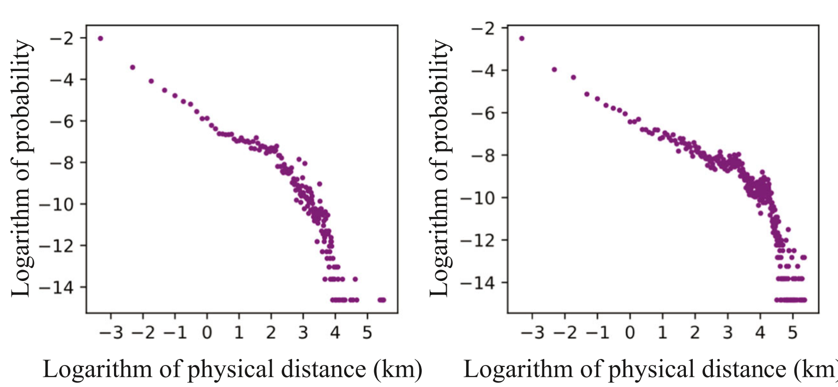


Fig. 1 Geographical distance probability on datasets

3 PR-RCUC: a POI recommendation using region-based collaborative filtering and user-based mobile context

POI recommendation aims to recommend new locations to users that they may be interested in by mining their historical check-in data. In this section, we will introduce our recommendation model PR-RCUC on the whole. Firstly, we cluster locations into different regions and add the region factor into traditional collaborative filtering, which is called PR-RC. The first part is to enhance the CF since users just check-in a few locations and that causes the data sparsity. Secondly, based on geographical distance and category transition matrix, we model the distance probability and category information respectively. The second part is to make full use of user-based mobile contexts because they will influence the decisions on visiting a new location. Finally, we compute the scores of predictions by combination of the two parts mention above to make lists of top-K recommendations for users. The framework is shown in Fig. 2.

3.1 Problem formulation

Suppose we have a set U of users $\{u_1, u_2, \cdots, u_n\}$, a set L of locations $\{l_1, l_2, \cdots, l_m\}$ and a set C of categories $\{c_1, c_2, \cdots, c_p\}$. Each location is denoted by a triple $l = (lon, lat, c)$, which includes the longitude and latitude of the location and the category it belongs to. The user-POI matrix is P: $U \times L$ and P_{ij} represents the number of times user u_i has visited the location l_j, which is defined as follows:

$$P_{ij} = \begin{cases} times, if\ l_j\ is\ visited\ by\ u_i \\ 0, if\ l_j\ is\ unvisited\ by\ u_i \end{cases} \tag{1}$$

Our goal is to predict the unknown $\widehat{P}_{ij}$ and produce a list Rec_{u_i} of recommendations for u_i, which is defined as follows:

$$\widehat{P}_{ij} = F\left(u_i, l_j|\theta\right) \tag{2}$$

$$Rec_{u_i} = \left\{l_j|sorted\ by\ \widehat{P}_{ij}, K\right\} \tag{3}$$

where F is our proposed recommendation model, θ is parameter set and K is the length of Rec_{u_i}. Our major task is to recommend satisfactory POIs to users and make them enjoy our mobile service.

3.2 Region-based collaborative filtering

It is intuitive and realistic to cluster locations in terms of distance. The more adjacent two locations are, the more likely they will gather to a region. Meanwhile, the number of regions is much smaller than that of locations and then region factor could alleviate the data sparsity to a certain. There are many clustering methods. Since we plan to cluster locations into regions according to the geographical distances between locations, we adopt classical K-means algorithm in our model. The goal of K-means is to minimize the sum of the distances, which exactly meets the definition of region. Suppose we have q regions $R = \left\{r_1, r_2, \cdots, r_q\right\}$

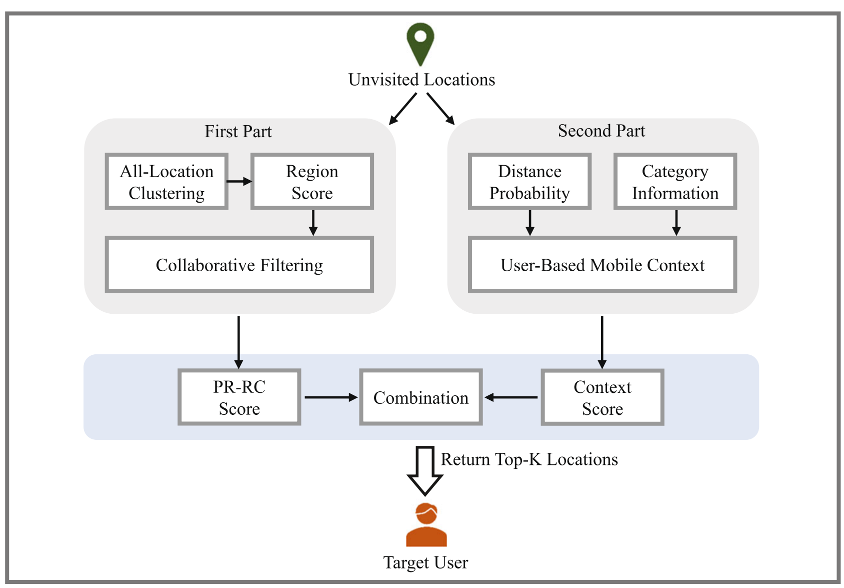


Fig. 2 The Framework of PR-RCUC

where $|R| \ll |L|$ and μ_k is the center location of region r_k. So the objective function is defined as follows:

$$\min \sum_{k=1}^{q} \sum_{l_j \in r_k} \left\| dis\left(l_j, \mu_k\right) \right\|^2 \tag{4}$$

where dis computes the distance between two locations according to their latitudes and longitudes. The region number q is the key parameter of K-means in terms of the influence of region. Hence, we need to conduct sufficient experiments to select the optimal q. According to K-means algorithm, the initial region centers are selected randomly. After continuous iteration, the region centers change dynamically to adapt to the objective function of K-means. After obtaining regions that is shown in Fig. 3, it is easy to know the overall check-ins of each region. Different colors in Fig. 3 represent different regions and it is obvious that all locations are clustered around their region centers. Based on region factor, the user-region matrix is $Z : U \times R$ and Z_{ik} represents the number of times user u_i has visited the region l_k, which is defined as follows:

$$Z_{ij} = \begin{cases} times, if\ r_k\ is\ visited\ by\ u_i \\ 0, if\ r_k\ is\ unvisited\ by\ u_i \end{cases} \tag{5}$$

In order to compare with the overall regional visits, we first have adopted a normalization on each user vector in Z in advance. The region similarity between the user u_i and other users is shown as follows:

$$sim\left(u_i, R\right) = \frac{\sum_{k=1}^{q} Z_{ik} \times cnt\left(r_k\right)}{\sqrt{\sum_{k=1}^{q} \left(Z_{ik}\right)^2} \times \sqrt{\sum_{k=1}^{q} cnt^2\left(r_k\right)}} \tag{6}$$

where $cnt(r_k)$ denotes the ratio of the total number of check-ins in the region r_k to the maximum of all the regions. $sim(u_i, R)$ indicates whether the user will follow others while choosing a region. While there are many methods for calculating similarity, cosine similarity compares two things in terms of their trends in vector and it will not be influenced by the real values in vector. If $sim(u_i, R)$ is large, it means the user is more willing to visit the regions that other users tend to visit. Otherwise, the user may have his own unique taste on visiting regions that is different from others. Next, the new prediction of $\widehat{Z}_{ik}$ is defined as follows, which considers both other users and target user himself:

$$\widehat{Z}_{ik} = sim\left(u_i, R\right) \times cnt\left(r_k\right) + \left(1 - sim(u_i, R)\right) \times Z_{ik} \tag{7}$$

Our goal is to enhance CF by adding region factor into that. The original CF that is based on locations is shown as follows:

$$cf\left(u_i, l_j\right) = \frac{\sum_{s=1}^{SU(u_i)} sim\left(u_i, u_s\right) \times P_{sj}}{\sum_{s=1}^{SU(u_i)} sim\left(u_i, u_s\right)} \tag{8}$$

where $SU(u_i)$ is the similar-user group of u_i. The traditional CF faces the data sparsity because most users just check-in a few locations and that may make recommendation results inaccurate. On the contrary, the number of regions is smaller than that of locations, which gives us chance to capture the preferences of users on regions. The regions-based collaborative filtering is defined as follows:

$$Score_{u_i, l_j}^{PR-RC} = \widehat{Z}_{ik} + \frac{cf\left(u_i, l_j\right)}{\max_{l_s \in L}\left(cf\left(u_i, l_s\right)\right)} \tag{9}$$

Different from geographical and category contexts, region factor is a further strategy for CF so we don't adopt weighted calculation for them. r_k of $\widehat{Z}_{ik}$ is corresponding region of l_j. Next, we will explore the influence of user-based mobile context.

3.3 Geographical distance context

The user-based mobile contexts will help us find the POIs that users may have interests in, such as geographical distance and location category. Generally speaking, geographical distance is the major context in POI recommendation since users tend to visit the locations that are closer to their current location. In order to model the distance probability, inspired by the work of Ye [28], we decide to adopt a power law distribution that is defined as follows:

$$Prob\left(l_i | l_j\right) = a \times dis\left(l_i, l_j\right)^b \tag{10}$$

where $Prob\left(l_i | l_j\right)$ denotes the probability of visiting location l_i while the current location is l_j. a and b are the parameters of the power law distribution. This is a non-linear mode and it is hard to solve directly. For obtaining the suitable a and b, we convert it to a linear model by using logarithmic representation, which is defined as follows:

$$\log Prob\left(l_i | l_j\right) = \log a + b \log dis\left(l_i, l_j\right) \tag{11}$$

$$y\left(x, W\right) = a' + bx \tag{12}$$

where $y\left(x, W\right)$ denotes $\log Prob\left(l_i \mid l_j\right)$ and W is the parameter set.a' is equal to $\log a$.x denotes the pair of two adjacent locations l_i and l_j.Now our distance probability

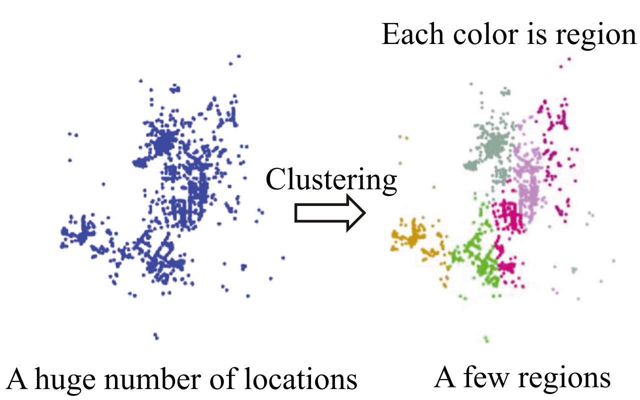


Fig. 3 Region

model is converted into a linear form. We adopt the least squares method in terms of its simplicity and the loss function that needs to be minimized is defined as follows:

$$\min_{W} \frac{1}{2} \sum_{x \in D} \left(y\left(x, W\right) - t\left(x\right)\right)^2 + \frac{\lambda}{2} \|W\|^2 \tag{13}$$

where D is our real-world dataset and $t(x)$ is the logarithmic value of the true distance probability derived from D. Moreover, the last term is a regularization that is controlled by its weight λ.

Although we cluster locations into regions, a user still visits a few locations in one region. Sometimes the mobile behaviors of users indicate regular patterns. For example, the regular activities of users always are around their workplace or home. So, it is necessary to compute predictions of all locations in a region. Now we present a new concept that is called activity area. The locations that user has visited in a region denote the activity areas of this user, which is shown in Fig. 4. Hence, in order to filter out candidate locations, we set 0 to those locations that is out of activity areas. We only calculate geographical score for locations belong to the activity areas of the user.

In a region, centered by location the user has visited, the circle with a radius of 0.5 kilometers composes an activity area. If two activity areas intersect, then merge them into one activity area. For target user, suppose we have a set A_{u_i} of his activity areas $\{a_1, a_2, \cdots, a_o\}$, where a_x represents one activity area that consists of many locations. Therefore, the geographical probability of the candidate location l_j is calculated as follows:

$$p\left(l_j|a_x\right) = \prod_{l_s \in a_x} p\left(l_j|l_s\right) \tag{14}$$

where l_j is an unvisited location in the activity area a_x of user and l_s is a location user has visited before in a_x. That means we use historical locations in activity area to get the probability of a new unvisited location. Next, the geographical score of unvisited l_j is defined as follows:

$$Score^{Geo}_{u_i,l_j} = \frac{count\left(a_x\right)}{\max\limits_{a \in A_{u_i}} count\left(a\right)} \times \frac{p\left(l_j|a_x\right)}{\max\limits_{l_k \in a_x} p\left(l_k|a_x\right)} \tag{15}$$

where $count\left(a_x\right)$ is the function that counts the total number of check-ins of the user in a_x. Even in the same region, user has different check-in behaviors in different activity areas. l_k denotes the unvisited location in a_x. For the sake of making all candidate locations comparable, we decide to normalize the geographical scores of them according to activity areas they belong to and also take the weight of each activity area into account.

3.4 Category context

Categories indicate the semantic information and reflect the functions of locations. Just category information cannot be called mobile context. However, in our model, we propose a pair-wise method which is based on the transition between categories. If user has visited both category c_i and c_j, then we can say there is a transition relationship between c_i and c_j. This kind of transfer has symmetry. The category transition matrix is $T : C \times C$ and T_{ij} represents the number of users who has visited both c_i and c_j, which is defined as follows:

$$T_{ij} = \begin{cases} user\ counts, if\ c_i\ and\ c_j\ are\ both\ visited \\ 0, otherwise \end{cases} \tag{16}$$

In a region, we use $C_{u_i,r}$ to denote the set of categories that user u_i has visited. Suppose candidate location l_j belongs to category c_k and its category score is defined as follows:

$$Score^{Cate}_{u_i,l_j} = \begin{cases} \frac{count(c_k)}{\sum_{c_v \in C_{u_i,r}} count(c_v)}, c_k \in C_{u_i,r} \\ f\left(c_k\right), c_k \notin C_{u_i,r} \end{cases} \tag{17}$$

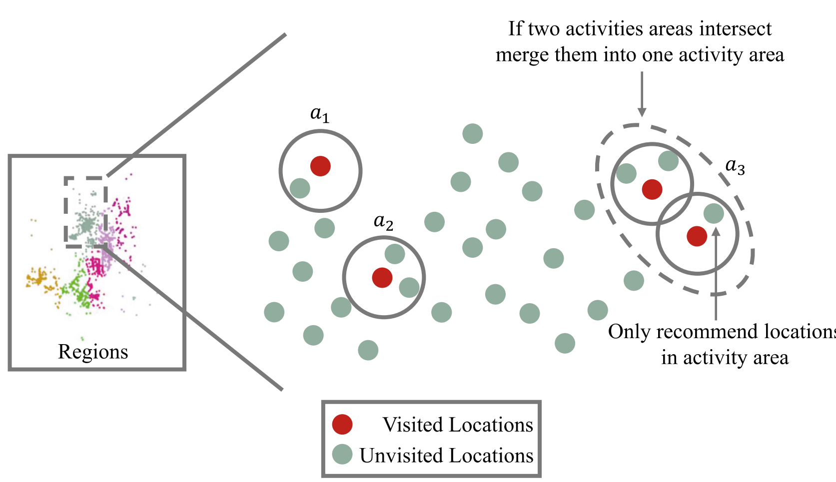


Fig. 4 Activity areas

$$f\left(c_k\right)=\frac{\frac{1}{\left|C_{u_i,r}\right|}\sum_{c_v\in C_{u_i,r}}T_{kv}}{\sum_{c_s\in C_{u_i,r}}f\left(c_s\right)} \tag{18}$$

where c_v denotes the category user has visited and c_s denotes the unvisited category. All computations are based the region that l_j belongs to. If user has visited the category before, then we calculate its probability according to the check-ins of target user in this region. Otherwise, we adopt function f to get its category score, which is based on the category transition matrix. After obtaining geographical distance score and category score, the final user-based mobile context score of location l_j is defined as follows:

$$Score_{u_i,l_j}^{Context}=\beta\times Score_{u_i,l_j}^{Geo}+(1-\beta)\times\frac{Score_{u_i,l_j}^{Cate}}{\max\limits_{l_x\in r}Score_{u_i,l_x}^{Cate}} \tag{19}$$

where l_x and l_j belong to the same region r and $\beta\in[0,1]$. We have adopted normalization when calculating geographical score, so now we only have to normalize category score.

3.5 Final recommendation

Firstly, we cluster locations into regions and add region factor into collaborative filtering in order to alleviate the negative effect cased by data sparsity. Secondly, we model the geographical distance probability and category information respectively since user-based mobile contexts could improve the accuracy of recommendation results. Now, we propose the final model called PR-RCUC by combing $Score_{u_i,l_j}^{PR-RC}$ with $Score_{u_i,l_j}^{Context}$, which is defined as follows:

$$\widehat{P}_{ij}=\alpha\times Score_{u_i,l_j}^{RCF}+(1-\alpha)\times Score_{u_i,l_j}^{Context} \tag{20}$$

where $\alpha\in[0,1]$. α is the weight that controls the region-based CF. If α is large, the region-based CF will play a greater role than context-based model. Finally, we sort the scores of all candidate locations and recommend top-K locations to users. $\widehat{P}_{ij}$ is the result of normalization on two kind of prediction scores and we just need to know the relative order of unvisited locations for making the recommendation list Rec_{u_i}.

4 Experiments

4.1 Datasets

We use two real-word datasets on two cities from Foursquare. One is Los Angeles and the other is London. We preprocess check-ins in Foursquare by removing users who have visited fewer than 5 POIs, and POIs which are visited by fewer than 5 users. After preprocessing, there are 48460 check-ins generated by 4746 users over 7135 POIs in the Los Angeles dataset and the average check-ins of each user is 10. There are 43912 check-ins generated by 3470 users over 7941 POIs in the London dataset and the average check-ins of per user is 12. Each POI in both datasets is associated with its longitude and latitude. Referring to most of existing papers of POI recommendation, we randomly select 70% of the locations of each user as training data and the remaining 30% as test data.

4.2 Evaluation metrics

We adopt three widely-used metrics for evaluation, namely, precision, recall and F1-score. Precision and recall are classical metrics in machine learning. F1-score combines precision with recall. All evaluation metrics are defined as follows:

$$Precision@K=\frac{1}{|U|}\sum_{u\in U}\frac{\left|Rec_u\cap Test_u\right|}{\left|Rec_u\right|} \tag{21}$$

$$Recall@K=\frac{1}{|U|}\sum_{u\in U}\frac{\left|Rec_u\cap Test_u\right|}{\left|Test_u\right|} \tag{22}$$

$$F1-score@K=2\frac{Precision@K\times Recall@K}{Precision@K+Recall@K} \tag{23}$$

where F1-score is based on the combination of precision and recall. K is the number of recommended POIs, which is always set to 10, 15 and 20. Rec_u is the recommendation list for user u and $Test_u$ is the test data of user u.

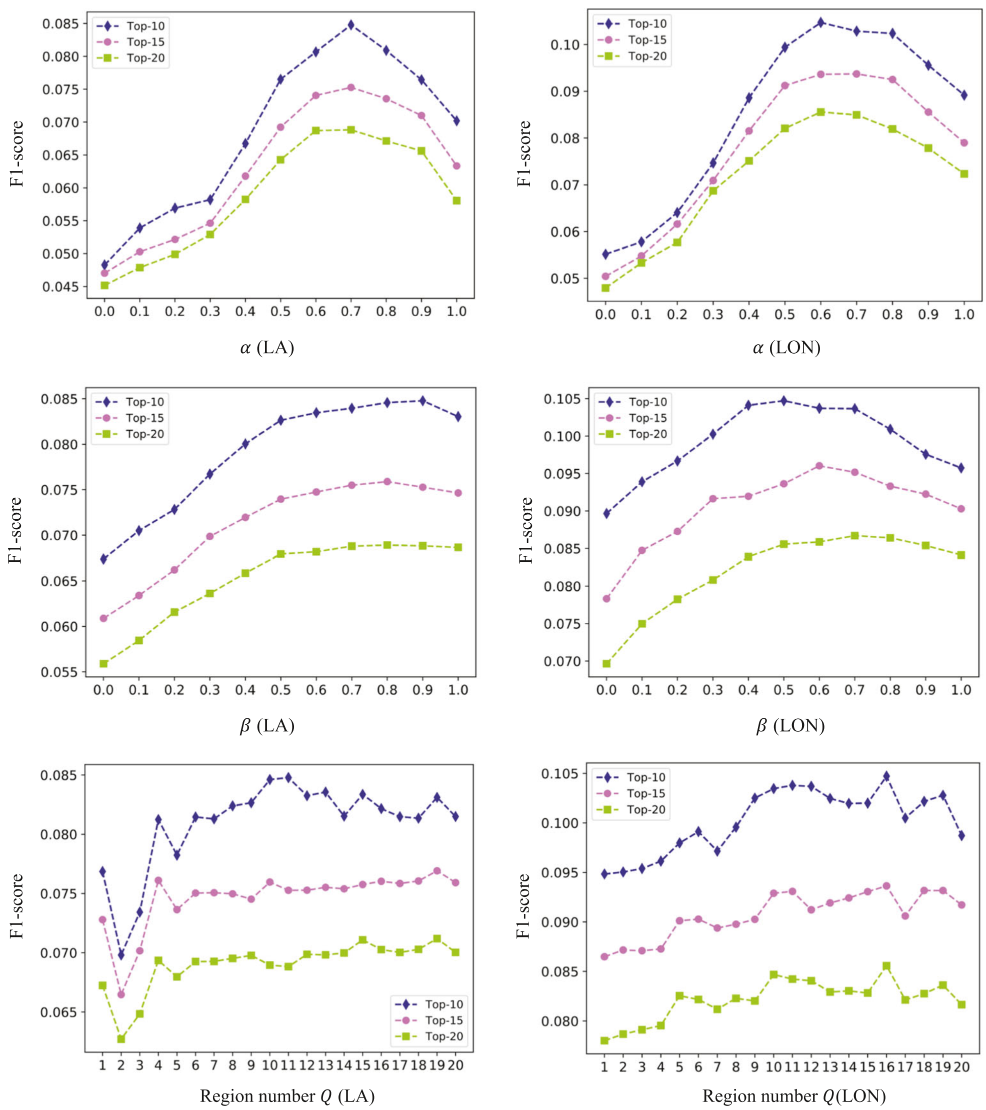
4.3 Experimental procedure

Firstly, we preprocess the datasets and make them become structured data. Secondly, we have three primary parameters in PR-RCUC. Q determines the region number which is important for region-based collaborative filtering. α weights the region-based CF and the influence of user-based mobile context. β weights the geographical distance and location category. Hence, we use F1-score to find the best values of them since F1-score is the combination of precision and recall and it will simplify the procedure. Thirdly, in order to evaluate the performance of our model, we compare it with other popular algorithms. The parameter setting of each comparative experiment is based on the optimal setting of their original paper.

4.4 Parameter study

We do experiments on different combinations of the three parameters of our model and the experimental results are shown in Fig. 5. As we mentioned before, Q determines

Fig. 5 Experimental result on parameters



how to cluster locations into regions. α weights the region-based CF and the influence of context and β weights the geographical distance and location category.

As for α, we observe performance variance when we change the hyper-parameter α. It is clear that all curves first go up and then go down in both datasets, which indicates the performance of PR-RCUC could be improved by the combination of region-based CF and user-based mobile context. For the LA dataset, the optimal α is 0.7 while that for the LON dataset is 0.6. After reaching their peaks, they start to decline and that means excessively ignoring the contexts will cause the performance to decrease. $\alpha = 0.0$ means our model depends only on the influence of user-based mobile context while $\alpha = 1.0$ means there is only region-based CF. The latter is better than former on both datasets, which proves clustering locations into CF will enhance the recommendation results of CF effectively. Mobile contexts could help us complete recommendation service but they should be used as a way of strategy to support the core algorithms.

As for β, we can see that with the increase with parameter β, both datasets are trending upward. However, the curves of LA dataset have clear upward trends before $\beta = 0.5$ but then become flat gradually. It reveals the geographical distance context has limited improvement and there will not be much changes after reaching peaks. On the contrary, while Top-K =10 and 15, the curves of LON dataset go down significantly after reaching their peaks, which indicates adding too much influence of geographical distance will weaken the performance. Since the two datasets have different distributions of users and locations, it is normal that they show different trends. For both datasets, region-based CF with only geographical distance context where $\beta = 1.0$ outperforms that with only category context where $\beta = 0.0$. This fully demonstrates that geographical distance is the most important user-based mobile context for POI recommendation. Although location category context dose not play the same role as geographical distance context, it is still worth studying and could be fused with geographical distance context.

As for region number Q, it is obvious that there are many significant fluctuations on both datasets. One possible reason is that, for each Q, the K-means method clusters locations into regions dynamically and the results of clustering cannot be predicted. Meanwhile, some differences exist between two datasets. A huge decline happens for LA dataset when $Q = 2$, which indicates that few regions may result in insufficient clusters especially at the beginning. However, with the increase of Q, the advantage of region comes into play gradually. Fortunately, $Q = 2$ does not destroy the performance, which may be due to characteristics of the dataset since distributions of locations will affect the clustering for regions. In general, a larger region number could improve our PR-RCUC even though the fluctuations exist. Besides, a larger region number help us understand the behaviors of users on regions.

4.5 Performance comparison

After obtaining the optimal parameters of PR-RCUC, we compare it with following popular recommendation models:

POP: The basic model that recommends the most popular POIs to users.

CF: The classical user-based collaborative filtering that is widely-used in industry.

NMF: Non-negative matrix factorization which aims to fill in the unknown items in user-POI matrix.

BPR [23]: The Bayesian personalized ranking model via optimizing the relative order of users and POIs.

DMF [26]: The novel deep matrix factorization that is designed for recommendation system, which learns a common low-dimensional space for users and recommendations.

NCF [11]: Neural collaborative filtering that is a general framework based on neural networks and could model the latent features of users and recommendation items.

PR-RC: The single recommendation model that only considers the region-based collaborative filtering, which we present in Section 3.2.

The model we finally propose in this paper is called PR-RCUC which combines the region-based CF with user-based mobile context. Although POP and CF are relatively old, they are still adopted by recommendation systems today. NMF and BPR could be regarded as baselines of matrix factorization and pair-wise recommendation. Nowadays, DMF and NCF are popular algorithms based on deep learning. Those recommendation models we choose indicate the different types when solving recommendation task. Besides, our model is based on multiple contexts and we are able to compare it with those types. For each model we choose for comparison, we explore the optimal hyper-parameters to suit our regular datasets. For example, since our dataset is not large, it is necessary to reduce the layers and neural-units for NCF and DMF.

We conduct comparison experiments on two datasets respectively. For the LA dataset, we set $Q = 11$, $\alpha = 0.7$ and $\beta = 0.9$. For the LON dataset, we set $Q = 16$, $\alpha = 0.6$ and $\beta = 0.5$. The performance comparison is shown in Fig. 6 and then we summarize the following observations.

First of all, compared with other recommendation algorithms, our proposed PR-RCUC is better in terms of both precision and recall. Since we combine the region-based CF and user-based mobile context, the performance of recommendation system could be improved and it is not hard to explain the results of recommendation. The PR-RC which only considers the region-based CF dose not outperform the BPR, which proves the Bayesian personalized ranking is good at solving the relative-order problem of users and POIs. BPR mines the correlations between users and POIs effectively by matching them instead of computing the values of unknown items. Although the DMF is state-of-art deep model that is designed for recommendation system, it dose not beat the traditional CF, just better than NMF and POP. One possible reason for that is DMF depends on the structure of neural network to a great extent and a small change will lead to a different result. Besides, DMF as a deep learning model, is very sensitive to the datasets since they could train the layer weights of neural network. DMF dose not adopt any contexts of recommendation system. In POI recommendation, if we put the unique characteristics into DMF, such as geographical distance and location category, it may achieve better performance.

Obviously, the classical CF performs well on both datasets, which tell us why it could become the widely-used basic model in industry. The performances of NCF are much different on two datasets. For the London dataset, NCF outperforms CF slightly in terms of precision. For the LA dataset, its performance is not good as CF on both precision and recall. Since NCF adopt neural networks to construct the general framework for CF, its training process depends on the check-ins we choose for computing CF. This inspires us to cluster locations into regions for further enhance the CF. The PR-RC adds region factor into CF to alleviate the negative effect caused by data sparsity of CF. Then we combine PR-RC with user-based mobile context and propose PR-RCUC which achieves excellent improvement as we expect. Since the PR-RC only considers the region-based CF, it dose not outperform the PR-RCUC. NMF and POP are baseline algorithms and it is normal that they

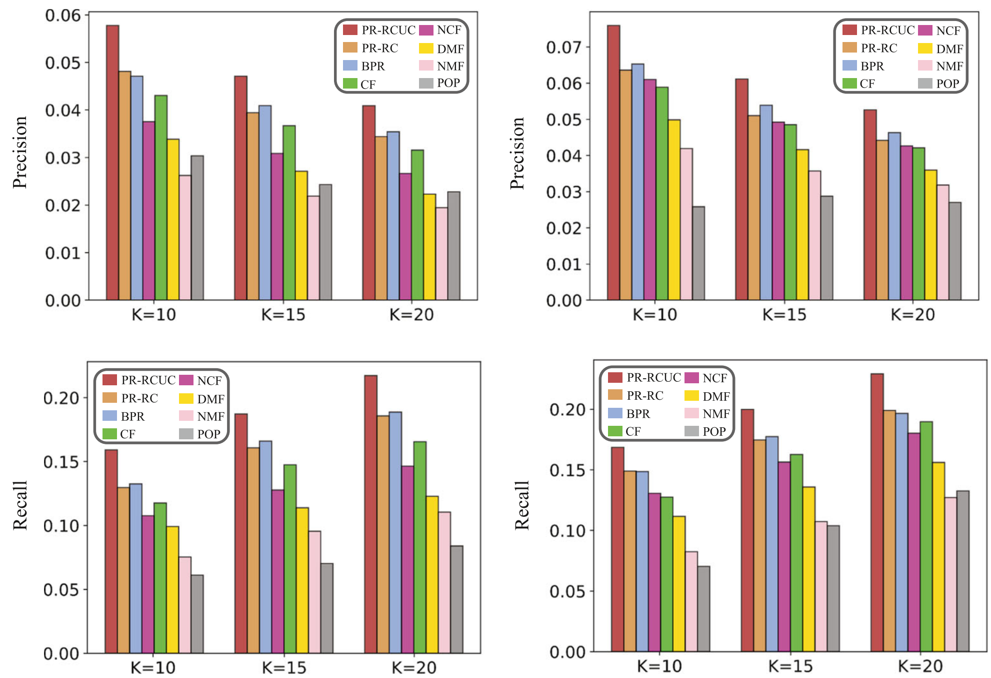


Fig. 6 Performance comparison in terms of precision and recall (Left: LA, Right: LON)

both have the worst performances. However, such baseline models could be completed and understood easily. In a word, our proposed PR-RCUC outperforms some popular recommendation models because of its combination of region-based CF and user-based mobile context. Moreover, the results of recommendations obtained by our model are reasonable according to the explanation of region and context.

5 Conclusion

Recent years have witnessed the rapid prevalence of location-based social networks, which facilitates the outdoor activities of people. Point of intertest (POI) recommendation that aims to recommend satisfactory locations to users has faces two major challenges. One is the traditional recommendation models is affected by the data sparsity since users always visit a few locations. The other is that it is difficult for recommendation system to make reasonable explanations on the results of recommendations. Hence, we propose a model called PR-RCUC that combines the region-based collaborative filtering and user-based mobile context. First, we cluster locations into different regions and add region factor to collaborative filtering to alleviate the negative effect caused by data sparsity. Second, we model the user-based mobile context such as geographical distance and location category. Third, we fuse the two parts mentioned above and propose our final model. Experimental results on two datasets collected from Foursquare demonstrate our proposed model outperforms some popular recommendation algorithms. Moreover, the recommendation of our model could be reasonable according to the explanation of region and context. In the future, on one hand, we will try different geo-clustering methods to get the regions. On the other hand, we only adopt two kinds of user-based mobile contexts in this paper and we will take into account more contexts such as weather and season.

Acknowledgements This research is sponsored by Natural Science Foundation of Chongqing, China (No.cstc2020jcyj-msxmX0900) and the Fundamental Research Funds for the Central Universities (Project No.2020CDJ-LHZZ-040).

References

1. Cheng C, Yang H, King I, Lyu MR (2012) Fused matrix factorization with geographical and social influence in location-based social networks. In: Aaai, vol 12, pp 17–23
2. Fletcher K (2019) Regularizing matrix factorization with implicit user preference embeddings for web API recommendation. In: 2019 IEEE International conference on services computing (SCC), IEEE, pp 1–8
3. Gao H, Duan Y, Shao L, Sun X (2019) Transformation-based processing of typed resources for multimedia sources in the IoT environment. Wirel Netw, pp 1–17
4. Gao H, Huang W, Duan Y (2020) The cloud-edge based dynamic reconfiguration to service workflow for mobile ecommerce environments: A QoS prediction perspective. ACM Transactions on Internet Technology
5. Gao H, Kuang L, Yin Y, Guo B, Dou K (2020) Mining consuming behaviors with temporal evolution for personalized recommendation in mobile marketing Apps. ACM/Springer Mobile Networks and Applications (MONET) 25(4):1233–1248

6. Gao H, Liu C, Li Y, Yang X (2020) V2VR: Reliable hybrid-network-oriented V2V data transmission and routing considering RSUs and connectivity probability. IEEE Transactions on Intelligent Transportation Systems
7. Gao H, Tang J, Hu X, Liu H et al (2015) Content-aware point of interest recommendation on location-based social networks. In: Aaai, vol 15, Citeseer, pp 1721–1727
8. Gao H, Xu Y, Yin Y, Zhang W, Li R, Wang X (2019) Context-aware QoS prediction with neural collaborative filtering for internet-of-things services. IEEE Internet Things J 7(5):4532–4542
9. He J, Li X, Liao L (2017) Category-aware next point-of-interest recommendation via listwise bayesian personalized ranking. In: IJCAI, vol 17, pp 1837–1843
10. He J, Li X, Liao L, Song D, Cheung WK (2016) Inferring a personalized next point-of-interest recommendation model with latent behavior patterns. In: Proceedings of the Thirtieth AAAI conference on artificial intelligence, pp 137–143
11. He X, Liao L, Zhang H, Nie L, Hu X, Chua TS (2017) Neural collaborative filtering. In: Proceedings of the 26th international conference on world wide web, pp 173–182
12. Hu S, Tu Z, Wang Z, Xu X (2019) A POI-sensitive knowledge graph based service recommendation method. In: 2019 IEEE International conference on services computing (SCC), IEEE, pp 197–201
13. Li H, Ge Y, Hong R, Zhu H (2016) Point-of-interest recommendations: Learning potential check-ins from friends. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, pp 975–984
14. Li H, Ge Y, Lian D, Liu H (2017) Learning user's intrinsic and extrinsic interests for point-of-interest recommendation: a unified approach. In: IJCAI, pp 2117–2123
15. Li H, Hong R, Zhu S, Ge Y (2015) Point-of-interest recommender systems: a separate-space perspective. In: 2015 IEEE International conference on data mining, IEEE, pp 231–240
16. Li X, Cong G, Li XL, Pham TAN, Krishnaswamy S (2015) Rank-geoFM: A ranking based geographical factorization method for point of interest recommendation. In: Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval, pp 433–442
17. Lian D, Wu Y, Ge Y, Xie X, Chen E (2020) Geography-aware sequential location recommendation. In: Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining, pp 2009–2019
18. Liu Q, Wu S, Wang L, Tan T (2016) Predicting the next location: a recurrent model with spatial and temporal contexts. In: Thirtieth AAAI conference on artificial intelligence
19. Liu S, Wang L (2018) A self-adaptive point-of-interest recommendation algorithm based on a multi-order markov model. Futur Gener Comput Syst 89:506–514
20. Liu W, Wang ZJ, Yao B, Yin J (2019) Geo-ALM: POI recommendation by fusing geographical information and adversarial learning mechanism. In: IJCAI, pp 1807–1813
21. Liu Y, Wei W, Sun A, Miao C (2014) Exploiting geographical neighborhood characteristics for location recommendation. In: Proceedings of the 23rd ACM international conference on information and knowledge management, pp 739–748
22. Manotumruksa J, Macdonald C, Ounis I (2017) A personalised ranking framework with multiple sampling criteria for venue recommendation. In: Proceedings of the 2017 ACM on conference on information and knowledge management, pp 1469–1478
23. Rendle S, Freudenthaler C, Gantner Z, Schmidt-Thieme L (2012) BPR: Bayesian personalized ranking from implicit feedback. arXiv preprint arXiv:1205.2618
24. Sarwar B, Karypis G, Konstan J, Riedl J (2001) Item-based collaborative filtering recommendation algorithms. In: Proceedings of the 10th international conference on World Wide Web, pp 285–295
25. Tran T, Lee K, Liao Y, Lee D (2018) Regularizing matrix factorization with user and item embeddings for recommendation. In: Proceedings of the 27th ACM international conference on information and knowledge management, pp 687–696
26. Xue HJ, Dai X, Zhang J, Huang S, Chen J (2017) Deep matrix factorization models for recommender systems. In: IJCAI, vol. 17, Melbourne, Australia, pp 3203–3209
27. Yang C, Bai L, Zhang C, Yuan Q, Han J (2017) Bridging collaborative filtering and semi-supervised learning: a neural approach for poi recommendation. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp 1245–1254
28. Ye M, Yin P, Lee WC, Lee DL (2011) Exploiting geographical influence for collaborative point-of-interest recommendation. In: Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval, pp 325–334
29. Zeng J, He X, Tang H, Wen J (2020) Predicting the next location: a self-attention and recurrent neural network model with temporal context. Transactions on Emerging Telecommunications Technologies, pp e3898
30. Zeng J, Li F, He X, Wen J (2019) Fused collaborative filtering with user preference, geographical and social influence for point of interest recommendation. Int J Web Serv Res (IJWSR) 16(4):40–52
31. Zeng J, Tang H, Li Y, He X (2019) A deep learning model based on sparse matrix for point-of-interest recommendation. In: SEKE, pp 379–492
32. Zhang JD, Chow CY (2015) GeoSoCa: Exploiting geographical, social and categorical correlations for point-of-interest recommendations. In: Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval, pp 443–452
33. Zhao K, Zhang Y, Yin H, Wang J, Zheng K, Zhou X, Xing C (2020) Discovering subsequence patterns for next poi recommendation. In: Proceedings of the Twenty-Ninth international joint conference on artificial intelligence, pp 3216–3222
34. Zhao S, Zhao T, King I, Lyu MR (2016) GT-SEER: Geo-temporal sequential embedding rank for point-of-interest recommendation. arXiv:1606.05859