Joint Geo-Spatial Preference and Pairwise Ranking for Point-of-Interest Recommendation

Fajie Yuan
University of Glasgow, UK
f.yuan.1@research.gla.ac.uk

Guibing Guo
Northeastern University, China
guogb@swc.neu.edu.cn

J. Jose, L. Chen, H. Yu, R. Alkhawaldeh
University of Glasgow, UK
{Joemon.Jose, Long.Chen, hait.yu, r.alkhawaldeh}@research.gla.ac.uk

***Abstract*—Recommending users with preferred point-of-interests (POIs) has become an important task for location-based social networks, which facilitates users' urban exploration by helping them filter out unattractive locations. Although the influence of geographical neighborhood has been studied in the *rating prediction* task (i.e. regression), few work have exploited it to develop a ranking-oriented objective function to improve *top-N item recommendations*. To solve this task, we conduct a manual inspection on real-world datasets, and find that each individual's traits are likely to cluster around multiple centers. Hence, we propose a co-pairwise ranking model based on the assumption that users prefer to assign higher ranks to the POIs near previously rated ones. The proposed method can learn preference ordering from non-observed rating pairs, and thus can alleviate the sparsity problem of matrix factorization. Evaluation on two publicly available datasets shows that our method performs significantly better than state-of-the-art techniques for the *top-N item recommendation* task.**

***Keywords*—top-N item recommendation; point-of-interest; spatial preference; pairwise ranking.**

I. Introduction

Location-based social networks have emerged as an application to assist users in their decision-making among a wide variety of point-of-interests (POIs), e.g. bars, stores, and cinemas. Typical location-based websites, such as yelp.com and foursquare.com, allow users to check-in POIs with mobile devices like smart phones and share tips with online friends [1]. Yelp, for instance, reaches a monthly average of 83 million unique access via mobile devices, with nearly hundred million reviews by the end of 2015[1]. The huge volume of location data contains valuable information about business popularity and customer preferences. However, how to effectively make a satisfactory decision among a large number of POIs has become a tough challenge for individuals. POI recommendation aims to solve such a problem by learning preference from users' previous visits.

In the field of personalized POI recommendation, the key tasks are to estimate users' preferences to unknown POIs and return the top-N POIs with highest rankings for them. Thus, most efforts focus on fitting a preference scoring function based on users' visiting profiles. Specifically, various types of contextual information, e.g. geographical coordinates [2], time stamps [1], social friends [3] have been studied within a single collaborative filtering (CF) model (e.g. matrix factorization [4]) or a unified framework [2][5]. However, all these methods are essentially based on the pointwise theory that aims to regress real-valued scores on item instances.

[1] http://www.yelp.co.uk/press

Unlike previous work, Rendle et al. [6] argue that the task of item recommendation is actually a classification (qualitative) problem rather than a regression (quantitative) one. Hence, they devise a Bayesian personalized ranking (BPR) model based on pairwise preference comparison over observed and non-observed feedback such that the Area Under the ROC Curves (AUC) can be maximized. Prior research has shown that the BPR-based approaches empirically outperform pointwise methods for implicit feedback data [6][7]. Nevertheless, it is reasonable to argue that the BPR-based ranking models do not explicitly exploit geographical influence. Hence, there seems a large marginal space left to improve the performance by extending it for POI recommendation.

We may infer that jointing geo-spatial preference and BPR optimization criterion creates new opportunities for POI recommendation. To motivate this work, we first conduct a manual inspection on two real-world datasets, and observe that a user's rating distribution represents a spatial clustering phenomenon. Thus, we presume that *unrated POIs surrounding a POI that users prefer are more likely to be assigned higher ranks over the distant unrated ones*. Based on this assumption, we propose a co-pairwise ranking model called geographical Bayesian personalized ranking (GeoBPR). Specifically, a user's geo-spatial preference is exploited as intermediate feedback, which is treated as weak preference relative to positive feedback while as strong preference in comparison to other non-positive feedback. In other words, we reformulate the *item recommendation* problem into a two-level joint pairwise ranking scheme. To the best of our knowledge, the reported work is the first to improve the BPR pairwise assumption by injecting users' geographical preference. Finally, we conduct extensive experiments to evaluate the effectiveness of GeoBPR, and the results indicate that our proposed model can significantly outperform an array of counterparts in terms of four popular ranking metrics.

II. Related Work

In this section, we briefly review the recent advances in POI recommendation, particularly those employing geographical influence for POI recommendation. As the major challenge of *top-N item recommendation* falls within the realm of One Class Collaborative Filtering (OCCF), we also review the related techniques.

A. POI recommendation

Recently, a number of valuable work have been presented in the realm of POI recommendation. Based on the type of additional information involved, POI recommendation algorithms have been classified into four categories [8], which are (1) pure check-in/rating based POI recommendation approaches [9], (2) social influence enhanced POI recommendation [2][5], (3) temporal influence enhanced POI recommendation [1][5], and (4) geographical influence enhanced POI recommendation [2][4]. In particular, in terms of geographical influence enhanced POI recommendation, usual approaches are to assume that users tend to visit nearby POIs and the probability of visiting a new place decreases as the distance increases. For example, Ye et al. [10] and Yuan et al. [11] modelled the check-in probability to the distance of the whole visiting history by power-law distribution; Chen et al. [2] pointed out that they ignored the geographical cluster phenomenon of users' check-ins, and computing all pairwise distance of the whole visiting history is time-consuming and thus cannot be adapted to large-scale datasets. In contrast, they suggested to model the probability of a user's check-ins as a multi-center Gaussian Model (MGM). Moreover, Lian et al. [4] incorporated the spatial clustering phenomenon into matrix factorization to improve recommendation performance. All discussed methods are essentially based on the pointwise theory that aims to regress a real-valued score, whereas few work attempt to build a ranking-based estimator for personalized recommendation, which is the main objective in this paper.

B. One-Class Collaborative Filtering

In the context of POI recommendation, it is well known that only positive feedback (e.g. check-ins) can be observed, whereas the non-observed feedback is mixed with both negative and unlabeled positive samples, referred to as One-Class Collaborative Filtering (OCCF) [12]. To solve the task, Rendle et al. [6] devised a Bayesian personalized ranking (BPR) model based on pairwise preference comparison over observed and non-observed rating pairs such that the Area Under the ROC Curves (AUC) can be maximized. Empirically, the pairwise ranking method achieves much better performance than traditional pointwise methods [6][7]. Following this, various ideas have been inspired by fusing other contextual information. For instance, the work in [13] extended BPR-based matrix factorization with tensor factorization (i.e. RTF). They further suggested to apply adaptive and context-dependent oversampling to replace the uniform sampling of BPR [14]. In [15][16], BPR criterion was extended by modeling social relations and social preference information. Similarly, Pan et al. [7] proposed an improved algorithm called group Bayesian personalized ranking by leveraging rich interactions among users to relax the pairwise assumption. More recently, Li et al. [17] designed a pairwise ranking model (Rank-GeoFM) based on the Ordered Weighted Pairwise Classification (OWPC) criterion that can incorporate different contextual information. However, the time complexity of Rank-GeoFM is largely increased because before each stochastic gradient descent (SGD) update, a number of samplings need to be drawn and their score is computed each time before an SGD update is performed. Moreover, within the proposed rejection sampler, the negative items are sampled uniformly, which might need a large number of draws before finding a negative item that has higher ranks than a positive one. In contrast, our proposed GeoBPR does not increase the time complexity in both learning and predicting processes.

TABLE I: Basic statistics of Datasets

DataSets	#Users	#POIs	#Ratings	Density
Phoenix	4510	16402	226351	0.31%
Las Vegas	4470	11376	207649	0.41%
DataSets	Avg.U	N.50	N.100	N.200
Phoenix	50.19	2.51	5.26	12.6
Las Vegas	46.45	5.54	9.44	20.8
DataSets	N.400	N.600	N.1000	N.2000
Phoenix	29.3	46.1	80.26	189.8
Las Vegas	47.6	77.3	146.1	392.3

The "Density" column is the density of each dataset (i.e. Density=#Ratings/(#Users × #Items)). The "Avg.U" column denotes the average number of visited POIs for each user. The "N.k" column refers to the average number of geographical neighbors for a POI at radius k.

It is worth to mention that after finishing this work, we notice that two recent work (i.e. [18][19]) have exploited BPR learning techniques for *next POI recommendation* task. However, we argue that our work is different in two aspects: (1) We model the geographical proximity influence from the fundamental BPR assumption, and propose a new and improved one. Accordingly, a two-level pairwise ranking model has been presented to learn the new assumption. By contrast, both work in [18][19] model the geographical influence by extending the preference scoring function from matrix factorization to tensor-based factorization [20] without modifying the basic pairwise preference assumption. (2) Our work falls in the area of standard *top-N item recommendation* task, while the above two work are referred to as *next POI recommendation* task by solving different research problems.

III. Geo-spatial Preference Analysis

A. Data Description

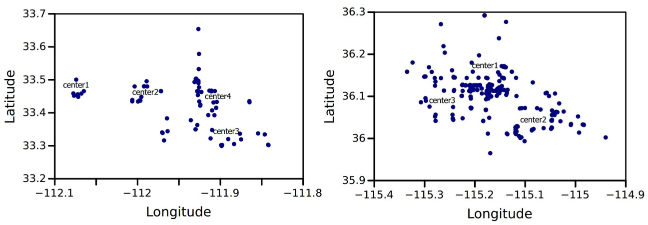
A recently released dataset Yelp[2] is used for data analysis. We extract data from two American cities (Phoenix and Las Vegas) that have the largest number of POIs and follow the common practice to remove users with less than 20 ratings and POIs with less than 5 ratings to reduce noise data[3], similarly as preprocessed in [2][21–23]. The basic statistics are shown in Table I.

B. Motivation

Motivation 1: Our first motivation is derived from the Tobler's First Law of Geography, which is "Everything is related to everything else, but near things are more related than distant things" [24]. This implies: (1) a user tends to visit nearby places [10]; (2) nearby places potentially have some relevance [21].

[2] www.yelp.co.uk/dataset_challenge

[3] The cold-start problem is beyond the concern of this work.



(a) Phoenix **(b)** Las Vegas

Fig. 1: The overview of a user's multi-center mobility behaviours on Phoenix and Las Vegas.

Motivation 2: Chen et al. [2] observed that users' check-in traces usually follows a multi-center distribution. Accordingly, they modelled the probability of a user's check-ins on a location as Multi-center Gaussian distribution and then fused the users' geographical preference and latent factor together in a unified framework. Ye et al. [10] argued that the probability of POI pairs visited by the same user approximately obeys power-law distribution with distance.

As a result, two main implications can be derived: (1) users usually visit POIs close to their activity centers, such as their homes and offices; (2) users may be interested in exploring POIs near a location they visited before, which have to be clustered together.

C. Proximity Analysis

We proceed to study if the above intuitive phenomena can be observed on our datasets. Figure 1 depicts the geographical distributions of POIs rated by two random users. It can be seen that the users' POIs indeed cluster around several spatial areas. That is, the above two implications are likely to hold on the Yelp datasets. Furthermore, we design the following statistical experiments to verify the intuitions.

Exp1: We randomly pick two POIs (l_a, l_b) from a city (e.g. Phoenix) and calculate the distance $d_{(a,b)}$ between l_a and l_b. We repeat the experiment 10000 times in order to yield the probability $\mathcal{P}$ that the distance $d_{(a,b)}$ is less than a threshold μ (e.g. $\mu = 200m$, where m is in meter).

Exp2: We randomly pick a user u from the same city in Exp1 and select two POIs that u has rated before, e.g. $(l_{a'}, l_{b'})$, then calculate the distance $d_{(a',b')}$ between $l_{a'}$ and $l_{b'}$. We repeat the experiment 10000 times to calculate the probability $\mathcal{P}^{'}$ that $d_{(a',b')}$ is less than the same threshold μ.

Table II shows the ratios of $\mathcal{P}^{'}$ to $\mathcal{P}$, which are indicators to demonstrate the proximity influence of individuals' rating behaviors. As shown, the ratios are much greater than 1 with all thresholds ([50, 2000]), which means $\mathcal{P}^{'}$ is higher than $\mathcal{P}$, in particular, $\mathcal{P}^{'}$ is about 20 times larger than $\mathcal{P}$ when μ is less than $200m$. This implies that users' visiting behaviors are highly affected by spatial distance and that the users' rated POIs are not geographically independent of each other. Second, all the ratios decrease with the increase of μ. This is consistent with intuition since the ratio should be close to 1 if μ is large enough. Moreover, this observation keeps consistent with our experimental results in Section VI-B.

TABLE II: Ratios of $\mathcal{P}^{'}/\mathcal{P}$

μ	50	100	200	400	600	1000	2000
Phoenix	44.7	28.8	23.3	18.9	13.6	10.2	8.0
Las Vegas	57.5	53.2	19.1	8.0	4.7	4.1	3.2

Unlike previous work, we do not model the distribution of multiple spatial cluster phenomenon directly since it is not proper to assume all users mobility patterns correspond to a prior distribution, e.g. Gaussian [2] or power-law [11] distribution. Nevertheless, the statistical analysis implies that an unrated POI surrounded a POI that one prefers is likely to be more appealing (to her) compared with other faraway and unrated places. The physical cost is the main difference that distinguishes POI recommendation from other product recommendations. Thus it becomes feasible to statistically model this intuition by a two-level pairwise preference comparison,

(Preference Rank of) a POI one rated $>$
nearby POIs she unrated $>$
unrated POIs far away from all rated POIs.

IV. Preliminary

First, we introduce several concepts used in this paper and define the research problem of geo-spatial preference enhanced POI recommendation. Then, we shortly recapitulate the basic idea of Bayesian Personalized Ranking (BPR). Table III lists the notations used in this work.

A. Problem Statement

In the context of typical POI recommendation, let $\mathcal{U} = \{u\}_{u=1}^{M}$ denote the set of all users, and $\mathcal{L} = \{i\}_{i=1}^{N}$ denote the set of all POIs, where M and N represent the number of users and POIs respectively, i.e. $M = |\mathcal{U}|$ & $N = |\mathcal{L}|$. Users' check-in/rating information is commonly expressed via user-POI check-in/rating action matrix $\mathcal{C}$, where each entry c_{ui} is the frequency or a binary value[4] made by u at i. Generally, the matrix $\mathcal{C}$ is extremely sparse (see Table I) since most users only rate a small portion of POIs. In contrast to other recommendation tasks, geographical information is available for each POI, which is usually geocoded by a pair of latitude and longitude.

In our scenario, we define the set of user-POI $(u,\ i)$ pairs as positive feedback if the rating behavior of u to i is observed, denoted as $\mathcal{L}_u^+ = \{(u,i)\}$. Unlike previous work (e.g. [6]) that defines the set of non-observed pairs $\mathcal{L}\backslash\mathcal{L}_u^+$ as negative feedback, we introduce a new geographical feedback by exploiting POI neighborhood information. In particular, assume we observe a POI geographical network $\mathcal{G} = (\mathcal{L}, \mathcal{L})$, where $(i, g) \in \mathcal{G}$ indicates that i and g are geographical neighbors. For each rated POI i, there is a neighbor $g \in \mathcal{L}$, which has not been rated by u. The set of (u, g) pairs is defined as geographical feedback, denoted as $\mathcal{L}_{ui}^{\mathcal{G}} = \{(u,g)\}$. Besides, there is a POI j neither rated by u nor a geographical neighbor of all rated POIs $i \in \mathcal{L}_u^+$. We define the set of (u, j) pairs as negative feedback, denoted as $\mathcal{L}_u^- = \{(u,j)\}$. For example, the circular area in

[4]For *item recommendation* task, it is common practice to handle explicit rating values as implicit binary values.

TABLE III: List of notations

Symbols	Meanings
$\mathcal{U}$	set of users $\{u_1, u_2, ..., u_{\|\mathcal{U}\|}\}$
$\mathcal{L}$	set of POIs $\{i_1, i_2, ..., i_{\|\mathcal{L}\|}\}$
$\mathcal{G}$	geographical network
$\mathcal{L}_u^+$	set of (u, i) pairs
$\mathcal{L}_u^-$	set of (u, j) pairs
$\mathcal{L}_{ui}^{\mathcal{G}}$	set of (u, g) pairs
Θ	model parameters
W	users' latent factor matrix
H	POIs' latent factor matrix
b	POI bias
λ, β	regularization parameters
η	learning rate
k	the dimension of latent factors
$\hat{y}$	the ranking score calculated from decomposed models
$\hat{r}$	the ranking relations of two POIs rated by user u

Figure 2 represents the range of geographical neighbors. On the left side, (u, i_1) and (u, i_2) are observed rating pairs, i.e. $\mathcal{L}_u^+ = \{(u, i_1), (u, i_2)\}$; (u, g_1) and (u, g_2) represent unobserved pairs, where g_1 and g_2 are neighbors of i_1 and i_2, respectively, i.e. $\mathcal{L}_{ui_1}^{\mathcal{G}} = \{(u, g_1)\}$, $\mathcal{L}_{ui_2}^{\mathcal{G}} = \{(u, g_2)\}$; (u, j) represents the remaining unobserved pairs, i.e. $\mathcal{L}_u^- = \{(u, j)\}$. On the right side of Figure 2, g is a common neighbor of i_1 and i_2[5].

The goal of this work is to recommend each user a personalized ranked list of POIs from $\mathcal{L} \backslash \mathcal{L}_u^+$. Motivated by geo-spatial proximity, the key challenge is to learn individuals' implicit preference by integrating positive, geographical and negative feedback.

B. BPR: Ranking with Implicit Feedback

POIs that a user has never visited are either really unattractive or undiscovered yet potentially appealing [4]. This is the key challenge of POI recommendation based on implicit feedback. To tackle it, Rendle et al. [6] proposed a well-known ranking-based optimization criterion Bayesian personalized ranking (BPR) that maximizes a posterior estimation with Bayesian theory. An intuitive assumption is made: user u prefers item (i.e. POI in our case) i to item j, provided that (u, i) rating pair is observed and (u, j) is unobserved, defined by:

$$\hat{r}_{uij}(\Theta) := \hat{y}_{ui}(\Theta) \succ \hat{y}_{uj}(\Theta), i \in \mathcal{L}_u^+, j \in \mathcal{L} \backslash \mathcal{L}_u^+ \quad (1)$$

where Θ denotes a set of parameters of a ranking function (i.e. matrix factorization in this work), $\hat{y}_{ui}(\Theta)$ and $\hat{y}_{uj}(\Theta)$ are the predicted score by the ranking function, $\hat{r}_{uij}(\Theta)$ says i is preferred over j by u[6]. A unique characteristic of BPR is to sort pairwise preference $\hat{y}_{ui}$ and $\hat{y}_{uj}$ instead of regressing a predictor to a numeric value.

V. The GeoBPR Model

A. Model Assumption

The pairwise preference assumption of BPR, holds in practice, empirically produces much better performance than pointwise prediction methods [6][7]. However, we

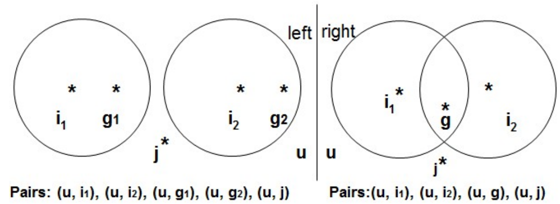


Fig. 2: Two scenarios of user-POI pairs

observe that there are two drawbacks in the BPR assumption for POI recommendation tasks: (1) The BPR algorithm is originally designed for general item recommendations[7], where the structure of geo-spatial preference has not been explicitly considered. Although the factors decomposed from matrix are semantically latent, there is no evidence showing that the latent space has included geographical features. Furthermore, leveraging geographical influence explicitly has been confirmed effectively in the regression task, e.g. [2][4]. (2) A large number of non-observed user-POI pairs cannot be employed for learning since BPR treats non-positive pairs equally. Thus we believe there is much room for improvement by exploiting geographical proximity influence between users and POIs. Specifically, we propose a novel assumption by explicitly modeling the structure of geographical proximity factors.

Assumption-a: As stated in section III-C, individuals tend to visit nearby places. Hence, we devise an intermediate process to enhance the BPR assumption: user u prefers POI i to POI g, provided that (u, i) rating pair is observed and (u, g) is unobserved, where g is one of the geographical neighbors of i; moreover, u prefers g to j, provided that (u, j) is unobserved and j is not a geographical neighbor of all rated POIs. This assumption can be formulated as follows:

$$\underbrace{\hat{y}_{ui} \succ \hat{y}_{ug}}_{:=\hat{r}_{uig}} \wedge \underbrace{\hat{y}_{ug} \succ \hat{y}_{uj}}_{:=\hat{r}_{ugj}}, i \in \mathcal{L}_u^+, g \in \mathcal{L}_{ui}^{\mathcal{G}}, j \in \mathcal{L}_u^- \quad (2)$$

It can be seen the preference orders of non-observed pairs, i.e. $(u, g), (u, j)$ are now possible to be compared using our assumption. Thus it seems promising that the sparsity problem is likely to be alleviated. Moreover, based on this assumption and the sound transitivity scheme [6], it is easy to infer as follows:

$$\underbrace{\hat{y}_{ui} \succ \hat{y}_{ug}}_{:=\hat{r}_{uig}} \wedge \underbrace{\hat{y}_{ui} \succ \hat{y}_{uj}}_{:=\hat{r}_{uij}} \quad (3)$$

We can see the assumption based on Eq.(3) reduces to that of BPR (see Eq.(1)). In other words, the new assumption leads to a more accurate interpretation than typical BPR assumption. Furthermore, the proposed assumption balances the contribution between geographical preference and latent factors[8]. Note that we cannot infer any preference relation from these pairs: $(\hat{y}_{ui_1}, \hat{y}_{ui_2})$, $(\hat{y}_{ui_1}, \hat{y}_{ug_2})$, $(\hat{y}_{ui_2}, \hat{y}_{ug_1})$, $(\hat{y}_{ug_1}, \hat{y}_{ug_2})$) (see Figure 2).

[5] Further details with Figure 2 can also be found in Section V-A.

[6] Throughout this work, we will write $\hat{r}_{uij}$ for $\hat{r}_{uij}(\Theta)$ to shorten notation, and the same applies to $\hat{r}_{uig}(\Theta)$, $\hat{r}_{ugj}(\Theta)$, $y_{ui}(\Theta)$, $y_{ug}(\Theta)$ and $y_{uj}(\Theta)$.

[7] An item can be anything, e.g. a book, a song as well as a POI.

[8] $\hat{y}$ is usually computed by a latent factor model, i.e. matrix factorization in this work.

Assumption-b: We are also interested in investigating an opposite assumption since unvisited POIs near a frequently visited POI are likely to be unattractive. This is because the user is likely to know about these POIs since they are close to her frequently visited ones, yet she has never chosen to patronize them before. This might be a signal that she dislikes them. In other words, geographical neighbors should be treated more negatively than other unvisited ones. Visiting frequency here is employed as the confidence of a user's preference. However, on the Yelp datasets, each POI has at most one rating by each user. Intuitively, this assumption may not hold without frequency information. For the sake of completeness of this work, we also verify the effectiveness of this assumption, formulated as follows:

$$\underbrace{\hat{y}_{ui} \succ \hat{y}_{uj}}_{:=\hat{r}_{uij}} \wedge \underbrace{\hat{y}_{uj} \succ \hat{y}_{ug}}_{:=\hat{r}_{ujg}}, i \in \mathcal{L}_u^+, g \in \mathcal{L}_{ui}^{\mathcal{G}}, j \in \mathcal{L}_u^- \quad (4)$$

Due to the space limitations, we merely elaborate the derivation process of our approach with assumption-a and report the final performance of the two assumptions in section VI.

B. Model Derivation

Based on the above assumptions, we employ a maximum posterior estimator to find the best ranking for a specific user u:

$$\arg\max_{\Theta} \mathcal{P}(\Theta| >_u) \quad (5)$$

where Θ represents a set of model parameters as mentioned before, $>_u$ is the total order, which represents the desired but latent preference structure for user u. According to Bayesian theory, the $\mathcal{P}(\Theta| >_u)$ can be inferred as:

$$\mathcal{P}(\Theta| >_u) \propto \mathcal{P}(>_u |\Theta)\mathcal{P}(\Theta) \quad (6)$$

where $P(>_u |\Theta)$ is the likelihood function and $\mathcal{P}(\Theta)$ is the prior distribution of parameters Θ. Here three intuitive assumptions are made: (1) Each user's rating actions are independent of every other user. (2) The preference ordering of each triple of items (i, g, j) for a specific user is independent of the ordering of every other triple. (3) The preference ordering of (i, g) pair for a specific user is independent of the ordering of (g, j) one. Based on the assumptions, Bernoulli distribution over the binary random variable can be used to estimate the likelihood function as follows:

$$\begin{aligned}\prod_{u\in\mathcal{U}} \mathcal{P}(>_u |\Theta) = \\ \prod_{(u,i,g,j)\in\mathcal{U}\times\mathcal{L}\times\mathcal{L}\times\mathcal{L}} \mathcal{P}\left(\hat{y}_{ui} \succ \hat{y}_{ug} \wedge \hat{y}_{ug} \succ \hat{y}_{uj}|\Theta\right)^{\delta((u,i,g,j)\in D_s)} \\ \cdot \left(1 - \mathcal{P}(\hat{y}_{ui} \succ \hat{y}_{ug} \wedge \hat{y}_{ug} \succ \hat{y}_{uj}|\Theta)\right)^{\delta((u,i,g,j)\notin D_s)}\end{aligned} \quad (7)$$

D_s is a poset of $>_u$, which expresses the fact that that user u is assumed to prefer i over g, and prefer g over j, i.e. $D_s = \left\{(u, i, g, j)| i \in \mathcal{L}_u^+ \wedge g \in \mathcal{L}_{ui}^{\mathcal{G}} \wedge j \in \mathcal{L}_u^-\right\}$. $\delta(x)$ is a binary indicator with $\delta(x) = 1$ if x is true and $\delta(x) = 0$, otherwise. Due to the totality and antisymmetry [6] of a pairwise ordering scheme, Eq.(7) can be simplified to:

$$\begin{aligned}\prod_{u\in\mathcal{U}} \mathcal{P}(>_u |\Theta) = \prod_{u\in\mathcal{U}, i\in\mathcal{L}_u^+, g\in\mathcal{L}_{ui}^{\mathcal{G}}} \mathcal{P}(\hat{y}_{ui} \succ \hat{y}_{ug}|\Theta) \\ \prod_{u\in\mathcal{U}, g\in\mathcal{L}_{ui}^{\mathcal{G}}, j\in\mathcal{L}_u^-} \mathcal{P}(\hat{y}_{ug} \succ \hat{y}_{uj}|\Theta)\end{aligned} \quad (8)$$

We employ a differential function, e.g. $\sigma(x) = \frac{1}{1+e^{-x}}$, to approximate the probability $\mathcal{P}(.)$ and map the value to probability range (0, 1). Unlike previous work, e.g. [6][7][20], which assign an equal weight to each training pair, in this paper, we design a weight function w_{ig} to control the contribution of the preference ordering between $\hat{y}_{ui}$ and $\hat{y}_{ug}$ so as to relax the assumption. Specifically, the two estimators can be derived as:

$$\begin{aligned}\mathcal{P}(\hat{y}_{ui} \succ \hat{y}_{ug}|\Theta) = \frac{1}{1 + e^{-w_{ig}(\hat{y}_{ui} - \hat{y}_{ug})}} \\ \mathcal{P}(\hat{y}_{ug} \succ \hat{y}_{uj}|\Theta) = \frac{1}{1 + e^{-(\hat{y}_{ug} - \hat{y}_{uj})}}, w_{ig} = \frac{1}{1 + n_{ig}}\end{aligned} \quad (9)$$

where w_{ig} is to control the contribution of sampled training pair (u, i) and (u, g) to the objective function, n_{ig} is the number of rated POIs that are geographical neighbors of POI g. w_{ig} equals 1 if no other rated POI shares g as a geographical neighbor, and the value decreases if g is a public geographical neighbor. The reason behind is that the above assumption of pairwise preference may not always hold in real applications. For example, POI g may be a geographical neighbor of more than one rated POIs (see right side of Figure 2). In this case, a user u may potentially prefer an POI g to POI i because g is close to the activity center of u. With this setting, the contribution of geographical preference works more reasonably.

Regarding prior density $\mathcal{P}(\Theta)$, it is common practice to design a Gaussian distribution with zero mean and model specific variance-covariance matrix $\lambda_\Theta I$, i.e. $\mathcal{P}(\Theta) \sim \mathcal{N}(0, \lambda_\Theta I)$.

Finally, we reach the objective loss function of our GeoBPR:

$$\begin{aligned}GeoBPR := argmax_\Theta \mathcal{P}(\Theta|\mathcal{D}_s) := argmin_\Theta (\lambda_\Theta ||\Theta||^2 \\ - \sum_{u\in\mathcal{U}, i\in\mathcal{L}_u^+, g\in\mathcal{L}_{ui}^{\mathcal{G}}} \ln\sigma(w_{ig}(\hat{y}_{ui} - \hat{y}_{ug})) \\ - \sum_{u\in\mathcal{U}, g\in\mathcal{L}_{ui}^{\mathcal{G}}, j\in\mathcal{L}_u^-} \ln\sigma(\hat{y}_{ug} - \hat{y}_{uj}))\end{aligned} \quad (10)$$

The prediction function $\hat{y}$ is modelled by matrix factorization, which is well known for discovering the underlying interactions between users and items.

$$\hat{y}_{ui} = W_u \cdot H_i^T + b_i = \sum_{f=1}^{k} w_{u,f} \times h_{i,f} + b_i \quad (11)$$

where W_u and H_i^T represent latent factors of user u and POI i resp., b_i is the bias term of i[9]., i.e. $\Theta = \{W \in$

[9]The user bias term vananishes for predicting rankings and for optimization as the pairwse comparison is based on one user level.

$R^{\mathcal{U}\times k}, H \in R^{\mathcal{L}\times k}, b \in R^{\mathcal{L}}\}$. The similar ranking functions apply to $\hat{y}_{ug}$ and $\hat{y}_{uj}$.

Algorithm 1: GeoBPR Learning

```
Input: D_s, G(L, L)
Output: model parameters Θ
Initialize Θ with Normal distribution N(0, 0.1)
for u ∈ U do
    Calculate L_u^+, L_ui^G, L_u^-
end
repeat
    for u ∈ U do
        Uniformly draw (i, g, j) from L_u^+, L_ui^G, L_u^-
        Calculate c_ig, c_gj, i.e.
        c_ig = 1 / (1 + e^{w_ig (ŷ_ui − ŷ_ug)}) · w_ig, c_gj = 1 / (1 + e^{ŷ_ug − ŷ_uj})
        W_u ←
        W_u + η(c_ig(H_i − H_g) + c_gj(H_g − H_j) − λ_u W_u)
        H_i ← H_i + η(c_ig W_u − λ_i H_i)
        H_g ← H_g + η(−c_ig W_u + c_gj W_u − λ_g H_g)
        b_i ← b_i + η(c_ig − β_i b_i)
        b_g ← b_g + η(−c_ig + c_gj − β_g b_g)
        b_j ← b_j + η(−c_gj − β_j b_j)
    end
until convergence;
return Θ
```

Discussion. According to Eq.(10) & Eq.(11), we observe that GeoBPR models the user's preference rankings by taking into account two types of factors: (1) due to the spatial proximity of (i, g) pair, the difference of ($\hat{y}_{ui}$, $\hat{y}_{ug}$) models the preference relations mostly based on general latent features[10], such as users' latent factors (i.e. the taste of the user) and POI latent factors (e.g. POI styles, item price, service reputation, etc.); (2) by explicitly modeling the difference of ($\hat{y}_{ug}$, $\hat{y}_{uj}$), the factorization model is likely to learn more about the structure of geo-spatial preference. Thus, by the optimal balance between the geographical influence and the latent features, even the POIs far away from the previously rated location have the chance to be recommended when personal preference dominates.

C. Model Learning

Since Eq.(10) is differentiable, we adopt the widely used stochastic gradient descent (SGD) for optimization. Specifically, for each user u, we randomly select (i, g, j) triples from positive, geographical and negative feedback, and then iteratively update parameters Θ. The update equations are given in Algorithm 1. Regarding the computational complexity, we can see that each update rule is $\mathcal{O}(k)$, where k is the number of latent dimensions in Eq.(11). The total complexity is $\mathcal{O}(\mathcal{T}|\mathcal{U}|k)$, where $\mathcal{T}$ is the number of iterations. For predicting a user's preference on a POI, the complexity is linear $\mathcal{O}(k)$. Both learning and predicting processes do not increase the time complexity in contrast with BPR.

VI. Experiments

A. Experimental Setup

Yelp datasets described in section III-A are used for evaluation. All experiments are conducted by using 5-fold cross-validation. Specifically, we randomly split each dataset into five folds and in each iteration four folds are used as the training set and the remaining fold as the testing set. The average results over 5 folds are reported as the final performance.

Baseline Methods. We compare GeoBPR[11] with an array of strong baselines described as follows. **Random (Rand)**: For the sake of understanding the ranking effectiveness of different algorithms, we launch a random one by randomly ordering the candidate POIs and then create a list of recommended locations. **Most Popular (MP)** [7][11]: It is to recommend users with the top-N most popular POIs. The popularity of POIs are computed by the number of ratings they received. **User-based Collaborative Filtering (UCF)** [4][21]: It is a typical memory-based collaborative filtering technique for both *rating prediction* and *item recommendation* tasks. The preference of a user to a candidate POI is calculated as an aggregation of some similar users' preference on POI. Pearson correlation is used to calculate user similarity. Then, the top-30 most similar users are selected as nearest neighbors. **MFM**: The basic idea of MFM is to fuse Multi-center features [2] with Factorization Machines (FM) [25]. Following [2], we produce several clusters based on a user's previously visited POIs and get the average coordinate of each cluster as a centroid. The distance between a candidate POI and each centroid is calculated as features. Then we apply FM to model users' latent preference and geo-spatial influence. **BPR** [6]: It is a state-of-the-art algorithm optimized for *top-N item recommendations*. As our proposed model is extended from BPR, we consider it as the main method used for comparison. **NBPR**: Inspired by [21], we implement the baseline by fusing geographical neighborhood with matrix factorization, and then adopts BPR criterion for learning.

Parameter Settings. Stochastic gradient descent (SGD) has several critical hyperparameters to be tuned, which are: (1) Learning rate η: To conduct a fair comparison, we apply the 5-fold cross-validation to find the best η for BPR ($\eta = 0.05$), and then employ the same value for GeoBPR. For MFM and NBPR, we apply the same procedure to tune η ($\eta = 0.005$). (2) Factorization dimension k: We fix $k = 30$ for all models based on matrix factorization. The effect of k value will be detailed later. (3) Regularization λ, β: In our paper, regularization parameters are grouped as GeoBPR has several parameters: λ_u represents the regularization parameter of W_u; λ_π represents the regularization parameters of H_i, H_g, H_j (i.e. λ_i, λ_g, λ_j resp.); β_π represents regularization parameters of b_i, b_g, b_j (i.e. β_i, β_g, β_j resp.). On Phoenix dataset, $\lambda_u = 0.03$, $\lambda_\pi = 0.03$, $\beta_\pi = 0.05$; on Las Vegas dataset, $\lambda_u = 0.08$, $\lambda_\pi = 0.02$, $\beta_\pi = 0.05$. (4) Initialization Θ: It is common practice to sample a zero-mean normal distribution with a small standard deviation σ. We set $\sigma = 0.1$ in this paper.

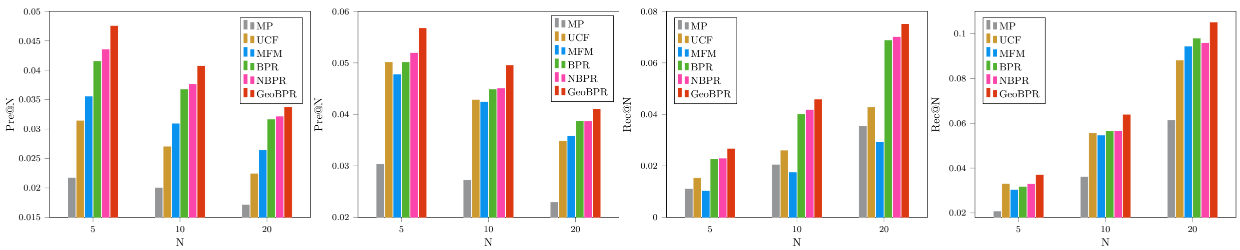
Evaluation Metrics. In order to measure the quality of *top-N recommendation* task, we choose four standard evaluation metrics, namely Precision@N and Recall@N (denoted

[10] Despite the success of latent factorization models, there is no literature to uncover the specific structure of latent factors, which is also beyond the scope of this paper.

[11] If not explicitly declared, GeoBPR is short for GeoBPR with assumption-a.

TABLE IV: Performance comparison where "*" means significant improvement in terms of paired t-test with p-value < 0.01, and symbols 'a' and 'b' of the GeoBPR model denote our two assumptions, respectively.

Dataset	Metrics	Rand	MP	UCF	MFM	BPR	NBPR	GeoBPRb	GeoBPRa	Improve
Phoenix	MAP	0.0012	0.0152	0.0187	0.0153	**0.0310**	0.0316	0.0095	**0.0335**	+8.06%*
	MRR	0.0043	0.0705	0.0939	0.1129	**0.1244**	0.1286	0.0507	**0.1406**	+13.02%*
Las Vegas	MAP	0.0016	0.0259	0.0372	0.0403	**0.0419**	0.0426	0.0167	**0.0462**	+10.26%*
	MRR	0.0055	0.0940	0.1422	0.1385	**0.1467**	0.1484	0.0702	**0.1656**	+12.88%*



(a) Phoenix **(b)** LasVegas **(c)** Phoenix **(d)** LasVegas

Fig. 3: Performance comparison with respect to top-N values in terms of Pre@N and Rec@N.

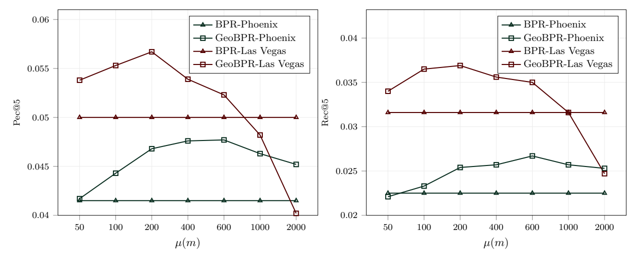
by Pre@N and Rec@N respectively) [7], Mean Average Precision (MAP) [14] and Mean Reciprocal Rank (MRR) [23]. For each metric, we first calculate the performance of each user from the testing data, and then obtain the average performance over all users. For saving space, we leave out detailed descriptions.

B. Experimental Results

Summary of Experimental Results. Table IV and Figure 3 present the experimental results of each algorithm in terms of the four ranking metrics.

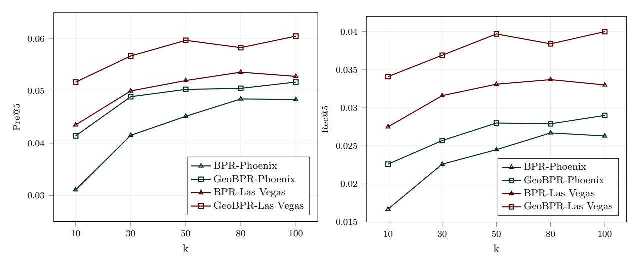
We highlight the results of BPR and GeoBPR in boldface for comparison in Table IV. The percentage in 'Improve' column represents the accuracy improvement of GeoBPR relative to BPR. As shown, BPR, NBPR and GeoBPR models perform much better than MP, MFM and UCF, which demonstrates the effectiveness of pairwise preference assumptions. Our approach outperforms the other baseline methods in terms of all the metrics on both datasets. In particular, our GeoBPR model achieves about 10% significant improvement compared to the BPR model in terms of MAP and MRR. The main reason is that BPR only learns one ranking order between observed and non-observed POI pairs, i.e. $(i,\ j)$. While our GeoBPR model learns two orders: rated POI i and nearby POI g which is unrated, i.e. $\hat{r}_{uig}$; both unrated POIs g and j, but j is distant from all rated POIs, i.e. $\hat{r}_{ugj}$. Intuitively, the assumption $\hat{r}_{uig}$ holds more accurately than $\hat{r}_{uij}$ in real scenarios; in addition, sparsity problem seems to be alleviated by the additional assumption $\hat{r}_{ugj}$. We can thus see that the assumption of GeoBPR by injecting geo-spatial preference is indeed more effective than that of simple pairwise preference assumed in BPR. Interestingly, one may observe that NBPR does not perform much better than BPR by modeling a new prediction function, i.e. fusing geographical neighborhood with matrix factorization. Our results potentially imply that algorithms optimized for *rating prediction* do not translate into accuracy improvements in terms of *top-N item recommendation*.

Impact of Neighborhood. Table IV shows the prediction quality of GeoBPR with two opposite assumptions, i.e.



(a) Pre@5 **(b)** Rec@5

Fig. 4: Performance comparison with different μ



(a) Pre@5 **(b)** Rec@5

Fig. 5: Performance comparison with different k

assumption-a and assumption-b (denoted by GeoBPRa, GeoBPRb resp.). As stated in section V-A, assumption-b intuitively cannot hold due to the lack of frequency information, we can see that GeoBPRb performs the worst over other approaches except for the Random one. To further investigate the contribution of geo-spatial preference, we propose using different thresholds μ mentioned in section III-C. The results, i.e. Pre@5 and Rec@5[12], are depicted in Figure 4. We observe the general trends are, both metrics increase with the increasing of threshold μ, when arriving at a certain threshold, the performance starts decreasing with a larger μ. The reason is that the size of nearby POIs ($g \in \mathcal{L}_{ui}^{\mathcal{G}}$) is not as large as required for training the

[12]The performance on other metrics follows similar trends.

model[13] when μ is small (e.g. $\mu \in [50, 200]$) (see Algorithm 1 and Table I). Once the number of training samples is large enough, the performance of GeoBPR keeps consistent with the ratio value ($\mathcal{P}^{'}/\mathcal{P}$), i.e. the larger ratio the training samples have, the better recommendation quality GeoBPR achieves. For example, the ratio on Las Vegas dataset become smaller when μ is larger than 600m, and accordingly the recommendation accuracy degrades rapidly when μ increases. Furthermore, we see GeoBPR always achieves better performance than BPR on Phoenix dataset when μ in $[100, 2000]$; on both datasets, it outperforms BPR when μ in [100, 600].

Impact of Factorization Dimensions. In this work, we apply a matrix factorization (MF) as the scoring function for GeoBPR (see section V-B). Thus it is important to investigate the impact of factorization dimension k to the prediction quality. As shown in Figure 5, the performance of BPR and GeoBPR steadily rise with the increasing number of dimensions, which keeps consistent with previous work, e.g. [6][20]. Furthermore, GeoBPR consistently outperforms BPR with the same number of dimensions on both datasets; in particular, the performance of GeoBPR in 30 dimensions is comparable with that of BPR in 100 dimensions.

VII. Conclusion

In this paper, we explored to leverage geographical influence to improve personalized POI recommendation. First, to motivate this work, we conducted proximity data analysis on two real-world datasets extracted from the Yelp Datasets and observed that a user's rated POIs tend to cluster together on the map. Thus it is reasonable to argue that users are likely to visit nearby places. Then, we presented a new pairwise preference assumption and proposed a co-pairwise ranking model (GeoBPR) by injecting the geo-spatial preference. The intermediate proximity preference introduced by geographical feedback leads to a more accurate interpretation than original BPR in the setting of POI recommendation, and makes the preference ordering of non-observed user-POI pairs possible to be inferred. Due to the optimal balance of geographical preference and latent factors, GeoBPR outperformed other state-of-the-art factorization models. Both the theoretical and empirical results indicated that GeoBPR was the right choice for personalized POI recommendation task.

References

[1] H. Gao, J. Tang, X. Hu, and H. Liu, "Exploring temporal effects for location recommendation on location-based social networks," in *RecSys*, 2013, pp. 93–100.

[2] C. Cheng, H. Yang, I. King, and M. R. Lyu, "Fused matrix factorization with geographical and social influence in location-based social networks," in *AAAI*, 2012.

[3] M. Ye, X. Liu, and W.-C. Lee, "Exploring social influence for recommendation: a generative model approach," in *SIGIR*, 2012, pp. 671–680.

[13] Once again, we emphasize that the ranking comparison of (i, g) pairs models general latent factors, while the comparsion of (g, j) explicitly models geo-spatial preference.

[4] D. Lian, C. Zhao, X. Xie, G. Sun, E. Chen, and Y. Rui, "Geomf: joint geographical modeling and matrix factorization for point-of-interest recommendation," in *SIGKDD*, 2014, pp. 831–840.

[5] J.-D. Zhang and C.-Y. Chow, "Geosoca: Exploiting geographical, social and categorical correlations for point-of-interest recommendations," in *SIGIR*, 2015, pp. 443–452.

[6] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme, "Bpr: Bayesian personalized ranking from implicit feedback," in *UAI*, 2009.

[7] W. Pan and L. Chen, "Gbpr: Group preference based bayesian personalized ranking for one-class collaborative filtering," in *IJCAI*, 2013, pp. 2691–2697.

[8] Y. Yu and X. Chen, "A survey of point-of-interest recommendation in location-based social networks," in *AAAI*, 2015.

[9] B. Berjani and T. Strufe, "A recommendation system for spots in location-based online social networks," in *Proceedings of the 4th Workshop on Social Network Systems*, 2011, p. 4.

[10] M. Ye, P. Yin, W.-C. Lee, and D.-L. Lee, "Exploiting geographical influence for collaborative point-of-interest recommendation," in *SIGIR*, ser. SIGIR '11, 2011, pp. 325–334.

[11] Q. Yuan, G. Cong, Z. Ma, A. Sun, and N. M. Thalmann, "Time-aware point-of-interest recommendation," in *SIGIR*, ser. SIGIR '13, 2013, pp. 363–372.

[12] R. Pan, Y. Zhou, B. Cao, N. N. Liu, R. Lukose, M. Scholz, and Q. Yang, "One-class collaborative filtering," in *ICDM*, 2008, pp. 502–511.

[13] S. Rendle, L. Balby Marinho, A. Nanopoulos, and L. Schmidt-Thieme, "Learning optimal ranking with tensor factorization for tag recommendation," in *SIGKDD*, 2009, pp. 727–736.

[14] S. Rendle and C. Freudenthaler, "Improving pairwise learning for item recommendation from implicit feedback," in *WSDM*, 2014, pp. 273–282.

[15] A. Krohn-Grimberghe, L. Drumond, C. Freudenthaler, and L. Schmidt-Thieme, "Multi-relational matrix factorization using bayesian personalized ranking for social network data," in *WSDM*, 2012, pp. 173–182.

[16] T. Zhao, J. McAuley, and I. King, "Leveraging social connections to improve personalized ranking for collaborative filtering," in *CIKM*, 2014, pp. 261–270.

[17] X. Li, G. Cong, X.-L. Li, T.-A. N. Pham, and S. Krishnaswamy, "Rank-geofm: a ranking based geographical factorization method for point of interest recommendation," in *SIGIR*, 2015, pp. 433–442.

[18] J. He, X. Li, L. Liao, D. Song, and W. K. Cheung, "Inferring a personalized next point-of-interest recommendation model with latent behavior patterns," in *AAAI*, 2016.

[19] C. Cheng, H. Yang, M. R. Lyu, and I. King, "Where you like to go next: Successive point-of-interest recommendation." 2013.

[20] S. Rendle and L. Schmidt-Thieme, "Pairwise interaction tensor factorization for personalized tag recommendation," in *WSDM*, 2010, pp. 81–90.

[21] L. Hu, A. Sun, and Y. Liu, "Your neighbors affect your ratings: on geographical neighborhood influence to rating prediction," in *SIGIR*, 2014, pp. 345–354.

[22] B. Liu, H. Xiong, S. Papadimitriou, Y. Fu, and Z. Yao, "A general geographical probabilistic factor model for point of interest recommendation," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 27, pp. 1167–1179, 2015.

[23] Y. Shi, A. Karatzoglou, L. Baltrunas, M. Larson, N. Oliver, and A. Hanjalic, "CLiMF: learning to maximize reciprocal rank with collaborative less-is-more filtering," in *RecSys*, 2012, pp. 139–146.

[24] W. R. Tobler, "A computer movie simulating urban growth in the detroit region," *Economic geography*, pp. 234–240, 1970.

[25] S. Rendle, "Factorization machines," in *ICDM*, 2010, pp. 995–1000.