Exploiting Ranking Consistency Principle in Representation Learning for Location Promotion

Siyuan Zhang[1(✉)], Yu Rong[2], Yu Zheng[3], Hong Cheng[1], and Junzhou Huang[2]

[1] The Chinese University of Hong Kong, Hong Kong, China
{syzhang,hcheng}@se.cuhk.edu.hk
[2] Tencent AI Lab, Shenzhen, China
{royrong,joehhuang}@tencent.com
[3] Microsoft Research, Beijing, China
yuzheng@microsoft.com

Abstract. Location-based services, which use information of people's geographical position as service context, are becoming part of our daily life. Given the large volume of heterogeneous data generated by location-based services, one important problem is to estimate the visiting probability of users who haven't visited a target Point of Interest (POI) yet, and return the target user list based on their visiting probabilities. This problem is called the *location promotion problem*. The location promotion problem has not been well studied due to the following difficulties: (1) the cold start POI problem: a target POI for promotion can be a new POI with no check-in records; and (2) heterogeneous information integration. Existing methods mainly focus on developing a general mobility model for all users' check-ins, but ignore the ranking utility from the perspective of POIs and the interaction between geographical and preference influence of POIs.

In order to overcome the limitations of existing studies, we propose a unified representation learning framework called *hybrid ranking and embedding*. The core idea of our method is to exploit the ranking consistency principle into the representation learning of POIs. Our method not only enables the interaction between the geographical and preference influence for both users and POIs under a ranking scheme, but also integrates heterogeneous semantic information of POIs to learn a unified preference representation. Extensive experiments show that our method can return a ranked user list with better ranking utility than the state-of-the-art methods for both existing POIs and new POIs. Moreover, the performance of our method with respect to different POI categories is consistent with the hierarchy of needs in human life.

Keywords: Location promotion · Ranking consistency
Graph embedding

J. Pei et al. (Eds.): DASFAA 2018, LNCS 10828, pp. 457–473, 2018.
https://doi.org/10.1007/978-3-319-91458-9_28

1 Introduction

Location-based services, such as Foursquare, Yelp and Facebook Place, are becoming increasingly popular these days. The data generated in a typical location-based service consist of two parts: the check-in records of users and the profiles of Points of Interest (POIs). Many methods have been developed for applications such as POI recommendation [2,6,9–11,13,14,19,21,22] and friend recommendation in location-based social networks [15,18]. Besides providing user-centric services as mentioned above, mining data from location-based services can also help local companies to promote their business more effectively. For example, a new restaurant owner at Pittsburgh would like to know the target users who are more likely to have dinner at his restaurant according to their check-in histories so that he can distribute coupons to them.

Different from existing studies on POI recommendation [2,6,9–11,13,14,19, 21,22], which recommends POIs to a target user as a user-centric task, we need to recommend users to a target POI in the coupon distribution scenario. In this paper, we define the task of ranking users according to their visiting probabilities to a target POI as a **location promotion problem**. There are two major challenges for this problem: (1) the cold start POI problem, i.e., a target POI for promotion can be a new POI with no check-in records; and (2) heterogeneous information integration, as there are both geographical and semantic information associated with POIs. Existing POI recommendation methods [10,11,13,22] cannot be directly applied to solve the location promotion problem with satisfactory performance.

Existing solutions [21,28] for location promotion is unsatisfactory due to two reasons. *First*, they build their models by maximizing the likelihood of observing all check-in records. They ignore "unobserved" POI-user pairs, i.e., those users with no check-in at certain POIs, in their models. But such "null" relationship can be combined with the check-in records to help infer the ranking of potential users to specific POIs. *Second*, they consider modeling interaction between users and POIs in only one space. [28] only considers the geographical proximity between users and POIs, i.e., geographical space. [21] only considers representing users and POIs in one POI latent space. However, using only one space is not enough for integrating heterogenous information in location-based service because different information may have different interaction patterns.

In order to overcome the limitations of existing studies, we propose a unified representation learning framework called **hybrid ranking and embedding (HRE)** to solve the location promotion problem. The core idea of our method is to **exploit the ranking consistency principle in the representation learning framework of POIs.** The *ranking consistency principle* states that users with more check-ins at the POI should be ranked higher than users with fewer check-ins or no check-in at the POI. With the ranking consistency principle, we can use the unobserved POI-user pairs to alleviate the data sparsity issue in check-in records. To measure the interaction between users and POIs, we use geographical embeddings and preference embeddings to represent both users and POIs. The geographical embeddings measure the geographical influence to users and POIs in each region. The preference embeddings measure the similarity of check-in patterns between users and between POIs. The final ranking score of users w.r.t. the target POI is estimated by their embeddings in both geographical space and preference space.

To learn the embeddings of new POIs in preference space as well as improve the ranking performance of existing POIs, we build different types of weighted bipartite POI-semantic graphs to capture and integrate heterogeneous semantic information of POIs. Our intuition is that POIs with similar semantic information should be close in preference space. Five kinds of information, including sequential visiting POIs, temporal check-in number in each time slot, region proximity, tags, and neighborhood visiting users, are considered for learning the preference embeddings for all POIs through multiple graph embeddings. A joint learning algorithm is proposed to train a unified preference representation for each POI.

Compared with existing solutions, our method has four advantages. *First*, we can gain better ranking performance for location promotion since we optimize from the POIs' perspective and integrate heterogeneous semantic information for POIs. *Second*, our hybrid model can learn preference embeddings for new POIs by sharing POI preference embeddings in different POI-semantic graphs. New POIs can utilize the collective check-in data of existing POIs to achieve better ranking performance. *Third*, the embeddings of users are divided into two spaces, the geographical and preference spaces, and trained by a joint learning algorithm, which enables the interaction between geographical influence and preference influence as well as preserving the heterogenous interaction pattern in different spaces. *Fourth*, our method can be easily extended to incorporate other kinds of semantic information such as the photos and comments posted by users at POIs.

We have made the following contributions in this paper.

– To the best of our knowledge, we are the first to incorporate the ranking consistency information into the representation learning framework of POIs, which has been ignored by current literatures.
– Our method can integrate both the geographical and semantic information of POIs. Existing POIs and new POIs can have a unified embedding in both geographical space and preference space.
– Extensive experiment results show that our method can boost the ranking performance measured by AUC and Tau for both existing POIs and new POIs, compared to several baselines and state-of-the-art methods. On average, we increase the AUC by 3.0% and Tau by 7.7% for all POIs in four cities of United States, compared to the second best solution.
– We conduct a case study to demonstrate that the performance of our method with respect to different POI categories is consistent with the hierarchy of needs in human life, which has not been reported by previous studies.

The remainder of this paper is organized as follows: Sect. 2 gives a formal definition of our problem and performs data analysis for geographical space embedding. Section 3 describes the hybrid ranking and embedding method and the parameter learning algorithm. Section 4 reports the experimental results. Section 5 reviews related work. Finally Sect. 6 concludes this paper.

2 Preliminaries

2.1 Problem Description

Assume we have a user set $U = \{u_1, u_2, \ldots, u_N\}$ and a POI set $L = \{l_1, l_2, ..., l_M\}$. A check-in record can be defined as a triple $c = (u, l, t)$ which indicates that user u performs a check-in action on POI l at time t. We denote $C_u = \{c_1, c_2, \ldots\}$ as the user u's check-in records, C_{ul} as the set of check-in records which are performed by user u on POI l, and U_l as the set of users that have check-in at l. Furthermore, we denote a POI as $l = (\omega_l, \tau_l)$, where ω_l is the coordinate of l and τ_l is the tag set of l. The *location promotion problem* can be defined as:

Definition 1 *(Location Promotion Problem). Given the user set* $U = \{u_1, u_2, \ldots, u_N\}$*, all users' check-in records* $\mathcal{C} = \{C_1, \ldots, C_N\}$ *and a target POI* l*, return a rank list of candidate users who have not visited POI* l *yet, i.e.,* $u \in U \setminus U_l$*, in descending order of their probability of visiting* l*.*

2.2 Data Analysis

Different from the traditional recommendation problem, user's check-in records not only contain the user's preference on POIs but also indicate the user's geographical preference. For example, the check-in records of a user usually cluster around his/her workplace and home [4].

In order to find a reasonable geographical embedding method for users and POIs, we perform data analysis on the check-in records from two cities Los Angeles and San Diego. We first use k-means algorithm to cluster POIs in a city into K regions according to their coordinates. Then we consider all sequential check-in records in pairs made by the same user, in the form of (l_{j-1}, l_j). Denote the region center where l_{j-1} belongs to as ω_k. We calculate the distance between ω_k and the next POI l_j. Then we plot the distribution of such distances from a region center to the next check-in POI in Fig. 1. We observe that the visiting probability decreases as the distance between the region center and next check-in POI increases.

To further choose a suitable distribution to describe the relationship, we try to fit the data with three alternatives: Pareto distribution, Exponential distribution and Log-normal distribution, and show the loglikelihood of different distributions in Table 1. We find that the Pareto distribution has the largest loglikelihood in modeling the distance distribution between the region center and the next check-in POI. Thus, we decide to use Pareto distribution in POIs' geographical embeddings.

Table 1. Loglikelihood of different distance distribution in Los Angeles and San Diego

City	Pareto	Exponential	Log-normal
Los Angeles	**−2.476**	−3.186	−2.879
San Diego	**−2.153**	−3.199	−2.839

3 Hybrid Ranking and Embedding Method

In this section, we first introduce the two building blocks of our unified model: ranking consistency model and graph based POI embedding. Then we describe our joint learning model.

3.1 Learning Geographical and Preference Embedding with Ranking Consistency

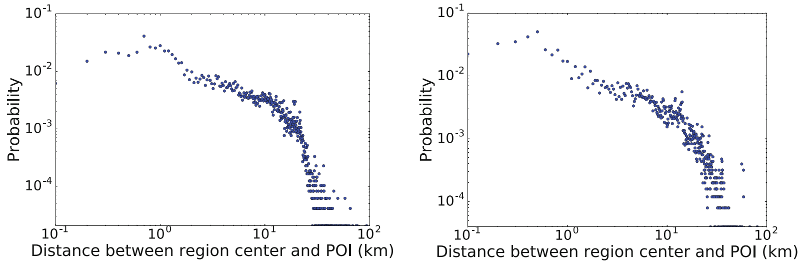
Based on the analysis on Sect. 2.2, we utilize two latent spaces $\mathcal{G}$ and $\mathcal{V}$ to model the geographical factor and the preference factor respectively.

Geographical Factor Embedding. We propose a geographical embedding method to encode the spatial proximity between candidate users and target POIs. For POI l_j, we denote its geographical feature by an embedding vector $\mathbf{g}_j \in \mathcal{G}$ as:

$$\mathbf{g}_j = [f(d(\omega_j, \omega_1)), \ldots, f(d(\omega_j, \omega_K))]^T. \quad (1)$$

In Eq. (1), $d(\cdot, \cdot)$ represents the distance between two coordinates and the k-th entry of $\mathbf{g}_j$ denotes the probability that a user moves from the k-th region to l_j. In general, $f(\cdot)$ can be any function that can output a probability. In our proposed framework, we adopt the Pareto distribution based on the analysis in Sect. 2.2. Following the setting in [28] where a similar observation is reported between sequentially visited POIs, the specific form of $\mathbf{g}_j$ is:

$$\mathbf{g}_j = [(1 + d(\omega_j, \omega_1))^{-\alpha}, \ldots, (1 + d(\omega_j, \omega_K))^{-\alpha}]^T. \quad (2)$$



(a) Los Angeles (b) San Diego

Fig. 1. The distance distribution between the sequential check-in region center and POI in two cities

α is a shape parameter in Pareto distribution and is estimated by maximum likelihood estimation.

The geographical embedding for user u_i can be defined as:

$$\mathbf{g}_i = [\gamma_{i,1}, \gamma_{i,2}, \ldots, \gamma_{i,K}]^T, \tag{3}$$

where $\gamma_{i,k}$ denotes the expected number of check-ins in the k-th region by u_i.

Using the geographical embedding of POIs and users, we define the ranking score y_{ij} that u_i will visit l_j as:

$$y_{ij} = \mathbf{g}_i \cdot \mathbf{g}_j. \tag{4}$$

Equation (4) can be considered as a two-layer model, where the inner product encompasses the weighted sum of the expected check-in number from user u_i's activity regions denoted by $\mathbf{g}_i$. The weight factor $\mathbf{g}_j$ of a target POI l_j only depends on its distance to each region center, which implies the spatial influence from each region to l_j.

Preference Factor Embedding. Though the geographical factor has a large impact on determining the ranking score of candidate users, there are still portions of check-in records that cannot be explained by the geographical factor. For example, when people choose a certain store they like to check-in in a shopping mall, the distance factor has less impact compared to users' preference. If we represent all users' check-in records with a POI-user matrix where each entry is the number of check-ins from a specific user at a POI, we can infer a user's preference by comparing his/her check-in records with those of other users who have similar check-in patterns.

To model the preference, we introduce an embedding vector $\mathbf{v}_i \in \mathcal{V}$ for u_i, and $\mathbf{v}_j \in \mathcal{V}$ for l_j to represent the preference factors of user u_i and POI l_j respectively. If u_i has check-ins at l_j, $\mathbf{v}_i \cdot \mathbf{v}_j$ should be larger than that of another user $u_{i'}$ who has no check-in at l_j. Under this assumption, two users who visit many common POIs should have close preference embeddings. On the other hand, POIs that share many common users should have close preference embeddings in $\mathcal{V}$. To combine the information from geographical coordinates and POI-user matrix, we define the final ranking score y_{ij} as:

$$y_{ij} = \mathbf{v}_i \cdot \mathbf{v}_j + \mathbf{g}_i \cdot \mathbf{g}_j. \tag{5}$$

In Eq. (5), $\mathbf{g}_i, \mathbf{g}_j \in \mathbb{R}_+^K$ indicate the feature vectors of user u_i and POI l_j in geographical space $\mathcal{G}$; $\mathbf{v}_i, \mathbf{v}_j \in \mathbb{R}^D$ indicate the feature vectors of user u_i and POI l_j in preference space $\mathcal{V}$.

Combining the scores in two separate spaces enables the interaction between geographical influence and preference influence. To illustrate this point, consider the following example. For a target shop in a shopping mall, users who have check-ins near the shopping mall are more likely to walk by and visit the target shop, i.e., from the perspective of the geographical embedding. On the other hand, users who have the similar preference to the target shop are also more likely to visit the target shop, i.e., from the perspective of preference embedding.

In our solution, the geographical embedding $\mathbf{g}_i$ for users and the preference embedding $\mathbf{v}_i$ for users and $\mathbf{v}_j$ for POIs are all trained using the check-in records because the influence of geographical embeddings and preference embeddings on check-in behaviors varies from user to user. Later we will describe how to learn the vectors $\mathbf{g}_i, \mathbf{v}_i, \mathbf{v}_j$ for all users and POIs in our inference algorithm.

Ranking Consistency Model. Given the definition of ranking score, we propose our ranking consistency model to learn the latent representation of users and POIs in both geographical and preference spaces simultaneously. The core idea behind our model lies in the *ranking consistency principle*, which indicates that the ranking score of different users should be consistent with their check-in records. Concretely, given a target POI l_j, the following constraints should be satisfied:

- A user who has performed check-in at l_j should be ranked higher than those who have not performed check-in at l_j.
- A user with more check-in records at l_j should be ranked higher than those with less check-in records at l_j.

Based on the *ranking consistency principle*, given POI l_j, our model can be defined as:

$$f(l_j|\Theta) = \prod_{u_i \in U_{l_j}} \prod_{u_{i'} \in U_{<|C_{ij}|}} P((y_{ij} - y_{i'j}) > 0|\Theta). \tag{6}$$

In Eq. (6), $U_{<|C_{ij}|} = \{u_{i'} || C_{i'j}| < |C_{ij}|\}$ is the set of users whose number of check-in records at POI l_j is less than that of user u_i. $\Theta = \{\mathbf{v}_i, \mathbf{v}_j, \mathbf{g}_i | u_i \in U, l_j \in L\}$ is the parameter set. $P((y_{ij} - y_{i'j}) > 0|\Theta)$ is the probability which indicates that the ranking score of user u_i is higher than that of user $u_{i'}$. Following the Bayesian personalized ranking scheme [16], we apply the logistic function $\sigma(x) = 1/(1 + \exp(-x))$ to output the probability $P((y_{ij} - y_{i'j}) > 0|\Theta)$:

$$P((y_{ij} - y_{i'j}) > 0|\Theta) = \frac{1}{1 + e^{-(y_{ij} - y_{i'j})}}. \tag{7}$$

Therefore, given a POI set L, the log-likelihood function $\mathcal{F}_{RC}(L, \Theta)$ is written as:

$$\mathcal{F}_{RC}(L, \Theta) = \sum_{l_j \in L} \log f(l_j|\Theta) - \lambda ||\Theta||^2. \tag{8}$$

In Eq. (8), $\lambda||\Theta||^2$ is the Gaussian prior for regularization. Compared with the methods that directly approximate the check-in frequency like rating based recommendation problems, Bayesian personalized ranking criterion learns the ranking models based on pairwise comparison of users such that the area under the ROC curve (AUC) can be maximized [16]. On the other hand, it alleviates the data sparsity problem in modeling check-in records by fully utilizing information of the number of check-in records of users at a POI. This happens to meet our goal for location promotion.

3.2 Learning POI Semantic with Graph Based Embedding

Although the ranking based embedding learning method can alleviate the data sparsity problem in check-in data, it cannot handle new POIs without any check-in record. Other than the geographical and user preference information, POIs in location-based services also contain rich semantic information, such as tags, spatial check-in relations, temporal check-in patterns and neighborhoods.

In order to handle new POIs and enhance the representation power of our POI embeddings, based on the intuition that *POIs with similar semantic information would share common potential target users*, we utilize the POI semantic information to infer the preference embeddings of all POIs. This would be particularly useful for new POIs with no check-in records.

POI-semantic Graph Construction. We first extract the semantic information for POIs by constructing POI-semantic graphs. A POI-semantic graph is a bipartite graph carrying weights on the edges. According to different semantics, we design five types of POI-semantic graphs as follows.

- POI-POI graph: It is denoted as $G_{LL} = (L, L, E_{LL}, W_{LL})$ and designed for capturing POI check-in sequential relationship. The two node sets are both the POI set L. If there exists a user who visits two POIs $l_j, l_k \in L$ sequentially and the time gap between the two visits is less than a threshold ΔT, we add an edge $e_{jk} \in E_{LL}$ from l_j to l_k. The weight $w_{jk} \in W_{LL}$ of e_{jk} is defined as the number of such sequential visits between l_j and l_k in all check-in records.
- POI-Time graph: It is denoted as $G_{LT} = (L, T, E_{LT}, W_{LT})$ and designed for capturing POI temporal visit patterns. One node set is the POI set L, and the other node set is T representing different time slots. We first divide all the check-in timestamps into 24 time slots and denote each time slot as a node in T. For a POI $l_j \in L$ and a time slot $t_k \in T$, an edge $e_{jk} \in E_{LT}$ denotes that there are check-ins in l_j at t_k. The weight $w_{jk} \in W_{LT}$ for e_{jk} is the number of check-ins in l_j at t_k.
- POI-Tag graph: It is denoted as $G_{LW} = (L, W, E_{LW}, W_{LW})$ and designed for capturing functions of POIs. One node set is the POI set L, and the other node set is W representing different tags (such as "*Chinese Restaurant*" and "*Coffee*") of POIs. For a POI $l_j \in L$ and a tag $w_k \in W$, an edge $e_{jk} \in E_{LW}$ exists if l_j has tag w_k, and the weight $w_{jk} \in W_{WL}$ is defined as the *tf.idf* value.
- POI-Region graph: It is denoted as $G_{LR} = (L, R, E_{LR}, W_{LR})$ and designed for capturing region influence to POIs. One node set is the POI set L, and the other node set is R representing the K regions defined in Sect. 3.1. For a POI $l_j \in L$ and a region $r_k \in R$, an edge $e_{jk} \in E_{LR}$ with a unit weight denotes that l_j belongs to r_k.

– POI-Neighborhood User graph: It is denoted as $G_{LU} = (L, U, E_{LU}, W_{LU})$ and designed for capturing the neighborhood visit patterns. One node set is the POI set L, and the other node set is the user set U. A user $u_i \in U$ may not have check-in records in $l_j \in L$, but has check-ins in the k-nearest neighbor POIs of l_j. Then we consider u_i as l_j's neighborhood user, i.e., whose check-ins are near l_j, and an edge $e_{ij} \in E_{LU}$ is added. The edge weight $w_{ij} \in W_{LU}$ is the number of check-ins at l_j's neighbor POIs by u_i.

Overall, there are five types of semantic information as described above, thus we use a semantic set $\mathcal{S} = \{L, T, R, W, U\}$ to denote them collectively. It is noted that the POI-POI, POI-Time and POI-Tag graphs are in the same form as defined in [21].

To solve the location promotion problem within a city, we further define POI-Region graph and POI-Neighborhood User graph to model different levels of geographical proximity between POIs. The construction of the two graphs helps us to learn a more geo-aware preference embedding for POIs and improve the ranking performance.

Learning POI and Semantic Embeddings. Based on the semantic information captured by the above POI-semantic graphs, we can learn the preference embeddings for new POIs. For the ease of presentation, we use a generic notation $G_{LS} = (L, S, E_{LS}, W_{LS})$ to denote a POI-semantic graph, which can be any specific type of the above five POI-semantic graphs. For the node set S in the POI-semantic graph, we call the nodes $s_1, s_2, \ldots, s_m \in S$ *semantic nodes*. Our target is to map these semantic nodes as well as the POIs to the preference latent space $\mathcal{V}$. We define the empirical conditional probability that s_k can be represented by l_j as:

$$\widehat{p}(s_k|l_j) = \frac{w_{jk}}{\sum_{s_{k'} \in S_{l_j}} w_{jk'}}, \tag{9}$$

where $S_{l_j} \subseteq S$ is the semantic node set related to POI l_j.

In the preference latent space $\mathcal{V}$, the embedding vector of s_k is $\mathbf{v}_k$. We use the softmax function to model the conditional probability that s_k can be represented by l_j in $\mathcal{V}$:

$$p(s_k|l_j) = \frac{\exp(\mathbf{v}_k \cdot \mathbf{v}_j)}{\sum_{s_{k'} \in S_{l_j}} \exp(\mathbf{v}_{k'} \cdot \mathbf{v}_j)}. \tag{10}$$

Since we have $\sum_{s_k \in S} \widehat{p}(s_k|l_j) = 1$ and $\sum_{s_k \in S} p(s_k|l_j) = 1$, given the POI l_j, the conditional probability over semantic information S can be treated as a distribution which is denoted as $\mathcal{P}_{l_j}$ (empirical distribution as $\widehat{\mathcal{P}}_{l_j}$). The objective of the embedding is to make the conditional distribution $\mathcal{P}_{l_j}$ close to the empirical distribution $\widehat{\mathcal{P}}_{l_j}$ for all POIs. We use the KL-divergence $KL(.,.)$ to measure the distance between two distributions. Thus the objective function for the semantic information S is written as:

$$\mathcal{F}(S, L|\Theta_S) = -\sum_{l_j \in L} \sigma_j KL(\widehat{\mathcal{P}}_{l_j}, \mathcal{P}_{l_j})$$
$$= \sum_{e_{jk} \in E_{LS}} \omega_{jk} \log p(s_k|l_j). \tag{11}$$

In Eq. (11), $\Theta_S = \{\mathbf{v}_k | s_k \in S\}$ is the embedding vector set of semantic nodes in S. $\sigma_j = \sum_{s_k \in S} \omega_{jk}$ is the importance of node l_j.

As there are five types of POI-semantic graphs, we construct a combinational graph-based POI embedding scheme. Given the semantic information set $\mathcal{S} = \{L, T, R, W, U\}$, the objective function can be written as:

$$\mathcal{F}_{EMB}(\mathcal{S}, L|\Theta_{\mathcal{S}}) = \sum_{S \in \mathcal{S}} \mathcal{F}(S, L|\Theta_S). \tag{12}$$

In Eq. (12), each POI in L has a unique preference embedding among different POI-semantic graphs. Therefore, the learned POI preference embeddings integrate heterogeneous information related to POIs and improve the ranking performance for all POIs as confirmed in experiments.

3.3 The Unified Model

We propose to jointly learn the embeddings of all elements simultaneously by sharing the preference embedding of POIs. The final objective function of our unified model is written as:

$$\mathcal{F}(L, \mathcal{S}, \Theta, \Theta_{\mathcal{S}}) = \mathcal{F}_{RC}(L, \Theta) + \beta \mathcal{F}_{EMB}(\mathcal{S}, L|\Theta_{\mathcal{S}}), \tag{13}$$

where $\Theta = \{\mathbf{v}_i, \mathbf{v}_j, \mathbf{g}_i | u_i \in U, l_j \in L\}$ and $\Theta_{\mathcal{S}} = \{\mathbf{v}_k | s_k \in S, S \in \mathcal{S}\}$. After learning the embeddings of POIs, users and all kinds of semantic information, we can calculate the visiting score for each unobserved POI-user pair and rank the unvisited users w.r.t. a POI according to Eq. (5). Our model can be trained efficiently by a stochastic gradient descent method with mixed unified and alias sampling. The training time for the largest data set of LA is less than 10 min with 4 threads.

4 Experimental Results

In this section, we conduct extensive experiments on the real data sets. Firstly we evaluate the overall performance of our algorithms and the other baselines. Specifically, we compare the performance of new POIs to demonstrate the effectiveness of our algorithms in cold start situation. Then we analyze the parameter sensitivity. At last, we conduct a case study on the real world data sets to show the explanation power of our algorithms.

4.1 Settings

Datasets. We conduct experiments on a foursquare data set collected from 4 cities in USA[1]. We list the statistics of the data set in Table 2. For each city, we only consider users with check-ins at more than 4 POIs as active users. For each user, we first aggregate the check-ins at each POI and sort the POIs in ascending order according to the first check-in timestamp. Then we select the earliest 80% POIs as training data and the remaining 20% as test data. As a result, the numbers of test POIs in four cities are 3686 in LA, 4035 in SF, 2149 in SD, and 1392 in NY.

Table 2. Statistics of our data set

City	#POIs	#Users	#Check-ins
Los Angeles (LA)	7986	1128	49616
San Francisco (SF)	7907	1204	58547
San Diego (SD)	4770	393	25618
New York City (NY)	3631	365	8523

Evaluation Metrics. We use two metrics to evaluate the performance of our model: AUC and Kendall's Tau Coefficient.

AUC. For a target POI l_j and candidate users $U \backslash U_{l_j}$, we consider candidate users that have check-ins at l_j in the test set as positive users and other candidate users as negative users. Then we can plot the ROC curve according to the predicted ranking scores of a model and calculate the Area Under ROC Curve (AUC) as $AUC(l_j)$. We compare the average AUC of all POIs in the test set produced by different models.

Kendall's Tau Coefficient. Kendall's Tau is used to measure the overall ranking accuracy when we consider the number of check-ins in the test data. For a target POI l_j and two candidate users $u_i, u_{i'} \in U \backslash U_{l_j}$, we can get user ranking scores y_{ij}, $y_{i'j}$ and check-in numbers $|C_{ij}|$, $|C_{i'j}|$. Then, the user pair $(u_i, u_{i'})$ is said to be *concordant*, if both $y_{ij} > y_{i'j}$ and $|C_{ij}| > |C_{i'j}|$, or if both $y_{ij} < y_{i'j}$ and $|C_{ij}| < |C_{i'j}|$. On the other hand, if both $y_{ij} > y_{i'j}$ and $|C_{ij}| < |C_{i'j}|$, or if both $y_{ij} < y_{i'j}$ and $|C_{ij}| > |C_{i'j}|$, $(u_i, u_{i'})$ is said to be *discordant*. We define $\#cons$ and $\#disc$ as the number of concordant and discordant user pairs in candidate users $U \backslash U_{l_j}$, and $Tau = \frac{\#cons - \#disc}{\#cons + \#disc}$. We compare the average Tau of all POIs in the test set produced by different models.

[1] The data set is publicly available at https://sites.google.com/site/dbhongzhi/.

Baseline Methods. Our method is denoted as **HRE**, which stands for *hybrid ranking and embedding*. We compare HRE with the following baselines.

- **User Popularity (POP)** ranks each candidate user according to the number of check-ins by that user at all POIs in the training data.
- **Distance-based Mobility Model (DMM)** [28] considers the probability that a candidate user moves from his/her visited POIs to a target POI. It estimates the probability of moving from a visited POI to the target POI by a Pareto distribution from the distances between sequentially visited POIs.
- **Graph Embedding (GE)** [21] is the state-of-the-art model for POI recommendation that learns the embedding of POIs by considering four types of information: POI sequential visit pattern, temporal visit pattern, regional proximity and content similarity. The user embedding in GE is the sum of POI embedding in a user's visited records. We rank each candidate user by the dot product of user embedding and the target POI embedding.
- **Multi-Context Embedding (MC)** [26] is also the state-of-the-art model for POI recommendation that incorporates user-level, trajectory-level, location-level, and temporal contexts for learning embeddings of users and POIs. However, this model cannot learn the embeddings for new POIs.

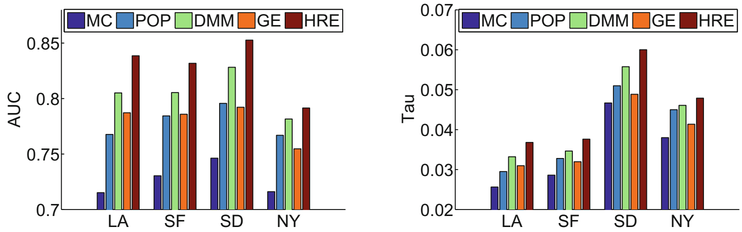
Parameter Settings. The major parameters for our method HRE include: (1) the dimension D for preference embeddings of users and POIs. D is also the dimension of semantic embeddings for time slots, regions, tags and neighborhood users; (2) the number of regions K. By default, we set $D = 50$, $K = 50$. We set the regularization parameter $\lambda = 0.001$ in Eq. (8), the number of nearest neighbor POIs $k = 30$ in constructing POI-neighborhood user graph, and the weight $\beta = 1.0$ in Eq. (13). We set the temporal threshold ΔT for constructing POI-POI graph to be 25 days as [21]. We conduct parameter sensitivity test in Sect. 4.4 to study the influence of parameters on the ranking performance.

4.2 Overall Ranking Performance

We report the performance of different methods for all test POIs in the four cities. The results for AUC and Tau are in Fig. 2. From Fig. 2, we observe that:

- Our method HRE has the best performance in terms of both AUC and Tau among all comparison methods in four cities, because HRE is designed to optimize the ranking consistency and incorporate the heterogeneous information. Compared with the second best method DMM, HRE increases the AUC by *3.0%* and Tau by *7.7%* on average.
- To our surprise, the ranking performance of two existing POI recommendation methods GE and MC are lower than those of DMM and HRE. The reasons are three-folds. *First*, GE and MC are not designed for optimizing ranking consistency. *Second*, GE and MC ignore the geographical proximity between regions, which are captured by geographical embedding and POI-Neighborhood User graph in HRE. *Third*, GE and MC ignore the interaction between geographical and preference influence for modeling user mobility.

– Regarding the results of four cities, we find that LA, SF and SD have much larger AUC than NY. This is because the hometowns of all users in our data set are located in California. These users' check-in behaviors become less regular and predictable when they travel to a new city such as New York City, which was also reported in [19]. The performance gain of HRE in LA and SF is larger than the gain in SD and NY, this is because there are more check-ins in LA and SF which contain more collaborative information between POIs for POI preference embedding enrichments.



(a) AUC for all POIs (b) Tau for all POIs

Fig. 2. Performance on all POIs

4.3 Performance on New POIs

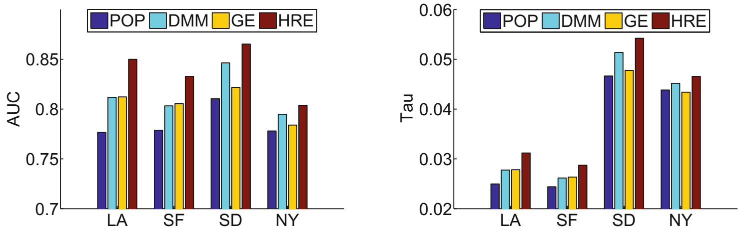
We consider POIs with no check-in users in the training set as new POIs in our evaluation. Since MC cannot be applied to new POIs, we only report the performance of the four methods for new POIs in Fig. 3. Again we can observe that HRE has the largest AUC and Tau among all methods, which confirms that incorporating the ranking information in the user-POI interaction can enrich the representation of POIs without check-in information.

4.4 Parameter Sensitivity Test

We evaluate the performance of HRE w.r.t. four parameters: the dimension D of preference embedding, the dimension K of geographical embedding, the number of k-nearest POIs, k, in selecting neighborhood users, and the coefficient β for tuning weight of graph-based POI embedding. The AUC values for all POIs and new POIs in all four cities are reported in Fig. 4. We observe that the ranking performance of our method is not very sensitive to the parameter change. This shows the robustness of our method HRE.

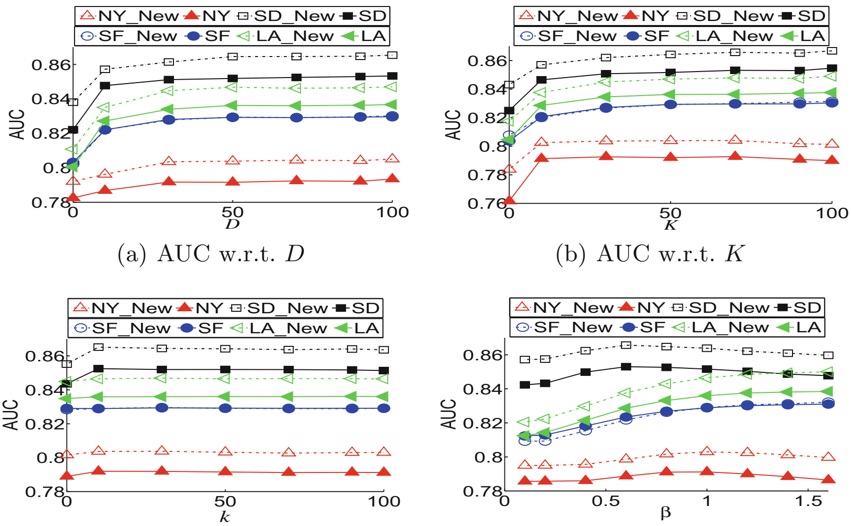
4.5 A Case Study

We report the AUC for each category of POIs in the test set. The category information is obtained from tags of POIs in the data set, where each POI may have



(a) AUC for new POIs (b) Tau for new POIs

Fig. 3. Performance on new POIs



(a) AUC w.r.t. D (b) AUC w.r.t. K

(c) AUC w.r.r. k (d) AUC w.r.t. β

Fig. 4. AUC w.r.t. different parameters in HRE

more than one tag because it may have multiple functions. We calculate the average AUC of one tag by aggregating the POIs in four cities. Then we sort the tags based on their AUC produced by our method in descending order. We plot a heat map recording the AUC produced by five methods in Fig. 5. We observed that the AUC rankings w.r.t. category produced by HRE are more consistent with the hierarchy of needs for human life. The top three categories with the largest AUC, which are "Food & Drink Shop", "Shop & Service", "Residence", stand for more fundamental needs. On the other hand, categories such as "Arts & Entertainment", "Office" and "Airport" stand for higher-level human needs related to spirit and self-actualization [5]. In contrast, ranking categories based on other methods' AUC fail to reveal the needs order of human due to the lack of user-POI interaction information in learning POIs' preference embeddings. For example, the AUC of "Clothing" in GE and DMM are close to "Art & entertainment". In fact, "Clothing" is a more fundamental need than "Arts & Entertainment".

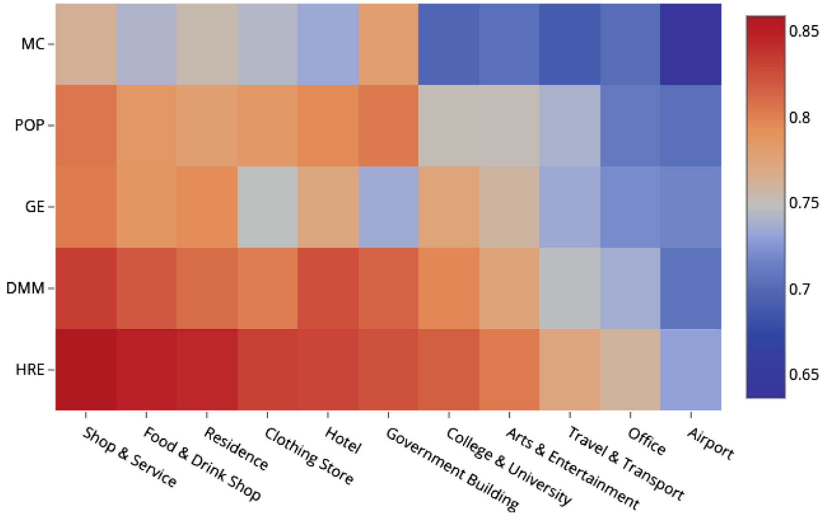


Fig. 5. AUC w.r.t. category

5 Related Works

Influential Users Selection in Location Based Social Network. Unlike traditional influence maximization problems in online social networks [1,7,17], users' locations need to be considered in finding influential users in location-based social networks [3,24,27]. [8] identifies a set of users who are most influential to other users in a specific region. [20] considers the user-POI distance in calculating the influence spread. Both works assume that the user's location is fixed and ignore the user mobility. [28] first studies the location promotion problem where the activation probability of each edge is determined by users' probability to visit a target location. They propose two mobility models for setting the users' visiting probability where the distance-based mobility model (DMM) is reported to have the largest AUC under different activation thresholds. As mentioned in Sect. 1, DMM only considers check-in records of one user in deriving his/her visiting probability, while our method considers heterogeneous information such as POI tags and neighborhood user visiting patterns in check-in records of all users. Besides, our method focuses on maximizing the ranking consistency on setting user visiting score so that it has the highest ranking performance as shown in the experiments.

User Mobility Modeling and POI Recommendation. Studies [4,12,25] related to user mobility modeling focus on maximizing the likelihood of observing all check-in records of individual users. [4] finds that most of the human movement is based on periodic behaviors. They propose a Periodic Mobility Model (PMM) that consists of two Gaussian distributions to denote user's states at work and home respectively. In POI recommendation [2,6,9–11,13,14,19,21,22], the search engine is asked to return a set of POIs that the query user may be interested in. Most of the works on POI recommendation [9,11,13,22] are based on a fused model that considers geographical influence as well as collaborative filtering information from check-in records. [10] first uses a ranking based loss function to optimize the precision and recall of top-k recommendation, but their method cannot deal with new POIs because they only use information of user-POI matrix. [21] further considers the context information in POI recommendation and proposes a multi-graphs embedding method for integrating heterogeneous information. It outperforms other cold start POI recommendation methods based on content information and geographical locations [19,23]. Our work is different from user mobility modeling and POI recommendation as our target is to recommend users for a target POI. We also confirm the effectiveness of our ranking based methods by comparing with the methods for POI recommendation [21,26].

6 Conclusion

In this paper, we study the location promotion problem in location-based services. In order to return a target user list w.r.t. a target POI, we propose a unified representation learning framework called hybrid ranking and embedding. Our framework maximizes the ranking consistency and integrates heterogeneous information of POIs, which alleviates the data sparsity problem in users' check-in data and solves the cold-start POI problem. Experiments on four cities of the United States show that our method has better ranking performance than the state-of-the-art methods.

Acknowledgement. The work is supported by a Microsoft Research Asia Collaborative Research Grant.

References

1. Arora, A., Galhotra, S., Ranu, S.: Debunking the myths of influence maximization: an in-depth benchmarking study. In: SIGMOD, pp. 651–666 (2017)
2. Bao, J., Zheng, Y., Wilkie, D., Mokbel, M.: Recommendations in location-based social networks: a survey. GeoInformatica **19**(3), 525–565 (2015)
3. Bouros, P., Sacharidis, D., Bikakis, N.: Regionally influential users in location-aware social networks. In: SIGSPATIAL, pp. 501–504 (2014)
4. Cho, E., Myers, S.A., Leskovec, J.: Friendship and mobility: user movement in location-based social networks. In: SIGKDD, pp. 1082–1090 (2011)
5. Fu, Y., Xiong, H., Ge, Y., Yao, Z., Zheng, Y., Zhou, Z.H.: Exploiting geographic dependencies for real estate appraisal: a mutual perspective of ranking and clustering. In: SIGKDD, pp. 1047–1056 (2014)
6. Kefalas, P., Symeonidis, P., Manolopoulos, Y.: Recommendations based on a heterogeneous spatio-temporal social network. WWW **21**(2), 345–371 (2018)
7. Kempe, D., Kleinberg, J., Tardos, É.: Maximizing the spread of influence through a social network. In: SIGKDD, pp. 137–146 (2003)

8. Li, G., Chen, S., Feng, J., Tan, K.L., Li, W.S.: Efficient location-aware influence maximization. In: SIGMOD, pp. 87–98 (2014)
9. Li, H., Ge, Y., Hong, R., Zhu, H.: Point-of-interest recommendations: learning potential check-ins from friends. In: SIGKDD, pp. 975–984 (2016)
10. Li, X., Cong, G., Li, X.L., Pham, T.A.N., Krishnaswamy, S.: Rank-GeoFM: a ranking based geographical factorization method for point of interest recommendation. In: SIGIR, pp. 433–442 (2015)
11. Lian, D., Zhao, C., Xie, X., Sun, G., Chen, E., Rui, Y.: GeoMF: joint geographical modeling and matrix factorization for point-of-interest recommendation. In: SIGKDD, pp. 831–840 (2014)
12. Lichman, M., Smyth, P.: Modeling human location data with mixtures of kernel densities. In: SIGKDD, pp. 35–44 (2014)
13. Liu, B., Fu, Y., Yao, Z., Xiong, H.: Learning geographical preferences for point-of-interest recommendation. In: SIGKDD, pp. 1043–1051 (2013)
14. Liu, Y., Pham, T.A.N., Cong, G., Yuan, Q.: An experimental evaluation of point-of-interest recommendation in location-based social networks. VLDB **10**(10), 1010–1021 (2017)
15. Pham, H., Shahabi, C., Liu, Y.: EBM: an entropy-based model to infer social strength from spatiotemporal data. In: SIGMOD, pp. 265–276 (2013)
16. Rendle, S., Freudenthaler, C., Gantner, Z., Schmidt-Thieme, L.: BPR: bayesian personalized ranking from implicit feedback. In: UAI, pp. 452–461 (2009)
17. Tang, Y., Shi, Y., Xiao, X.: Influence maximization in near-linear time: a martingale approach. In: SIGMOD, pp. 1539–1554 (2015)
18. Wang, H., Li, Z., Lee, W.C.: PGT: measuring mobility relationship using personal, global and temporal factors. In: ICDM, pp. 570–579 (2014)
19. Wang, W., Yin, H., Chen, L., Sun, Y., Sadiq, S., Zhou, X.: Geo-SAGE: a geographical sparse additive generative model for spatial item recommendation. In: SIGKDD, pp. 1255–1264 (2015)
20. Wang, X., Zhang, Y., Zhang, W., Lin, X.: Distance-aware influence maximization in geo-social network. In: ICDE, pp. 1–12 (2016)
21. Xie, M., Yin, H., Wang, H., Xu, F., Chen, W., Wang, S.: Learning graph-based poi embedding for location-based recommendation. In: CIKM, pp. 15–24 (2016)
22. Ye, M., Yin, P., Lee, W.C., Lee, D.L.: Exploiting geographical influence for collaborative point-of-interest recommendation. In: SIGIR, pp. 325–334 (2011)
23. Yin, H., Zhou, X., Shao, Y., Wang, H., Sadiq, S.: Joint modeling of user check-in behaviors for point-of-interest recommendation. In: CIKM, pp. 1631–1640 (2015)
24. Zhang, C., Shou, L., Chen, K., Chen, G., Bei, Y.: Evaluating geo-social influence in location-based social networks. In: CIKM, pp. 1442–1451 (2012)
25. Zhang, C., Zhang, K., Yuan, Q., Zhang, L., Hanratty, T., Han, J.: GMove: group-level mobility modeling using geo-tagged social media. In: SIGKDD, pp. 1305–1314 (2016)
26. Zhou, N., Zhao, W.X., Zhang, X., Wen, J.R., Wang, S.: A general multi-context embedding model for mining human trajectory data. TKDE **28**(8), 1945–1958 (2016)
27. Zhou, T., Cao, J., Liu, B., Xu, S., Zhu, Z., Luo, J.: Location-based influence maximization in social networks. In: CIKM, pp. 1211–1220 (2015)
28. Zhu, W.Y., Peng, W.C., Chen, L.J., Zheng, K., Zhou, X.: Modeling user mobility for location promotion in location-based social networks. In: SIGKDD, pp. 1573–1582 (2015)