Learning User's Intrinsic and Extrinsic Interests for Point-of-Interest Recommendation: A Unified Approach

Huayu Li[1], Yong Ge[2], Defu Lian[3*], Hao Liu[1]
[1] University of North Carolina at Charlotte
[2] Nanjing University of Finance and Economic
[3] Big Data Research Center, University of Electronic Science and Technology of China
hli38@uncc.edu, ygestrive@gmail.com, dove.ustc@gmail.com, hliu32@uncc.edu

Abstract

Point-of-Interest (POI) recommendation has been an important service on location-based social networks. However, it is very challenging to generate accurate recommendations due to the complex nature of user's interest in POI and the data sparseness. In this paper, we propose a novel unified approach that could effectively learn fine-grained and interpretable user's interest, and adaptively model the missing data. Specifically, a user's general interest in POI is modeled as a mixture of her intrinsic and extrinsic interests, upon which we formulate the ranking constraints in our unified recommendation approach. Furthermore, a self-adaptive location-oriented method is proposed to capture the inherent property of missing data, which is formulated as squared error based loss in our unified optimization objective. Extensive experiments on real-world datasets demonstrate the effectiveness and advantage of our approach.

1 Introduction

Recent years have witnessed the rapid prevalence of location-based social network (LBSN) services, such as Foursquare, Yelp, and Facebook, which can significantly facilitate users' outdoor activities by providing a large number of nearby Point-of-Interests (POIs) in a real-time fashion. The availability of large-scale user interaction data with these LBSN services, such as sharing check-in information, provides unparalleled opportunities for developing personalized POI recommender systems [Li *et al.*, 2015; 2016b].

However, the complex nature of user interest and the sparsity of check-in data present significant challenges to develop POI recommender systems. First, only with check-in records, it is difficult to explain which reason impels a user to check-in a location. Thus, modeling user's true and interpretable interest becomes a thorny issue. For example, when an unvisited POI is far away from a user, she may not visit it due to the external geographical restriction, even though she likes the POI. This imposes a challenge on interpreting and modeling a user's decision making on POI check-in. Second, location recommender systems usually suffer from another critical challenge caused by the extremely sparse data. In a real system, there are over millions of locations and users. However, each user only has limited historical check-ins, significantly increasing the difficulty of recommendation.

Recently, a variety of approaches have been proposed for POI recommendation with matrix factorization from squared error based loss. For example, [Hu *et al.*, 2008; Pan *et al.*, 2008; Devooght *et al.*, 2015] treat user's preferences for observed and unobserved locations as binary values with varying weights. [Liu *et al.*, 2014] models geo-neighboring influence in both instance and region levels, where a user's choice to a location is affected by its neighboring locations, and locations in a region share a similar sparsity structure. [Li *et al.*, 2016a] first utilizes additional knowledge, such as social network, to learn a set of potential locations from a user's all unobserved locations, and then assigns a small fixed value to fit this user's preference for these potential locations. However, all these existing approaches have two limitations. First, they model user's preference in a too general way, and thus fail to capture user's true interest, not to mention interpreting user's check-in decision making process. For example, a low predicted rating of a user for an unvisited POI can not reveal the reason why this user does not like the location. Is it because she does not like or due to the external environment restriction? Second, most existing methods treat user's all unobserved feedbacks as negative in the same way, and thus cannot capture the inherent property of missing data, i.e., a mixture of missed negative and positive values.

To address the aforementioned issues, in this paper, we propose a unified approach that could effectively learn fine-grained and interpretable user's interest, and adaptively model missing data. Each user's general interest is modeled as a mixture of her intrinsic and extrinsic interests, where the former one is personal-taste driven and characterizes her own satisfaction regardless of any restriction, and the latter one is environment driven and influenced by external environment, i.e., geographical distance. To capture and distinguish both of them, we first formally define a user's activity area as a set of locations geographically accessible by this user, and then formulate them into pairwise ranking constraints in our unified recommendation approach. Specifically, upon intrinsic interest, one user prefers each visited POI over any unvisited one within the corresponding activity area. On the other hand,

*The corresponding author.

upon extrinsic interest, she prefers each visited POI over any unvisited one outside activity area. Moreover, a self-adaptive location-oriented method is proposed to capture the characteristic of missing data, which is formulated as squared error based loss in our unified optimization objective. Finally, the proposed model is evaluated via different validation metrics and compared with several state-of-the-art baseline models on real-world datasets. The experimental results illustrate the superiority of our model for POI recommendation. The major contributions of this paper can be summarized as follows.

- We propose to learn user interest in a precise and interpretable way, which is a mixture of the intrinsic and extrinsic interests. Based on these two types of interests, we formulate pairwise ranking constraints.
- We propose a location-oriented method to adaptively model the missing data, upon which we formulate the squared error based loss.
- We conduct extensive evaluations with real-world datasets to demonstrate the effectiveness of our model.

2 The Proposed Method

Problem Statement. The recommendation task is defined as: *given the check-in behaviors of n users over m locations, we aim at recommending each user with top-K new locations that she might be interested in but has never visited before.*

Notations. Scalars, vectors and matrices are denoted by lower case letters, bold face lower case letters and bold face capital letters, respectively. Sets are represented by calligraphic capital letters. $\mathbf{u}_i$ denotes the i-th column of matrix $\mathbf{U}$. Frobenius and Euclidean norms are denoted by $||\cdot||_F$ and $||\cdot||$, respectively. $\mathbb{R}_n$ is defined as the set $\{1, \cdots, n\}$. A predicted value is denoted with a ˆ(hat) over it. $\mathbf{C}$ is used to denote the n-by-m check-in frequency matrix.

2.1 Our Framework

Matrix factorization techniques have been popularly used to solve recommendation tasks by mapping both user and location into latent low-dimensional spaces [Salakhutdinov and Mnih, 2007; Hu *et al.*, 2008; He *et al.*, 2016]. Specifically, each user-specific hidden vector is used to model this user's interest, and is then learned via an appropriate loss function. However, unlike traditional product consumption, user's check-in behaviors are constrained by many external factors, such as geographical distance. The complex nature in human mobility leads to the incapability of previous approaches for modeling a user's true interest in POI and interpreting her decision making process. Inspired by the studies in psychology and sociology about distinguishing whether or not a user's behavior is affected by external factors [Ryan and Deci, 2000; Calder and Staw, 1975], we propose to model each user's general interest from two aspects: (1) intrinsic interest, where she visits a location for the sake of her own inherent likeness regardless of any restriction, (2) extrinsic interest, where her check-in decision making process is influenced by geographical distance. Formally, they are are defined as:

Definition 1 (Intrinsic Interest) *It is an internal form of interest and driven by personal taste. For example, it is the self-desire for a user to visit a POI with her own satisfaction.*

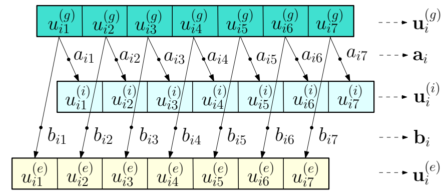


Figure 1: An example of the illustration for $\mathbf{u}_i^{(g)}$, $\mathbf{u}_i^{(i)}$, and $\mathbf{u}_i^{(e)}$, where $\mathbf{u}_i^{(g)}$ is a mixture of $\mathbf{u}_i^{(i)}$ and $\mathbf{u}_i^{(e)}$ with corresponding mixture weights $\mathbf{a}_i$ and $\mathbf{b}_i$.

Definition 2 (Extrinsic Interest) *It is an external form of interest and driven by environment. For example, a user's preference for a POI is influenced by geographical distance.*

These two types of user interest have different contributions to each individual user's check-in decision. Hence, each user's general interest is regarded as a mixture of her intrinsic and extrinsic interests. Suppose the general, intrinsic and extrinsic interests of user i are represented by d-dimensional vectors $\mathbf{u}_i^{(g)}$, $\mathbf{u}_i^{(i)}$, and $\mathbf{u}_i^{(e)}$, respectively. Their relationship can be formulated as follows,

$$\mathbf{u}_i^{(g)} = \mathbf{a}_i \odot \mathbf{u}_i^{(i)} + \mathbf{b}_i \odot \mathbf{u}_i^{(e)}, \quad (1)$$
$$s.t.\ a_{ik} \in [0,1], b_{ik} \in [0,1], a_{ik} + b_{ik} = 1, \forall k \in \mathbb{R}_d,$$

where $\odot$ denotes the element-wise multiplication. And $\mathbf{a}_i \in \mathbb{R}^{d\times 1}$ and $\mathbf{b}_i \in \mathbb{R}^{d\times 1}$ are the mixture weights of intrinsic and extrinsic interest, respectively. Figure 1 provides an illustration example for $\mathbf{u}_i^{(g)}$, $\mathbf{u}_i^{(i)}$, and $\mathbf{u}_i^{(e)}$.

We also propose to capture the characteristic of missing data (or unobserved data), i.e., a mixture of negative and missed positive values, for addressing the data sparseness issue in location recommendations. Let us assume each location j is also characterized by a latent vector $\mathbf{v}_j \in \mathbb{R}^{d\times 1}$ as in the matrix factorization technique [Salakhutdinov and Mnih, 2007]. Consequently, towards modeling user's intrinsic and extrinsic interest and the missing data, the overall loss function of our framework is formulated as:

$$\min \ell(\mathbf{P}, \mathbf{U}^{(g)}, \mathbf{V}) + \theta^c(\mathbf{U}^{(e)}, \mathbf{U}^{(i)}, \mathbf{V}) + \theta^r(\cdot), \quad (2)$$

where $\theta^r(\cdot)$ is the regularization term placed on all variables to avoid over-fitting. $\theta^c(\cdot)$ incorporates additional constraints that model user's intrinsic interest $\mathbf{U}^{(i)}$ and extrinsic interest $\mathbf{U}^{(e)}$, and will be introduced in Section 2.2. $\ell(\cdot)$ is the empirical loss used to model both observed and unobserved data, and introduced in Section 2.3 ($\mathbf{P}$ is aslo introduced).

2.2 Modeling User Intrinsic and Extrinsic Interests

Based on the definition 1 and 2, the distinction between user's intrinsic and extrinsic interests sheds light on whether there exists the involvement of external influence. In POI recommendation task, a user's check-in decision making process is significantly affected by geographical factor. Thus, in this paper, we focus on modeling user's interests with geographical influence. Before introducing how to model user's two types of interests, we first define user's activity area as follows:

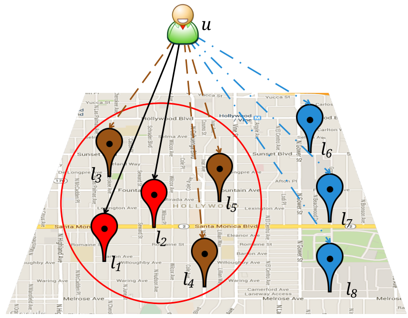


Figure 2: An example of a user's observed and unobserved locations. The solid lines indicate observed behaviors, and other dashed lines represent unobserved ones. The red circle represent a activity area of this user, where locations $l_1 \sim l_5$ is assumed to be geographically accessible by this user.

Definition 3 (User's Activity Area) *Each individual user has one or more activity areas, within each of which this user is capable of geographically accessing each POI regardless of any geo-restriction.*

Each user's activity areas can be calculated through many ways according to real-world application scenarios. One method is using clustering technique based on user's historical locations. In each cluster, we first select all locations within a circle with a specified distance as radius and each visited location as a center, and then merge them as the activity area. Formally, we define n_i^a as the number of activity areas for the i-th user, where each of them $h \in \mathbb{R}_{n_i^a}$ comprises a set of her visited locations $\mathcal{A}_{ih}$. Based on the definition of activity area, we will introduce how to model user's intrinsic and extrinsic interests in followings of this section.

Modeling User Intrinsic Interest $\mathbf{U}^{(i)}$. Based on the definition 1, each user's intrinsic interest suggests her to choose any locations that she likes regardless of any external geo-restriction. Within each activity area, the user is able to geographically access each location. It indicates that a user's check-in decision making process on those locations within her activity areas is restriction-free. Thus, within each activity area, compared to those unobserved locations, one user's intrinsic interest in an observed location plays an important role in her decision making process. In other words, upon intrinsic interest, each individual user i prefers each observed location j over any unobserved location l in each activity area, which can be formulated as follows:

$$(\mathbf{u}_i^{(i)})^T\mathbf{v}_j > (\mathbf{u}_i^{(i)})^T\mathbf{v}_l, \ \ \forall j \in \mathcal{A}_{ih}, l \in \bar{\mathcal{A}}_{ih}, h \in \mathbb{R}_{n_i^a}, \quad (3)$$

where $\bar{\mathcal{A}}_{ih}$ is a set of unvisited locations of user i in her h-th activity area, and $(\mathbf{u}_i^{(i)})^T\mathbf{v}_j$ indicates her preference for location j driven by intrinsic interest. Figure 2 shows an example to illustrate Eq.(3), where user u would like to check-in $l_1 \sim l_2$ rather than $l_3 \sim l_5$ upon her intrinsic interest.

Modeling User Extrinsic Interest $\mathbf{U}^{(e)}$. For those locations outside a user's activity areas, she has a small chance to visit them due to long distance. For example, although a user living at California likes a restaurant in New York, she would not go to check-in this restaurant for the sake of distance restriction. Thus, compared to those unobserved locations outside a user's activity areas, her extrinsic interest in an observed location has more impact on her check-in decision than her intrinsic one. Thus, upon extrinsic interest, each user i prefers each observed location j over any unobserved one outside the activity areas, which is formulated as follows:

$$(\mathbf{u}_i^{(e)})^T\mathbf{v}_j > (\mathbf{u}_i^{(e)})^T\mathbf{v}_l, \ \ \forall j \in \mathcal{A}_{ih}, l \in \mathcal{A}_i^*, h \in \mathbb{R}_{n_i^a}, \quad (4)$$

where $\mathcal{A}_i^*$ is a set of unvisited locations of user i outside activity areas, and $(\mathbf{u}_i^{(e)})^T\mathbf{v}_j$ indicates her preference for location j driven by extrinsic interest. Figure 2 shows an example to illustrate Eq.(4), where user u would like to check-in $l_1 \sim l_2$ rather than $l_6 \sim l_8$ upon her extrinsic interest.

Our Method V.S. Existing Approaches. We would clarify that the existing approaches can be viewed as a special case of our method. If the size of activity area is infinite, which means the activity area of each user includes the whole location set, the constraints in Eq.(4) then will be eliminated, i.e., $\mathbf{U}^{(e)} = \mathbf{0}$. If the size of activity area is zero, indicating that each location is a single activity area, the constraints in Eq.(3) then will be eliminated, i.e., $\mathbf{U}^{(i)} = \mathbf{0}$. In both situations, only either intrinsic or extrinsic interest of user i contributes to her general interest, i.e., $\mathbf{U}^{(g)} = \mathbf{U}^{(i)}$ or $\mathbf{U}^{(g)} = \mathbf{U}^{(e)}$, which is the same as existing approaches without distinguishing user's two types of interests. Thus, introducing the activity area can provide us finer-grained granularity to explore better accuracy of POI recommender systems.

The advantage of modeling user's general interest from two aspects is as follows: (1) It makes location recommendation systems behave in an explainable way by interpreting user's choice for locations from both internal and external perspective. (2) It provides a fine-grained and accurate way to learn user's interest. Finally, we can obtain the ranking constraint $\theta^c(\cdot)$ of Eq.(2) by penalizing those violated constraints shown in Eq.(3) and Eq.(4) as follows:

$$\theta^c(\cdot) = \sum_{i,h,j \in \mathcal{T}} \left(\lambda_c^i \sum_{l \in \bar{\mathcal{A}}_{ih}} (\hat{r}_{il}^{(i)} - \hat{r}_{ij}^{(i)})_+ + \lambda_c^e \sum_{l \in \mathcal{A}_i^*} (\hat{r}_{il}^{(e)} - \hat{r}_{ij}^{(e)})_+ \right) \quad (5)$$

where $\hat{r}_{ij}^{(i)} = (\mathbf{u}_i^{(i)})^T\mathbf{v}_j$, $\hat{r}_{ij}^{(e)} = (\mathbf{u}_i^{(e)})^T\mathbf{v}_j$, $\mathcal{T} = \{i, h, j | i \in \mathbb{R}_n, h \in \mathbb{R}_{n_i^a}, j \in \mathcal{A}_{ih}\}$, $(x)_+ = \max(x, 0)$ is plus function, and λ_c^i, λ_c^e are parameters weighting two ranking constraints.

2.3 Modeling the Missing Data

Due to the large set of locations, it is crucial to model the missing data as well for addressing data sparseness and improving learning accuracy. Most recent approaches [Hu *et al.*, 2008; Liu *et al.*, 2014] focus on squared error based loss and treat all unobserved feedbacks as negative in the same way. However, it is not realistic in real-world scenario. An unvisited location of one user does not necessarily indicate that she dislikes it, whereas it happens possibly due to her unawareness. In other words, some of the unvisited locations might be those users are interested in, while others are actually those they dislike. Thus, each user's preferences for unobserved locations are a mixture of negative and missed positive values. Motivated by this intuition, we propose a location-oriented method to adaptively learn the potential values for missing entries, instead of treating them equally as a predefined value. To achieve this, we introduce an augmented matrix $\mathbf{P} \in \mathbb{R}^{n \times m}$ that is designed only for unobserved feedbacks and learned during the training. Suppose the predicted preference of user i for location j is approximated by $\hat{r}_{ij} = (\mathbf{u}_i^{(g)})^T \mathbf{v}_j$ [1]. The empirical loss $\ell(\cdot)$ of Eq.(2) with squared error is formulated as:

$$\ell(\cdot) = \frac{1}{2} ||\mathbf{W} \odot (\mathbf{R} + \mathbf{P} - \hat{\mathbf{R}})||_F^2, \tag{6}$$

where $\mathbf{W} \in \mathbb{R}^{n \times m}$ is a weight matrix associated with check-in frequency and a parameter $\gamma \in (0, +\infty)$, and defined as $w_{ij} = \sqrt{1 + \gamma * c_{ij}}$. $\mathbf{R} \in \mathbb{R}^{n \times m}$ is a one-class feedback matrix with observed feedback as one and unobserved feedback as zero, i.e., $r_{ij} = 1$ if $c_{ij} \neq 0$ and $r_{ij} = 0$ if $c_{ij} = 0$. Specifically, $\mathbf{P}$ has two properties: (1) It is location-oriented, i.e., comprising m latent factors, for efficient computing; (2) It has a small variation range for decreasing the noise in model learning. Thus, $\mathbf{P}$ can be formulated by introducing a m-dimensional vector $\mathbf{q}$ as follows,

$$p_{ij} = \begin{cases} 0 & \text{if } r_{ij} = 1, \\ q_j & \text{if } r_{ij} = 0, \end{cases} \tag{7}$$
$$s.t. \quad \mathbf{q} \in [q_{min}, q_{max}],$$

where q_{min} and q_{max} are parameters used to bound $\mathbf{P}$ into a small range around zero. In Eq.(6), it is clear that $\mathbf{R}$ is used to model observed feedbacks, while $\mathbf{P}$ accounts for unobserved feedbacks and explains the difference of missing data.

Discussion. If $q_{min} = q_{max} = 0$, it leads to the basic implicit-feedback based approaches by fitting a binary matrix [Hu *et al.*, 2008; Pan *et al.*, 2008]. If $q_{min} = q_{max} \neq 0$, it leads to the constant imputation-based approaches by assigning a predefined non-zero numeric value for unobserved feedback [Yao *et al.*, 2014; Li *et al.*, 2016a]. If $q_{min} \neq q_{max}$, our method is distinct from existing approaches by adaptively separating all the unobserved feedbacks, which has two advantages: (1) capturing the inherent property exhibited in unobserved data, i.e., a mixture of negative and positive values, and (2) automatically learning the optimal values for unobserved data instead of specifying a fixed value.

2.4 Optimization Algorithm

So far we have introduced our solutions to capture user's intrinsic and extrinsic interests, and address missing data issue in POI recommendation. With these solutions, the loss function of the proposed model, denoted as *IEMF*, is achieved by integrating Eq.(5) and Eq.(6) into the framework in Eq.(2):

$$\underset{\mathbf{U}^{(i)},\mathbf{U}^{(e)},\mathbf{V},\mathbf{A},\mathbf{B},\mathbf{q}}{\text{argmin}} \frac{1}{2} ||\mathbf{W} \odot (\mathbf{R} + \mathbf{P} - \hat{\mathbf{R}})||_F^2 + \theta^r(\cdot) + \tag{8}$$
$$\sum_{i,h,j \in \mathcal{T}} w_{ij} \left(\lambda_c^i \sum_{l \in \bar{\mathcal{A}}_{ih}} (\hat{r}_{il}^{(i)} - \hat{r}_{ij}^{(i)})_+ + \lambda_c^e \sum_{l \in \mathcal{A}_i^*} (\hat{r}_{il}^{(e)} - \hat{r}_{ij}^{(e)})_+ \right),$$
$$s.t. \ \mathbf{A} \in [0,1], \mathbf{B} \in [0,1], \mathbf{A} + \mathbf{B} = \mathbf{1}_{n \times d}, \mathbf{q} \in [q_{min}, q_{max}],$$

Algorithm 1: IEMF Optimization

Input: $\lambda_c^i, \lambda_c^e, \lambda_u^i, \lambda_u^e, \lambda_v, \lambda_z, \lambda_p, \alpha, \gamma, q_{min}, q_{max}, \eta, maxIter$
Output: $\mathbf{U}^{(i)}, \mathbf{U}^{(e)}, \mathbf{V}, \mathbf{Z}, \mathbf{q}$
Randomly initialize $\mathbf{U}^{(i)}, \mathbf{U}^{(e)}, \mathbf{V}, \mathbf{Z}, \mathbf{q}, t \leftarrow 1$
while $t \leqslant maxIter$ **do**
  Randomly sample a tuple $(i, h, j, k, l, \mathcal{S}_-)$, where i is a user, h is activity area index, j is a visited location within the acticity area, k is an unvisited location within this activity area, l is an unvisited location outside activity areas, and $\mathcal{S}_-$ is a set of other unvisited locations with size as $|\mathcal{S}_-|$.
  $e_{io} \leftarrow w_{io}(\hat{r}_{io} - \widetilde{r}_{io}), \forall o \in \mathcal{S}$
  $e_{jk} \leftarrow \lambda_c^i w_{ij} y'(\hat{r}_{ik}^{(i)} - \hat{r}_{ij}^{(i)}), e_{jl} \leftarrow \lambda_c^e w_{ij} y'(\hat{r}_{il}^{(e)} - \hat{r}_{ij}^{(e)})$
  $\mathbf{u}_i^{(i)} \leftarrow \mathbf{u}_i^{(i)} - \eta(\sum_{o \in \mathcal{S}} e_{io}(\mathbf{a}_i \odot \mathbf{v}_o) + e_{jk}(\mathbf{v}_k - \mathbf{v}_j) + \lambda_u^i \mathbf{u}_i^{(i)})$
  $\mathbf{u}_i^{(e)} \leftarrow \mathbf{u}_i^{(e)} - \eta(\sum_{o \in \mathcal{S}} e_{io}(\mathbf{b}_i \odot \mathbf{v}_o) + e_{jl}(\mathbf{v}_l - \mathbf{v}_j) + \lambda_u^e \mathbf{u}_i^{(e)})$
  $\mathbf{z}_i \leftarrow \mathbf{z}_i - \eta(\sum_{o \in \mathcal{S}} e_{io}(\mathbf{a}_i' \odot \mathbf{u}_i^{(i)} + \mathbf{b}_i' \odot \mathbf{u}_i^{(e)}) \odot \mathbf{v}_o + \lambda_z \mathbf{z}_i)$
  $\mathbf{v}_j \leftarrow \mathbf{v}_j - \eta(e_{ij}\mathbf{u}_i^{(g)} - e_{jk}\mathbf{u}_i^{(i)} - e_{jl}\mathbf{u}_i^{(e)} + \lambda_v \mathbf{v}_j)$
  $\mathbf{v}_k \leftarrow \mathbf{v}_k - \eta(e_{ik}\mathbf{u}_i^{(g)} + e_{jk}\mathbf{u}_i^{(i)} + \lambda_v \mathbf{v}_k)$
  $\mathbf{v}_l \leftarrow \mathbf{v}_l - \eta(e_{il}\mathbf{u}_i^{(g)} + e_{jl}\mathbf{u}_i^{(e)} + \lambda_v \mathbf{v}_l)$
  $\mathbf{v}_o \leftarrow \mathbf{v}_o - \eta(e_{io}\mathbf{u}_i^{(g)} + \lambda_v \mathbf{v}_o), \forall o \in \mathcal{S}_-$
  $q_o \leftarrow (\mathbf{v}_o - \eta(-e_{io} + \lambda_q q_j))_q, \forall o \in \mathcal{S}_-^*$
  $t \leftarrow t + 1$
end
return $\mathbf{U}^{(i)}, \mathbf{U}^{(e)}, \mathbf{V}, \mathbf{Z}, \mathbf{q}$

where $\mathbf{1}_{n \times d}$ is a n-by-d matrix of ones, and the weight matrix $\mathbf{W}$ is also used to balance the optimization of squared error and ranking error. To solve above optimization problem, we need some preprocessing and approximation steps. First, we eliminate the constraints imposed on $\mathbf{A}$ and $\mathbf{B}$ by introducing a helper matrix $\mathbf{Z} \in \mathbb{R}^{n \times d}$ with $z_{ik} \in (-\infty, +\infty)$, and using a sigmoid function to bound the value to $[0, 1]$ as follows,

$$a_{ik} = \sigma(z_{ij}), \quad b_{ik} = 1 - \sigma(z_{ij}), \tag{9}$$

where $\sigma(x) = \frac{1}{1+\exp(-x)}$ is sigmoid function. Second, the plus function used in Eq.(5) is not twice differentiable and can be smoothly approximated by the integral to a smooth approximation of the sigmoid function [Lee and Mangasarian, 2001; Chen and Mangasarian, 1993], given by:

$$(x)_+ \approx y(x) = x + \frac{1}{\alpha} log(1 + \exp(-\alpha x)), \tag{10}$$

where $\alpha \in (0, +\infty)$ is a parameter. With these two steps, the optimization problem shown in Eq.(8) can be transformed to,

$$\underset{\mathbf{U}^{(i)},\mathbf{U}^{(e)},\mathbf{V},\mathbf{Z},\mathbf{q}}{\text{argmin}} \frac{1}{2} ||\mathbf{W} \odot (\widetilde{\mathbf{R}} - \hat{\mathbf{R}})||_F^2 + \theta^r(\mathbf{U}^{(i)}, \mathbf{U}^{(e)}, \mathbf{V}, \mathbf{Z}, \mathbf{q}) +$$
$$\sum_{i,h,j \in \mathcal{T}} w_{ij} \left(\lambda_c^i \sum_{l \in \bar{\mathcal{A}}_{ih}} y(\hat{r}_{il}^{(i)} - \hat{r}_{ij}^{(i)}) + \lambda_c^e \sum_{l \in \mathcal{A}_i^*} y(\hat{r}_{il}^{(e)} - \hat{r}_{ij}^{(e)}) \right),$$

where $\widetilde{\mathbf{R}} = \mathbf{R} + \mathbf{P}$, $\mathbf{q} \in [q_{min}, q_{max}]$, and $\theta^r(\cdot)$ is defined with regularization parameters λ_* as follows:

$$\theta^r(\cdot) = \frac{\lambda_u^i}{2} ||\mathbf{U}^{(i)}||_F^2 + \frac{\lambda_u^e}{2} ||\mathbf{U}^{(e)}||_F^2 + \frac{\lambda_v}{2} ||\mathbf{V}||_F^2 + \frac{\lambda_z}{2} ||\mathbf{Z}||_F^2 + \frac{\lambda_q}{2} ||\mathbf{q}||^2.$$

As there is no close-form for each variable with ALS approach, a Stochastic Gradient Descent (SGD) using the boostrap sampling with replacement algorithm is developed to solve the optimization problem. The optimization algorithm is iteratively performed by sampling a tuple $(i, h, j, k, l, \mathcal{S}_-)$ and updating corresponding variables. More details of optimization are provided in Algorithm 1, where $\mathcal{S} = \{j, k, l\} \cup$

[1]We will also incorporate geographical influence into final prediction with $\hat{r}_{ij}$ in a multiplicative or additive manner as [Li *et al.*, 2016a; Ye *et al.*, 2011].

$\mathcal{S}_-$, $\mathcal{S}_-^* = \{k,l\} \cup \mathcal{S}_-$, $\mathbf{a}_i' = \sigma'(\mathbf{z}_i)$, $\mathbf{b}_i' = -\sigma'(\mathbf{z}_i)$, $(\cdot)_q$ is bounded value between q_{min} and q_{max}, and $\sigma'(x) = \sigma(x)(1-\sigma(x))$, $y'(x) = \frac{1}{1+exp(-\alpha x)}$ are the derivatives of $\sigma(x)$ and $y(x)$, respectively.

Complexity Analysis. Sampling a tuple $(i,h,j,k,l,\mathcal{S}_-)$ has a constant cost $\mathcal{O}(|\mathcal{S}_-|)$ in each update, where $|\mathcal{S}_-|$ is the size of sampled unobserved POIs and usually very small. Hence, the overall complexity of optimization algorithm is $\mathcal{O}(d|\mathcal{S}_-|\#iter)$, where $\#iter$ is the total number of iterations. In practical, $\#iter \gg d|\mathcal{S}_-|$ is proportional to the number of the observed check-ins.

3 Experiments

3.1 Datasets

We use Gowalla and Foursquare datasets to evaluate model performance. Gowalla and Foursquare contain check-in data, ranging from January 2009 to August 2010, and from December 2009 to June 2013, respectively. Each check-in record in the datasets includes a user ID, a location ID and a timestamp, where each location has latitude and longitude information. We split the training and testing data as follows: for each individual user, (1) aggregating check-ins for each location; (2) sorting locations according to the first checked-in timestamp; (3) selecting the earliest 80% to train the model and using the next 20% as testing. The data statistics are shown in Table 1.

Table 1: The statistics of data sets.

Data Set	#Users	#Locations	#Records	Sparsity
Gowalla	52,216	98,351	2,577,336	0.0399%
Foursquare	2,551	13,474	124,933	0.2910%

3.2 Parameter Settings

All regularization parameters are set as 0.01. The parameter λ_u^i, α, η and d are set as 0.01, 5, 0.001, and 10. The λ_u^e, $|\mathcal{S}_-|$, q_{min}, and q_{max} are set as 0.1 (or 1), 10 (or 150), -0.3 (or -0.05) and 0 (or 0.05) in Gowalla (or Foursquare) dataset. The number and radius of activity area are set as 1 and $2km$.

3.3 Evaluation Metrics

We quantitatively evaluate model performance in terms of top-K recommendation performance, i.e., Precion@K and Recall@K, and ranking performance, i.e., MAP. They are formally defined as follows:

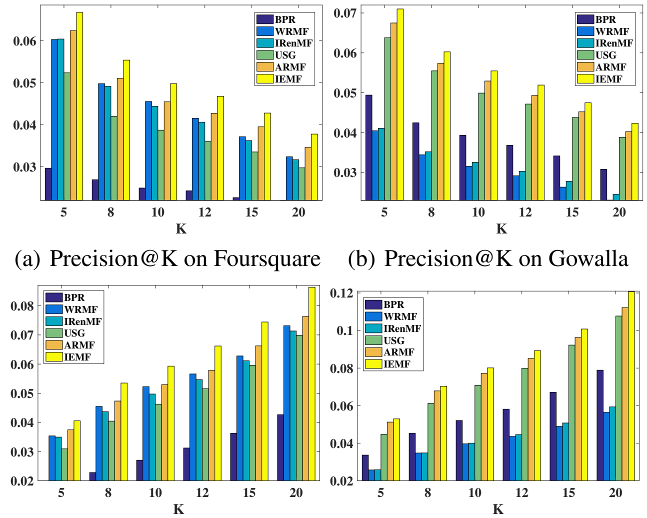
$$Precision@K = \frac{1}{n}\sum_{i=1}^{n}\frac{S_i(K)\cap\mathcal{T}_i}{K}, Recall@K = \frac{1}{n}\sum_{i=1}^{n}\frac{S_i(K)\cap\mathcal{T}_i}{|\mathcal{T}_i|},$$

$$MAP = \frac{1}{n}\sum_{i=1}^{n}\frac{\sum_{j=1}^{m}p(j)\times rel(j)}{|\mathcal{T}_i|},$$

where $S_i(K)$ is a set of top-K unvisited locations recommended to user i excluding those locations in the training, and $\mathcal{T}_i$ is a set of locations that are visited by user i in the testing. $p(j)$ is the precision of a cut-off rank list from 1 to j, and $rel(j)$ is an indicator function that equals to 1 if the location is visited in the testing, otherwise equals to 0.

3.4 Baseline Methods

To comprehensively demonstrate the effectiveness of our model, we compare them with the following popular models:



(a) Precision@K on Foursquare (b) Precision@K on Gowalla

(c) Recall@K on Foursquare (d) Recall@K on Gowalla

Figure 3: Performance comparison in terms of precision@K and recall@K on Foursquare and Gowalla datasets.

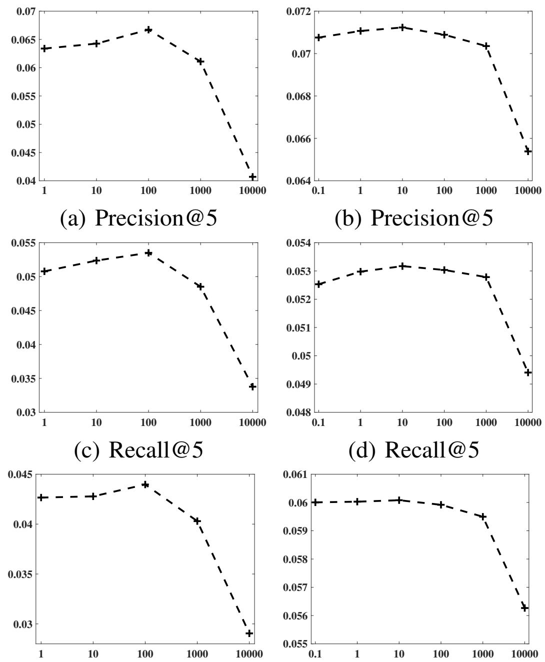
- **ARMF** [Li *et al.*, 2016a], which learns a set of user's potential locations from her unobserved locations using social network, and then incorporates them into matrix factorization with category and geographical information.
- **IRenMF** [Liu *et al.*, 2014], which incorporates neighboring characteristics in both instance level and region level into weighted matrix factorization;
- **USG** [Ye *et al.*, 2011], which incorporates geographical influence, social network and user interest into user-based collaborative filtering in an additive manner;
- **BRP** [Rendle *et al.*, 2009], which optimizes the ordering relationship of user's preferences for the observed location and the unobserved location;
- **WRMF** [Hu *et al.*, 2008], which minimizes the squared error loss by assigning both observed and unobserved check-ins with different weights based on matrix factorization.

3.5 Performance Comparison

In this section, we evaluate model performance from Precision@K, Recall@K and MAP on two datasets shown in Figure 3 and Table 2. We summarize the following observations.

First, our method outperforms all the other baseline methods. This superior result is for the sake of modeling fine-grained user interest, unobserved data and geographical influence together. Our better performance over those baseline methods with geo-influence, i.e., ARMF, USG, and IRenMF, further illustrates the benefit of capturing user's two forms of interest and adaptively learning unobserved data.

Second, WRMF and BRP perform differently in two datasets, where the former one is based on squared error and the latter is based on ranking error. Until now, there is no explicit clue to demonstrate which method is suitable for which type of data from Precision@K, Recall@K and MAP. From the technical viewpoint, our approach can be viewed as a unified method for ranking and squared error based loss functions. It is a tradeoff between these two popular matrix



(a) Precision@5 (b) Precision@5

(c) Recall@5 (d) Recall@5

(e) MAP (f) MAP

Figure 4: The influence of λ_u^i and λ_u^e on Foursquare dataset (left) and Gowalla dataset (right). The x-axis is λ_u^e/λ_u^i.

factorization methods. Although our approach does not have closed-form solution as WRMF does, the effective SGD with sampling can achieve the same effectiveness.

Third, those methods with geo-influence, i.e., IEMF, ARMF, USG, and IRenMF, are better than other methods without geo-influence, i.e., BPR and WRMF. This result further clarifies the importance of geo-influence in location recommendation, which also distinguishes location recommendation task from traditional item recommendation task.

Table 2: Performance comparison in terms of MAP.

Foursquare Dataset					
IEMF	ARMF	USG	IRenMF	WRMF	BPR
0.0440	0.0391	0.0346	0.0368	0.0363	0.0192

Gowalla Dataset					
IEMF	ARMF	USG	IRenMF	WRMF	BPR
0.0601	0.0571	0.0521	0.0255	0.0247	0.0365

3.6 Study of Influence of Parameters λ_u^i and λ_u^e

We study the influence of parameters λ_u^i and λ_u^e. Specifically, we set $\lambda_u^i = 0.01$ and then train our model with different λ_u^e, where the results are shown in Figure 4. From the results, we observe that the optimal λ_u^e/λ_u^i is 100 and 10 on Foursquare and Gowalla datasets, respectively. It indicates that the weight placed on the ranking constraint about extrinsic interest should be larger than the one about intrinsic interest. There are two reasons: for each user, (1) the number of unobserved locations outside her activity areas is usually larger than those within activity areas, and thus, a large λ_u^e can allow IEMF to model more unobserved data; (2) she has a large probability to check-in the locations within her activity areas due to geo-influence, and thus, a too large λ_u^i will increase the noise for model learning. If λ_u^e/λ_u^i is too large, IEMF almost focuses on optimizing the ranking constraint related to extrinsic interest, definitely resulting in a poor result.

4 Related Work

Related work of this paper can be grouped into two categories. The first category is about matrix factorization (MF) [He *et al.*, 2016; Chen *et al.*, 2016; Kabbur *et al.*, 2013; Balakrishnan and Chopra, 2012; Pilászy *et al.*, 2010; Salakhutdinov and Mnih, 2007; Pan and Chen, 2013]. The core of MF is to map user and item with two into low dimensional latent space. Different loss has been developed to model the implicit feedback. For example, [Hu *et al.*, 2008; Pan *et al.*, 2008] propose to minimize the sum-of-squared error with different weight over all user-item pairs, where the observed and unobserved feedbacks are assigned to one and zero, respectively. Another loss function is based on ranking error [Rendle *et al.*, 2009; Rendle and Freudenthaler, 2014] by optimizing the ranking order between observed examples and unobserved ones.

The second category is about POI recommendation with geo-influence [Li *et al.*, 2015; Lian *et al.*, 2014; Lichman and Smyth, 2014; Li *et al.*, 2016c]. For example, [Ye *et al.*, 2011; Li *et al.*, 2016a] propose to use a power law distribution to estimate the check-in probability with distance. With geo-influence, one user's interest in a location then can be calculated by a user-based collaborative filtering [Ye *et al.*, 2011] or learned by MF models [Li *et al.*, 2016a]. Also, [Liu *et al.*, 2014] proposes to model geographical neighboring influence from both instance level and region level.

Different from the aforementioned methods, in this paper, we propose a new fine-grained approach to model a user's general interest from both intrinsic and extrinsic perspectives. In addition, the unobserved feedbacks (i.e., missing data) are modeled by a self-adaptive location-oriented approach.

5 Conclusion

In this paper, we proposed a unified approach to integrate squared error loss and ranking error loss for solving location recommendation task by effectively learning fine-grained and interpretable user interest, and adaptively modeling the missing data. Specifically, each user's general interest is modeled as a mixture of her intrinsic and extrinsic interests, upon which we formulated the ranking constraints in our unified approach. Additionally, a self-adaptive location-oriented method is proposed to capture the characteristic of missing data, and is then formulated as the squared error loss in our unified optimization objective. To evaluate our model, we conducted extensive experiments on real-world datasets and compared our method with several baselines. The experimental results have shown the effectiveness of our model.

Acknowledgments

This work is partially supported by the NIH (1R21AA023975-01) and NSFC (61602234, 61572032, 91646204, 61502077).

References

[Balakrishnan and Chopra, 2012] Suhrid Balakrishnan and Sumit Chopra. Collaborative ranking. In *Proceedings of WSDM*, pages 143–152, 2012.

[Calder and Staw, 1975] Bobby J. Calder and Barry M. Staw. *Journal of personality and social psychology*, 31(4):599–605, 1975.

[Chen and Mangasarian, 1993] Chunhui Chen and O. L. Mangasarian. Smoothing methods for convex inequalities and linear complementarity problems. *Mathematical Programming*, 71:51–69, 1993.

[Chen *et al.*, 2016] Haolan Chen, Di Niu, Kunfeng Lai, Yu Xu, and Masoud Ardakani. Separating-plane factorization models: Scalable recommendation from one-class implicit feedback. In *Proceedings of CIKM*, pages 669–678, 2016.

[Devooght *et al.*, 2015] Robin Devooght, Nicolas Kourtellis, and Amin Mantrach. Dynamic matrix factorization with priors on unknown values. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 189–198, 2015.

[He *et al.*, 2016] Xiangnan He, Hanwang Zhang, Min-Yen Kan, and Tat-Seng Chua. Fast matrix factorization for online recommendation with implicit feedback. In *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 549–558, 2016.

[Hu *et al.*, 2008] Yifan Hu, Yehuda Koren, and Chris Volinsky. Collaborative filtering for implicit feedback datasets. In *Proceedings of ICDM*, pages 263–272, 2008.

[Kabbur *et al.*, 2013] Santosh Kabbur, Xia Ning, and George Karypis. Fism: Factored item similarity models for top-n recommender systems. In *Proceedings of SIGKDD*, pages 659–667, 2013.

[Lee and Mangasarian, 2001] YUH-JYE Lee and O. L. Mangasarian. Ssvm: A smooth support vector machine for classification. *Computational Optimization and Applications*, 20(1):5–22, 2001.

[Li *et al.*, 2015] Huayu Li, Richang Hong, Shiai Zhu, and Yong Ge. Point-of-interest recommender systems: A separate-space perspective. In *Proceedings of ICDM*, pages 231–240, 2015.

[Li *et al.*, 2016a] Huayu Li, Yong Ge, Richang Hong, and Hengshu Zhu. Point-of-interest recommendations: Learning potential check-ins from friends. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 975–984, 2016.

[Li *et al.*, 2016b] Huayu Li, Richang Hong, Defu Lian, Zhiang Wu, Meng Wang, and Yong Ge. A relaxed ranking-based factor model for recommender system from implicit feedback. In *Proceedings of IJCAI*, pages 1683–1689, 2016.

[Li *et al.*, 2016c] Huayu Li, Richang Hong, Zhiang Wu, and Yong Ge. A spatial-temporal probabilistic matrix factorization model for point-of-interest recommendation. In *Proceedings of SDM*, pages 117–125, 2016.

[Lian *et al.*, 2014] Defu Lian, Cong Zhao, Xing Xie, Guangzhong Sun, Enhong Chen, and Yong Rui. Geomf: Joint geographical modeling and matrix factorization for point-of-interest recommendation. In *Proceedings of SIGKDD*, pages 831–840, 2014.

[Lichman and Smyth, 2014] Moshe Lichman and Padhraic Smyth. Modeling human location data with mixtures of kernel densities. In *Proceedings of SIGKDD*, pages 35–44, 2014.

[Liu *et al.*, 2014] Yong Liu, Wei Wei, Aixin Sun, and Chunyan Miao. Exploiting geographical neighborhood characteristics for location recommendation. In *Proceedings of CIKM*, pages 739–748, 2014.

[Pan and Chen, 2013] Weike Pan and Li Chen. Gbpr: group preference based bayesian personalized ranking for one-class collaborative filtering. In *Proceedings of IJCAI*, pages 2691–2697, 2013.

[Pan *et al.*, 2008] Rong Pan, Yunhong Zhou, Bin Cao, Nathan N. Liu, Rajan Lukose, Martin Scholz, and Qiang Yang. One-class collaborative filtering. In *Proceedings of ICDM*, pages 502–511, 2008.

[Pilászy *et al.*, 2010] István Pilászy, Dávid Zibriczky, and Domonkos Tikk. Fast als-based matrix factorization for explicit and implicit feedback datasets. In *Proceedings of the Fourth ACM Conference on Recommender Systems*, pages 71–78, 2010.

[Rendle and Freudenthaler, 2014] Steffen Rendle and Christoph Freudenthaler. Improving pairwise learning for item recommendation from implicit feedback. In *Proceedings of WSDM*, pages 273–282, 2014.

[Rendle *et al.*, 2009] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. Bpr: Bayesian personalized ranking from implicit feedback. In *the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 452–461, 2009.

[Ryan and Deci, 2000] Richard M. Ryan and Edward L. Deci. Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, 25(1):54–67, January 2000.

[Salakhutdinov and Mnih, 2007] Ruslan Salakhutdinov and Andriy Mnih. Probabilistic matrix factorization. In *NIPS*, pages 1257–1264, 2007.

[Yao *et al.*, 2014] Yuan Yao, Hanghang Tong, Guo Yan, Feng Xu, Xiang Zhang, Boleslaw K. Szymanski, and Jian Lu. Dual-regularized one-class collaborative filtering. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, pages 759–768, 2014.

[Ye *et al.*, 2011] Mao Ye, Peifeng Yin, Wang-Chien Lee, and Dik-Lun Lee. Exploiting geographical influence for collaborative point-of-interest recommendation. In *Proceeding of SIGIR*, pages 325–334, 2011.