Uncertainty-Aware Heterogeneous Representation Learning in POI Recommender Systems

Fan Zhou, *Member, IEEE*, Tangjiang Qian, Yuhua Mo, Zhangtao Cheng, Chunjing Xiao, Jin Wu, and Goce Trajcevski, *Member, IEEE*

***Abstract*—Discovering interesting yet unvisited point-of-interests (POIs) is among the most practical applications but challenging problems in location-based social networks (LBSNs). Popular approaches face several issues, such as data sparsity and difficulties in modeling latent nonlinearity between users and POIs. Furthermore, the uncertainty in LBSNs poses additional obstacles to learning good representations of users' general and current interests. To effectively address these issues, we postulate that fusing multiple sources of information is paramount. Toward that, we propose a novel deep generative recommender system—Wasserstein autoencoder for POI recommendation (WaPOIR). It unifies the information from users' personal preference, social influence, and geographical data, and captures users' general interests from historical check-ins, while modeling users' current interests from recently visited POIs. Unlike previous methods, WaPOIR learns the latent distribution of data in the Wasserstein space as a potential representation for each POI and each user in LBSNs. This enables simultaneous maintenance of social and POI interactions and modeling the uncertainty of their relationships. WaPOIR is a stochastic recommendation approach that allows Bayesian inference and approximation of variational posterior distribution. Extensive experiments conducted on real-world LBSN datasets demonstrate that WaPOIR achieves better performance over the state-of-the-art approaches.**

***Index Terms*—Adversarial learning, collaborative filtering (CF), point-of-interest (POI) recommendation, variational inference, Wasserstein autoencoder.**

Manuscript received 18 February 2022; revised 6 November 2022; accepted 25 February 2023. Date of publication 15 March 2023; date of current version 16 June 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 62176043 and Grant 62072077; in part by the Science and Technology Foundation of Henan Province of China under Grant 212102210387; in part by the Natural Science Foundation of Sichuan Province under Grant 2022NSFSC0505; and in part by the National Science Foundation SWIFT under Grant 2030249. This article was recommended by Associate Editor J. Lu. *(Corresponding authors: Chunjing Xiao; Jin Wu.)*

Fan Zhou, Tangjiang Qian, Yuhua Mo, Zhangtao Cheng, and Jin Wu are with the University of Electronic Science and Technology of China, Chengdu 610054, China (e-mail: fan.zhou@uestc.edu.cn; tj.qian@std.uestc.edu.cn; moyuhua95@gmail.com; zhangtao.cheng@std.uestc.edu.cn; wj@uestc.edu.cn).

Chunjing Xiao is with the Henan Key Laboratory of Big Data Analysis and Processing, Henan University, Kaifeng 475004, China (e-mail: chunjingxiao@gmail.com).

Goce Trajcevski is with the Department of Electrical and Computer Engineering, Iowa State University, Ames, IA 50011 USA (e-mail: gocet25@iastate.edu).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TSMC.2023.3252079.

Digital Object Identifier 10.1109/TSMC.2023.3252079

I. Introduction

The popularity of location-aware social media has brought about a number of location-based social networks (LBSNs), such as Yelp and Foursquare. These, in turn, provide references for users to make an informed visiting decision. A large amount of user-point-of-interest (POI) interaction data facilitates a range of promising services, among which personalized POI recommendation is one of the most important applications and consistently received attention in recent years. The goal of POI recommendation is to learn the users' visiting preferences from the visiting history based on their spatial, temporal, and other contextual information—and then provide a ranked list of POIs that the user may be interested in but has not visited yet [1], [2].

In general, most of the existing models [3], [4], [5], [6], [7] predict the user preference relying on the user-POI rating matrix. Collaborative filtering (CF) techniques exploit the user-POI interactive patterns (the rating matrix) via a matrix factorization [8] to make relevant predictions. A common drawback in these methods is that they ignore negative samples which might help improve the recommendation performance [9], [10]. Thus, researchers have incorporated the ranking methods into the recommendation procedure. For example, a ranking-based geographical factorization method (Rank-GeoFM) leveraging both visited POIs and unvisited POIs for POI recommendation was developed in [6]. Subsequently, a listwise ranking system by injecting users' geosocial preferences was proposed in [11]. Such CF-based methods typically suffer from: 1) sparsity issues because the preference dynamics and auxiliary information related to users' interaction are ignored and 2) user-POI interactions are modeled in a simple linear way, rather than using more complicated nonlinear relations [12].

Our perspective for the problem settings is illustrated in Fig. 1. We consider three graph-structured data sources, characterizing the relationships among the entities of interest: 1) users' check-ins; 2) social friendships among users; and 3) geographical relationships among POIs. We postulate that, while each of them can be used to learn the latent factors for POI recommendation in LBSNs applications, fusing them in an integrated approach can greatly improve the quality/effectiveness of such recommendations.

A plethora of POI recommendation approache have attempted to improve the performance by adding rich auxiliary information available in LBSNs with memory-based CF. For

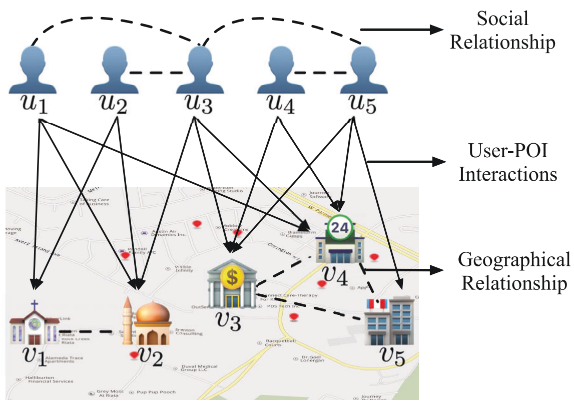


Fig. 1. Illustration of three different types of relationships.

example, some works [4], [5] incorporate geographical, temporal, and social information for POI recommendation to tackle the data sparsity issue. Recently, Qian et al. [13] proposed a translation-based recommender framework to model the relationship among users, POIs, and spatiotemporal contexts for POI recommendation. Though achieving improvement to some extent, these methods still have certain limitations, especially in their modeling capacity—in the sense that they are not able to learn suitable representation from the primary data.

Recently developed deep learning-based recommendation approaches are capable of not only modeling the nonlinearity, but also showing a great potential for learning effective representations, which enabled them to achieve promising performance improvements in recommendation systems. For example, a novel deep neural network with a co-attention mechanism by leveraging rich meta-path-based context was proposed in [14]. It was able to learn interaction-specific representations for users, items and meta-path context. The Bayesian recommendation model proposed in [15] can help alleviate the issues of sparsity and cold start. However, the formation and evolution of real-world LBSNs abound with uncertainties from various (heterogeneous) sources [16]. For example, users tend to change behavior due to the new POI exploration, user experience, POIs' popularity, and social influence [17], which makes existing Bayesian methods inapplicable for large-scale POI recommendation.

Inspired by the recent advances of coupling neural networks with the CF and the deep generative models such as variational autoencoder (VAE) [18], we tackle the POI recommendation problem by: 1) learning deep latent representations from geographical influence along with user preference and social influence and 2) capturing implicit relationships between POIs, users and spatiotemporal contexts from both users' historical and recent check-ins with an attention network. Different from conventional recommendation techniques directly learning user preference, social, and geographical influence, we present a novel auxiliary information fusion method with stochastic inference for POI recommendation.

We propose the Wasserstein autoencoder for POI recommendation (WaPOIR) model, which learns the Gaussian distribution embeddings of users and POIs, and employs Wasserstein distance (also known as the Earth Mover's distance) to measure the similarity between the model distribution and the real distribution. We also use Wasserstein distance to calculate the latent representation similarity between users and the similarity between POIs, aiming to identify more relevant users or POIs from any user pairs and POI pairs [19]. The main contributions of this work are listed.

1) We present a new CF model that leverages Bayesian inference and probabilistic generative model with VAE-based networks for POI recommendation, which can capture the nonlinear and significant user-POI relationships. Moreover, our model can alleviate the data sparsity issue by incorporating the geographical influence along with user preference and social influence. In addition, our novel neural attention network can simultaneously capture users' general and current interests.
2) Different from conventional VAE-based recommender systems, we learn the Gaussian distribution embedding of input data in the Wasserstein space. Our model satisfies the triangle inequality (cf. [20]), which can well preserve the transitivity in LBSNs. More importantly, it introduces the representation uncertainty of both users and POIs, as well as their similarities which, in turn, improves the learning of their latent representations and interactions.
3) When constructing the training set that computes the similarity of latent representations between users and POIs, we use the same constructive method as BPR-based methods. However, for calculating the stochastic gradient descents of the objective function, we develop a sampling scheme, which not only optimizes the objective function, but also significantly alleviates the time complexity issue of learning the ranks of POIs.

We conducted a comprehensive experimental evaluation on three real-world datasets, demonstrating that our proposed WaPOIR model significantly outperforms the state-of-the-art baselines for POI recommendation.

This article is structured as follows. In Section II, we formally define the studied problem and introduce necessary backgrounds. In Section III, we detail our model framework. We present the results from the comprehensive experimental evaluations in Section IV and review the related literature in Section V. Section VI concludes this work.

II. Preliminaries

We now formalize the POI recommendation problem and summarize the basic notations used throughout this article. Subsequently, we describe the necessary background regarding VAEs and Wasserstein distance.

A. Problem Definition

The setting that we consider is represented by two basic sets: a set of users $\mathcal{U} = \{u_1, \ldots, u_m\}$ and a set of POIs $\mathcal{V} = \{v_1, \ldots, v_n\}$.

For a given user u_i, we assume that there is a record of his history of visited POIs (i.e., check-ins)—$\mathcal{V}_{u_i} = [v_{(u_i,1)}, \ldots, v_{(u_i,l)}]$, ordered by the time of the visit—i.e., $v_{(u_i,j)}.t < v_{(u_i,j+1)}.t$, whereby $v_{(u_i,l)}$ is the most recently visited

TABLE I
LIST OF NOTATIONS

Notation	Description
$\mathbf{C}$	user-POI check-in matrix.
$\mathcal{U}/\mathcal{V}$	user/POI set.
D	the latent dimension of $\mathbf{U}$ and $\mathbf{V}$.
$\mathbf{U}/\mathbf{V}$	the latent factors of users/POIs.
$\mathcal{N}(v_j)$	the neighbor set of POI v_j.
$\mathcal{N}(u_i)$	the neighbor set of user u_i.
$\mathcal{V}_{u_i}$	the historical check-ins of user u_i.
G_p	distance matrix between POIs.
G_u	social relationship between users.
$\mathcal{W}$	geographical influence matrix.
$\mathcal{S}$	user similarity matrix.
Q	user preference matrix.
$\mathbf{x}, \mathbf{y}/\tilde{\mathbf{x}}, \tilde{\mathbf{y}}$	input data and reconstructed data.
$\mathbf{z}/p(\mathbf{z})$	latent variable and its prior.
$\mathbf{u_i}/\mathbf{v_j}$	user/POI representation vector.
l_j	the geographical coordinate of j-th POI.

POI by u_i. The POI recommendation problem aims at determining (i.e., recommending) the top-K *new* POIs $\mathcal{V}/\mathcal{V}_{u_i}$ that user u_i would like to visit in the future given his/her historical visited check-in set.

The user-POI relationship is represented by a feedback matrix (i.e., check-in matrix) $\mathbf{C} \in \mathbb{R}^{m\times n}$. Each entry c_{ij} represents the check-in history of user u_i with respect to the POI v_j, i.e., $c_{ij} = 1$ means that the POI v_j has been visited by the user u_i, and $c_{ij} = 0$ denotes that u_i has not yet visited v_j.

We assume that all of the latent representations are defined as a lower-dimensional Gaussian distribution embeddings and use D to denote the dimension of the latent feature factors. The corresponding representations for the users and POIs are denoted as $\mathbf{U} = \{\mathbf{u_1}, \dots, \mathbf{u_m}\} \in \mathbb{R}^{m\times D}$ and $\mathbf{V} = \{\mathbf{v_1}, \dots, \mathbf{v_n}\} \in \mathbb{R}^{n\times D}$, respectively.

In accordance with [21], [22], we use side information about users and POIs to train our model. For each POI, we use $l_j = \{\text{longitude, latitude}\}$ to denote its geographical location as a pair of the corresponding coordinates. Let the matrix $G_p \in \mathbb{R}^{n\times n}$ represent the pair-wise distances—i.e., $g_{jk} = d_{jk}(l_j, l_k)$—between POIs v_j and v_k. If $d_{jk}(l_j, l_k)$ is less than a given threshold d_t, we consider them to be neighbors. We use the symbol $\mathcal{N}_{v_j} = \{v_k | d_{jk}(l_j, l_k) < d_t\}$ to indicate the neighborhood set of the POI v_j. We assume that there exist other types of information context for each POI—e.g., tag words, spatio-temporal visitors density, etc. For users, we use an undirected graph $G_u = (\mathcal{U}, \mathcal{E})$ to represent a social network, where edges in $\mathcal{E}$ denote friendship between users in $\mathcal{U}$ and $\mathcal{N}(u_i) = \{u_j | (u_i, u_j) \in \mathcal{E}\}$ denote the neighbor set of user u_i. A summary of the symbols used in this article is presented in Table I.

B. Background

VAE [18]: Is a probabilistic generative network which models the data distribution using amortized variational inference. It consists of an encoder and a decoder, aiming at utilizing $p_\theta(\mathbf{x}) := \int_z p_\theta(\mathbf{x}|\mathbf{z})p(\mathbf{z}))d\mathbf{z}$ to maximize

$$\log p_\theta(\mathbf{x}) \geq \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}\big[\log p_\theta(\mathbf{x}|\mathbf{z})\big] - D_{\mathbb{KL}}\big[q_\phi(\mathbf{z}|\mathbf{x})||p(\mathbf{z})\big] = \mathcal{L}_{\text{VAE}}(\mathbf{x}; \theta, \phi) \quad (1)$$

with respect to $p_\theta(\mathbf{x}|\mathbf{z})$. It describes a generative model parameterized by θ, which usually is chosen to follow a conditional Gaussian distribution. The Kullback–Leibler ($D_{\mathbb{KL}}$) in (1) can be viewed as a regularizer. Normally, in order to calculate $D_{\mathbb{KL}}$ easily, VAE assumes that the prior distribution $p(\mathbf{z})$ is a standard Gaussian distribution. And the term $q_\phi(\mathbf{z}|\mathbf{x})$ describes a recognition model that encodes the input data to a latent factor $\mathbf{z}$, which encourages the posterior to match the prior distribution $p(\mathbf{z})$. That is, VAE is minimizing the *evidence lower bound* (ELBO) on the negative marginal log-likelihood or, equivalently, on the Kullback-Leibler (KL)-divergence $D_{\mathbb{KL}}[q_\phi(\mathbf{z}|\mathbf{x})||p(\mathbf{z})]$.

Since the latent variables are stochastic, the probability distributions are learned for each data sample. It has been proved that the learned data variation may lead to more robust representations and yield enhanced recommendation performance compared to the deterministic neural network-based methods [15], [23], [24].

Wasserstein Distance or Kantorovich–Rubinstein metric, is a distance function defined between probability distributions on a given metric space. It can be used to measure the distance (similarity) between two distributions, which is defined as follows [25], [26]:

$$W\big(p_x, p_g\big)_p = \left(\inf_{\pi\sim\Pi(p_x,p_g)} \mathbb{E}_{(a,b)\sim\pi}\big[d(a,b)^p\big]\right)^{1/p} \quad (2)$$

where $p \geq 1$ and $\Pi(p_x, p_g)$ is a set of all possible joint distributions with marginals p_x and p_g. It is a.k.a the Earth-mover distance (EMD) when $p = 1$.

The main advantage of the *Wasserstein distance* compared to the KL divergence and the Jensen–Shannon (JS) divergence is that the distances of the two distributions can still be reflected even if the support sets of the two distributions do not overlap or have very little overlap. In this work, we use the Wasserstein distance as the similarity measure because it is able to calculate the similarity between two distributions while simultaneously satisfying the triangle inequality. Since we use distributions rather than deterministic vectors to embed the POIs and users, the Wasserstein distance is suitable to be a similarity measure of the latent representations of POIs and users, which can guarantee the model to preserve the transitivity of similarity between POI/user pairs.

III. METHOD: WAPOIR

This section presents the details of the proposed model—WaPOIR—the main components of which are illustrated in Fig. 2. It is inherently a deep generative model for item recommendation, which learns the Gaussian distribution as a potential representation of each input data in the Wasserstein space. Specifically, we use ranking-based methods to calculate similarity between entities (both users and POIs) using the Wasserstein distance. Through combining users' historical check-in information, along with the auxiliary information associated with the entities, we obtain two features matrices as input and use our WaPOIR to learn the Wasserstein distance. This is the main difference of our model from the conventional Bayesian recommendation models [23], [24], [27].

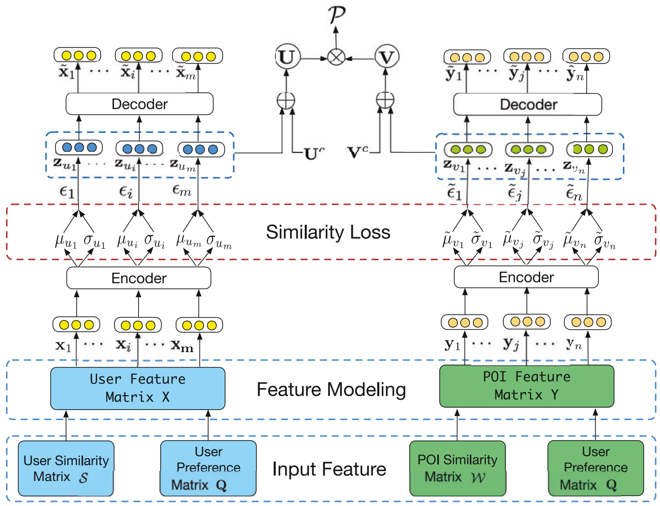


Fig. 2. Overview of the WaPOIR model.

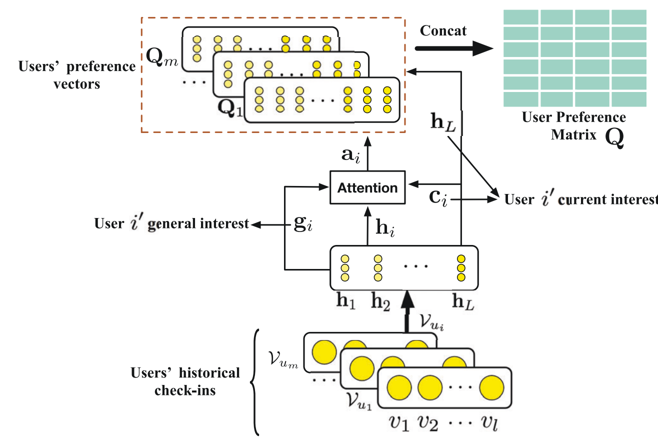


Fig. 3. Illustration of user preference learning.

We first discuss the method of measuring user similarity and POI similarity, and present the way of modeling users' long-short interests with attention mechanism. Then, we describe the details of uncertainty-aware user and POI representation learning, followed by the pairwise similarity measure and the method for computing the Wasserstein distance. Finally, we present the implementation details of the proposed model.

A. User Similarity and POI Similarity Measure

Recall that there are three graph-like data structures that we rely upon: 1) users' check-ins $\mathbf{C}$; 2) social friendship of users G_u; and 3) geographical relationship among POIs G_p. In WaPOIR, we provide a deep variational model to train the representations of users and POIs, and then use CF to recommend POIs for users. In particular, let $\mathbf{X} = \{\mathbf{x_1}, \ldots, \mathbf{x_m}\} \in \mathbb{R}^{m\times n}$ be users' feature matrix, and $\mathbf{Y} = \{\mathbf{y_1}, \ldots, \mathbf{y_n}\} \in \mathbb{R}^{n\times m}$ be the corresponding matrix for the POIs

$$\mathbf{X} = \mathrm{Norm}(\mathcal{S} \otimes \mathbf{Q}), \mathbf{Y} = \mathrm{Norm}\big(\mathcal{W} \otimes \mathbf{Q}^\mathsf{T}\big) \tag{3}$$

where Norm$(\cdot)$ means normalizing the matrix and $\otimes$ denotes a multiplication of matrices; $\mathcal{S} \in \mathbb{R}^{m\times m}$ denotes user similarity matrix. $\mathcal{S}_{ij}$ represents the similarity between user u_i and user u_j, and its value is between 0 and 1, which is calculated by

$$\mathcal{S}_{ij} = \gamma U_{ij} + (1-\gamma)F_{ij}. \tag{4}$$

In the equation above, γ is a hyperparameter for balancing the social influence and user influence; F_{ij} denotes the friendship similarity; and U_{ij} is the cosine similarity between user i and user j. According to [28], they can be computed as follows:

$$F_{ij} = \beta \cdot \frac{\left|F_{u_i} \cap F_{u_j}\right|}{\left|F_{u_i} \cup F_{u_j}\right|} + (1-\beta) \cdot \frac{\left|\mathcal{V}_{u_i} \cap \mathcal{V}_{u_j}\right|}{\left|\mathcal{V}_{u_i} \cup \mathcal{V}_{u_j}\right|}$$
$$U_{ij} = \frac{\sum_{v_k\in\mathcal{V}} c_{ik}\cdot c_{jk}}{\sqrt{\sum_{v_k\in\mathcal{V}} {c_{ik}}^2}\cdot\sqrt{\sum_{v_k\in\mathcal{V}} {c_{jk}}^2}}. \tag{5}$$

The purpose of β is to balance the friend influence based on social connections and check-in activities; F_u and $\mathcal{V}_u$ denote the friend set and user historical check-in set, respectively.

In (3), $\mathcal{W} \in \mathbb{R}^{n\times n}$ is the geographical influence matrix, where each value $W_{ij} \in \mathcal{W}$ represents the mutual impacts between POI v_i and POI v_j. In our settings, this implies that users are likely to visit POIs close to their homes or offices, and they may also like to explore POIs that near other POIs they are in favor of. In order to incorporate geographical influence on nearby *unvisited* POIs, we use the Gaussian radial basis function (RBF) to calculate the value of W_{ij} as follows:

$$W_{ij} = \exp\Big(-\xi\left\|l_i - l_j\right\|^2\Big) \tag{6}$$

where ξ is a hyper-parameter.

B. Modeling Users' Long-Short Interests

We first present a novel attention net for capturing users' long (general) and short (current) interests, as illustrated in Fig. 3. Given a user u_i and his historical check-in set $\mathcal{V}_{u_i} = [v_1, \ldots, v_k, \ldots, v_l] = [v_{(u_i,1)}, \ldots, v_{(u_i,k)}, \ldots, v_{(u_i,l)}]$ of length L, we denote the first check-in as one-hot vector $\mathbf{h}_1 \in \mathbb{R}^{1\times n}$, the kth check-in as $\mathbf{h}_k \in \mathbb{R}^{1\times n}$. We use the vectors $\mathbf{c}_i$ and $\mathbf{g}_i$ to denote the current (the latest) interest and general interest of user u_i, respectively

$$\mathbf{c}_i = \mathbf{h}_L, \mathbf{g}_i = \frac{1}{L}\sum_{k=1}^{L}\mathbf{h}_k. \tag{7}$$

According to (7), we can calculate the attention coefficient vector $\alpha_{u_i} = \{\alpha_1, \ldots, \alpha_k, \ldots, \alpha_L\}$ of historical check-ins with respect to user u_i as follows:

$$\alpha_k = \delta\big(\mathbf{c}_i\mathbf{w}_c + \mathbf{h}_k\mathbf{w}_h + \mathbf{g}_i\mathbf{w}_g + \mathbf{b}_i\big) \tag{8}$$

where $\delta(\cdot)$ denotes a nonlinear activation function (we use sigmoid in this work), $\mathbf{c}_i, \mathbf{g}_i \in \mathbb{R}^{1\times n}$, $\mathbf{w}_c, \mathbf{w}_h, \mathbf{w}_g \in \mathbb{R}^{n\times n}$ are weighting matrices, and $\mathbf{b}_i$ is a bias vector.

Then, a general interest vector about user u_i can be recalculated as follows:

$$\mathbf{a}_i = \sum_{k=1}^{L}\alpha_k\mathbf{h}_k. \tag{9}$$

Considering user's current interest, we can obtain the preference vector of user u_i for all POIs as follows:

$$\mathbf{Q}_i = \mathbf{a}_i + \alpha_L \mathbf{h}_L \tag{10}$$

where matrix $\mathbf{Q} \in \mathbb{R}^{m\times n}$ denotes the preference for all users. For all users, the loss function is defined as the cross-entropy

$$\mathcal{L}_{ec} = \sum_{i=1}^{m}\left(-\sum_{j=1}^{n} c_{ij}\log\big(\mathbf{Q}_{ij}\big) + \big(1-c_{ij}\big)\log\big(1-\mathbf{Q}_{ij}\big)\right) \tag{11}$$

where $c_{ij} \in \mathbf{C}$ (i.e., c_{ij} is either 0 or 1).

C. Uncertainty-Aware Representation Learning

When it comes to recommendation systems, there are many Bayesian collaborative methods [24], [29] that purely use user rating matrix to predict missing rating. Recent work [23] has incorporated feedback matrix and side information—which achieved a positive effect on improving the inference of users' and items' latent factors. An important assumption, though, is that the links between entities in the network are deterministic. However, the real-world user-POI interactions are full of uncertainty, which renders these recommendation approaches to be nonoptima. Inspired by the network embedding method [30], the proposed WaPOIR exploits the stochastic variational inference that can learn the mean and uncertainty of the entity through the mathematical correlation of mean and variance by the sampling process, respectively. For all the users and POIs in the network, their corresponding latent representations are generated through the respective latent variables $\mathbf{z}_{u_i} \in \mathbb{R}^{1\times D}$ and $\mathbf{z}_{v_j} \in \mathbb{R}^{1\times D}$, where $\mathbf{z}_{u_i}$ and $\mathbf{z}_{v_j}$ are inferred by input features with user preference matrix and auxiliary information, respectively. The user and POI generative latent variable models are, respectively, parameterized by θ_u and θ_v. Similarly, the user and POI inference networks are, respectively, parameterized by ϕ_u and ϕ_v. Then, we can obtain users' and POIs' representations as follows:

$$\mathbf{u}_i \sim p_{\theta_u}\big(\mathbf{u}_i|\mathbf{z}_{u_i}\big),\ \ \mathbf{z}_{u_i} \sim q_{\phi_u}\big(\mathbf{z}_{u_i}|\mathbf{x}_i\big) \tag{12}$$

$$\mathbf{v}_j \sim p_{\theta_v}\big(\mathbf{v}_j|\mathbf{z}_{v_j}\big),\ \ \mathbf{z}_{v_j} \sim q_{\phi_v}\big(\mathbf{z}_{v_j}|\mathbf{y}_j\big). \tag{13}$$

Similar to traditional probabilistic matrix factorization (PMF) approaches (cf. [31]), we also consider collaborative information associated with both users and POIs. We note that the collaborative information is embedded in the collaborative latent variables subjected to normal distribution

$$\mathbf{u}_i^c \sim N(0, I_D), \mathbf{v}_j^c \sim N(0, I_D). \tag{14}$$

Thus, the latent representations of users and POIs are combined with two latent variables as follows:

$$\mathbf{u}_i = \mathbf{u}_i^c + \mathbf{z}_{u_i}, \mathbf{v}_j = \mathbf{v}_j^c + \mathbf{z}_{v_j}. \tag{15}$$

The probability of user u_i visiting POI v_j is drawn from the inner product between their latent representations

$$\mathcal{P}_{ij} = \mathbf{u}_i \mathbf{v}_j^{\mathsf{T}} \tag{16}$$

where $\mathcal{P} \in \mathcal{R}^{m\times n}$ is the probability matrix that all users visit all POIs.

With the WaPOIR model constructed (cf. Fig. 2), the joint distribution of our model is given as follows:

$$p(\mathbf{U}, \mathbf{V}, \mathcal{P}, \mathbf{Z}_u, \mathbf{Z}_v) = \prod_{i,j} \underbrace{p\big(\mathcal{P}_{ij}|\mathbf{u}_i, \mathbf{v}_j\big)}_{\text{Bernoulli}} \cdot \underbrace{p\big(\mathbf{u}_i|\mathbf{z}_{u_i}\big)p\big(\mathbf{z}_{u_i}\big)}_{\text{for users}} \cdot \underbrace{p\big(\mathbf{v}_j|\mathbf{z}_{v_j}\big)p\big(\mathbf{z}_{v_j}\big)}_{\text{for POIs}} \tag{17}$$

where we ignore the subscripts without creating ambiguity.

In the remainder of this article, we only present the method for user-aspect representation learning when there is no ambiguity for the sake of simplicity. We note that the POI-aspect learning is in the same way.

To further improve the POI recommendation, we need to draw the latent factors of user u_i and POI v_j from $\mathbf{z}_{u_i}$ and $\mathbf{z}_{v_j}$ through a generative neural network. This suggests that the joint probability distribution in (17) is not our main task. Instead, we are interested in the posterior over users' and POIs' latent factors. Inspired by recent successes of VAEs [18], we use stochastic gradient variational Bayes [32] to approximate the posteriors. In our model, we introduce two VAEs, each consisting of an inference network (encoder) and a generation network (decoder). One of the inference networks is used for approximating posteriors of the latent variables of user u_i, whereas the other one targets the latent variables related to POI v_j. More specifically, we have

$$q\big(\mathbf{z}_{u_i}, \mathbf{z}_{v_j}|\mathbf{x}_i, \mathbf{y}_j\big) = q_{\phi_u}\big(\mathbf{z}_{u_i}|\mathbf{x}_i\big)\cdot q_{\phi_v}\big(\mathbf{z}_{v_j}|\mathbf{y}_j\big) \tag{18}$$

where the prior distributions of $\mathbf{z}_{u_i}$ and $\mathbf{z}_{v_j}$ are usually assumed to be diagonal-covariance Gaussian distributions $N(\mu, \sigma)$, satisfying $\mu \in \mathbb{R}^D$, $\sigma \in \mathbb{R}^{D\times D}$. Thus, in our model, we set the latent variables related to users as follows:

$$q_{\phi_u}\big(\mathbf{z}_{u_i}|\mathbf{x}_i\big) = N\Big(\mu_{\phi_u}(\mathbf{x}_i), \mathrm{diag}\Big({\sigma_{\phi_u}}^2(\mathbf{x}_i)\Big)\Big). \tag{19}$$

The input data $\mathbf{x}_i$ and $\mathbf{y}_j$ are obtained from the user preference matrix and the side information for users and POIs according to (3). When optimized, we push the variational posterior distribution $q_{\phi_u}(\mathbf{z}_{u_i}|\mathbf{x}_i)$ and $q_{\phi_v}(\mathbf{z}_{v_j}|\mathbf{y}_j)$ to approximate the intractable distribution $p(\mathbf{z}_{u_i})$ and $p(\mathbf{z}_{v_j})$, respectively. As is standard when learning latent-variable models with variational inference [33], we can lower-bound the log marginal likelihood (described in Section II-B). That is, the maximization of ELBO $\mathcal{L}_{\mathrm{VAE}}(\mathbf{x}; \theta, \phi)$ can be alternatively done by minimizing the KL divergence between the posterior and the prior. This is the well-known variational lower bound which provides a theoretically grounded framework successfully employed by VAEs [18], [34]. It makes use of amortized variational inference—trained with the reparameterization trick [18]—to propagate stochastic gradients from the decoder to the inference model.

Similar to VAE-based recommendation models [15], [23], [24], the objective of our model also contains two terms: one is the loss reconstruction, which aims at capturing the information of the users' and the POIs' input and the other is the regularization term [cf. (1)]—it encourages the encoded training distribution to match the prior distribution. However, instead of KL divergence used in previous works [15], [23], [24], we use a Wasserstein distance to minimize the distance between the model distribution and the target distribution, which leads to a different regularization than the one used in typical VAEs [18]. According to [20], the Wasserstein distance allows our model to compute the similarity between any distributions.

D. Pairwise Ranking

In many real-world networks, if there exists an observed edge linking two entities, then these two entities are considered to have similar characteristics. For example, if two users are friends in LBSNs, they may have a common interest. Complementary, users typically visit POIs that are geographically close to those POIs they have visited. However, real-world LBSN data are usually too sparse to capture a large number of links. Toward that, we introduce similarity calculations between a pair of entities based on their neighborhood network structure, even if there is no edge between them using link prediction techniques [35].

As proximity-criteria for entities in LBSNs, we would like the entities to be closer to their neighbors in the latent space. In this way, we can alleviate the sparsity and better capture a kind of a global network structure [36]. Following the Bayesian ranking with implicit feedback [19], we first create two training datasets, $\mathcal{D}_u$ and $\mathcal{D}_v$ for users and POIs, respectively, e.g.,

$$\mathcal{D}_u = \left\{ \left(u_i, u_j, u_k\right) | u_j \in \mathcal{N}(u_i) \wedge u_k \in \mathcal{U}/\mathcal{N}(u_i) \right\}. \tag{20}$$

Then, we combine all the pairwise constraints as follows:

$$\mathcal{L}_{\text{pair}}^{u} = \sum_{(u_i, u_j, u_k) \in \mathcal{D}_u} \left[W\left(\mathbf{u}_i, \mathbf{u}_j\right) + \exp\left(-W\left(\mathbf{u}_j, \mathbf{u}_k\right)\right) \right] \tag{21}$$

where $\mathbf{u}_i = N(\mu_{u_i}, \sigma_{u_i})$ denotes the D-dimensional Gaussian distribution for each user; the mean vectors μ_{u_i} captures the position of entities in the Wasserstein embedding space; the variance vector σ_{u_j} preserves the uncertainty of the entities; and $W(\cdot)$ is the 2nd Wasserstein distance, which can estimate two distributions' similarity, besides that the calculation speed is faster than general-formed Wasserstein distance. According to [20], we can calculate the Wasserstein distance between the two Gaussian distributions as follows:

$$\begin{aligned} W(N(\mu_1, \sigma_1), N(\mu_2, \sigma_2)) &= W(N(\mu_1, \sigma_1), N(\mu_2, \sigma_2))_2 \\ &= \|\mu_1 - \mu_2\|^2 + \mathrm{Tr}\left(\sigma_1 + \sigma_2 - 2\left(\sigma_1^{\frac{1}{2}} \sigma_2 \sigma_1^{\frac{1}{2}} \right)^{\frac{1}{2}} \right). \end{aligned} \tag{22}$$

Now, per (22), we can get the objective functions (21), and then penalize ranking errors by the energy of the pairs, which makes the similarity between positive cases greater than the similarity of negative examples.

E. Wasserstein Distance for Generative Models

As stated before, WaPOIR is a variant of the Bayesian recommendation model using the user feature matrix $\mathbf{X}$ and POI feature matrix $\mathbf{Y}$ as model inputs. It therefore has many properties of variational CF methods, such as stable training and nice latent manifold structure, while generating better latent representations about users and POIs.

Let $P_{\mathbf{X}}$ and $P_{\mathbf{Y}}$ denote the input data distributions, satisfying $\mathbf{x}_i \sim P_{\mathbf{X}}$ and $\mathbf{y}_j \sim P_{\mathbf{Y}}$. Like VAE, our model includes reconstruction losses and the regularization. In our work, we reconstruct the input data $\mathbf{X}$ and $\mathbf{Y}$, aiming at preserving the neighborhood structure for users and POIs, while the regularizer fosters the posterior to approximate the prior during training. Due to data sparsity, we only need to consider the nonzero entries in input feature matrices $\mathbf{X}$ and $\mathbf{Y}$, e.g., we can obtain the objective function regarding users as follows:

$$\begin{aligned} \mathcal{L}_u = & \inf_{q_{\phi_u}(\mathbf{Z}_u|\mathbf{X}) \in \Re} \mathbb{E}_{P_{\mathbf{X}}} \mathbb{E}_{q_{\phi_u}(\mathbf{Z}_u|\mathbf{X})} \left[\left\| \mathbf{X} \cdot \left(\mathbf{X} - \tilde{\mathbf{X}} \right) \right\|^2 \right] \\ & + \lambda_u \cdot D_{\mathbf{Z}}\left(q_{\phi_u}(\mathbf{Z}_u|\mathbf{X}), p(\mathbf{Z}_u) \right) \end{aligned} \tag{23}$$

where $\Re$ is the set of probabilistic encoders, $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{Y}}$ are generated by generative models $p_{\theta_u}(\tilde{\mathbf{X}}|\mathbf{Z}_u)$ and $p_{\theta_v}(\tilde{\mathbf{Y}}|\mathbf{Z}_v)$, respectively. Let $\mathbf{Z}_u \in \mathbb{R}^{m \times D}$ and $\mathbf{Z}_v \in \mathbb{R}^{n \times D}$ be the latent variables we are interested in. The $q_{\phi_u}(\mathbf{Z}_u|\mathbf{X})$ and $q_{\phi_v}(\mathbf{Z}_v|\mathbf{Y})$ are posterior distributions related to users and POIs, respectively. $D_{\mathbf{Z}}(\cdot)$ is an arbitrary divergence between posterior distribution $q_\phi(\mathbf{Z}_*|\star)$ and prior distribution $p(\mathbf{Z}_*)$ ($*$ means u or v, $\star$ denotes $\mathbf{X}$ or $\mathbf{Y}$). Similar to VAEs [18], we need to find a similarity metric like KL-divergence to measure the distance between $q_\phi(\mathbf{Z}_*|\star)$ and $p(\mathbf{Z}_*)$. Inspired by [26], we use a different penalty to calculate the distance, that is $D_{\mathbf{Z}}(q_\phi(\mathbf{Z}_*|\star), p(\mathbf{Z}_*)) = D_{\mathbb{JS}}(q_\phi(\mathbf{Z}_*|\star), p(\mathbf{Z}_*))$. In each iteration, we sample a batch of latent representations from the user (or POI) prior distribution $p(\mathbf{Z}_u)$ (or $p(\mathbf{Z}_v)$) as "true" data $\{\mathbf{z}_{u_1}, \ldots, \mathbf{z}_{u_M}\}$ (or $\{\mathbf{z}_{v_1}, \ldots, \mathbf{z}_{v_N}\}$), and sample "fake" data $\{\tilde{\mathbf{z}}_{u_1}, \ldots, \tilde{\mathbf{z}}_{u_M}\}$ (or $\{\tilde{\mathbf{z}}_{v_1}, \ldots, \tilde{\mathbf{z}}_{v_N}\}$) from the user (or POI) posterior distribution $q_\phi(\mathbf{Z}_u|X_M)$ (or $q_\phi(\mathbf{Z}_v|Y_N)$)—M and N are the related batch-size about users and POIs. Then, we use adversarial training to estimate it

$$\begin{aligned} & D_{\mathbf{Z}}\left(q_{\phi_u}(\mathbf{Z}_u|\mathbf{X}), p(\mathbf{Z}_u) \right) \\ & \quad = \frac{\lambda_u}{M} \sum_{i=1}^{M} \log D_{\lambda_u}\left(\mathbf{z}_{u_i}\right) + \log\left(1 - D_{\lambda_u}\left(\tilde{\mathbf{z}}_{u_i}\right)\right) \end{aligned} \tag{24}$$

where λ_u is the regularization coefficient. At the same time, we can update the encoder parameters and decoder parameters of users as follows:

$$\phi_u, \theta_u \sim \frac{1}{M} \sum_{i=1}^{M} \left\| \mathbf{x}_i - \tilde{\mathbf{x}}_i \right\|^2 - \lambda_u \cdot \log D_{\lambda_u}\left(\tilde{\mathbf{z}}_{u_i}\right). \tag{25}$$

This idea is very similar to generative adversarial net (GAN) [37]. However, in our model, we set the prior distribution to a Gaussian, which makes our optimization task easier than matching an uncertain intricate, and possibly multimodal distributions as usually done in GANs. Similar to VAEs, we use a deep neural network to parameterize the encoder and decoder of our model. However, the encoder is different from VAEs, which allows the input vector to be mapped directly to a latent variable.

F. Implementation Details

1) Optimization: In LBSNs, there are many users and POIs. If we calculate all valid triples for $\mathcal{D}_u$ and $\mathcal{D}_v$ in each training epoch, then the optimization objective function (21) is computationally expensive. To address this problem, we need to sample triplets from $\mathcal{D}_u$ uniformly. In each epoch, for each user, we take one triplet from $\mathcal{D}_u$; for each POI, we take one triplet from $\mathcal{D}_v$. In each iteration, we sample m triplets for users and n triplets for POIs to compute the gradient, and replace $\sum_{(u_i,u_j,u_k)\in\mathcal{D}_u}$ with $\mathbb{E}_{(u_i,u_j,u_k)\sim\mathcal{D}_u}$ in (21). Thus, we modify the loss function of (21) as follows:

$$\mathcal{L}^u_{\text{pair}} = \mathbb{E}_{(u_i,u_j,u_k)\sim\mathcal{D}_u}\left[W(\mathbf{u}_i,\mathbf{u}_j) + \exp\left(-W(\mathbf{u}_j,\mathbf{u}_k)\right)\right]. \quad (26)$$

For (23), we can obtain an unbiased estimate by sampling from $q_\phi(\mathbf{Z}_*|\star)$, which is a noncontinuous operation and cannot trivially take gradients with respect to ϕ through this sampling process. In this case, it is difficult to perform stochastic gradient ascent to optimize these loss functions. Inspired by [18], [32], we use a reparameterization trick to sidestep this issue. First, we sample $\epsilon \sim N(0, I_D)$, and then reparameterize $\mathbf{Z}_* = \mu_{\phi_*}(\star) + \sigma_{\phi_*}(\star) \odot \epsilon$. By giving the input data $\star$ and ϵ, the stochasticity in the sampling process is isolated and the objective functions (23) is deterministic and continuous in the parameters of encoder and decoder, which can be optimized using stochastic gradient descent.

In this way, we jointly optimize the model by using stochastic gradient descent, the loss function is presented as follows:

$$\mathcal{L} = \mathcal{L}_u + \mathcal{L}_v + \rho\mathcal{L}^u_{\text{pair}} + (1-\rho)\mathcal{L}^v_{\text{pair}} \quad (27)$$

where ρ is a hyper-parameter balancing the weight of social and geographical influence.

2) Prediction: We now describe how we make POI recommendation for users given a trained generative network. After training, we obtain all parameters and weights of the encoder and decoder. Most importantly, we have latent representations of users and POIs. Therefore, the prediction distribution $p(\mathcal{P}_{ij})$ can be made as follows:

$$p\left(\mathcal{P}_{ij}|\mathbf{x}_i,\mathbf{y}_j\right) = \int p\left(\mathcal{P}_{ij}|\mathbf{u}_i,\mathbf{v}_j\right)q(\mathbf{u}_i|\mathbf{x}_i)q\left(\mathbf{v}_j|\mathbf{y}_j\right)d\mathbf{u}_i d\mathbf{v}_j. \quad (28)$$

We can use the expectation of $p(\mathcal{P}_{ij}|\mathbf{u}_i,\mathbf{v}_j)$ as the predictive value for user u_i and POI v_j. Then, (28) can be rewritten as follows:

$$\mathbb{E}\left(\mathcal{P}_{ij}|\mathbf{x}_i,\mathbf{y}_j\right) = \left(\mathbb{E}\left[\mathbf{z}_{\mathbf{u_i}}|\mathbf{x_i}\right] + \mathbb{E}\left[\mathbf{u}^{\mathbf{c}}_{\mathbf{i}}|\mathbf{x_i}\right]\right)^{\mathsf{T}}\left(\mathbb{E}\left[\mathbf{z}_{\mathbf{v_j}}|\mathbf{y_j}\right] + \mathbb{E}\left[\mathbf{v}^{\mathbf{c}}_{\mathbf{j}}|\mathbf{y_j}\right]\right)$$

where $\mathbb{E}[\mathbf{z}_{\mathbf{u_i}}|\mathbf{x_i}] = \mu_{\phi_u}(\mathbf{x}_i)$ and $\mathbb{E}[\mathbf{z}_{\mathbf{v_j}}|\mathbf{y_j}] = \mu_{\phi_v}(\mathbf{y}_j)$ are produced by the user inference network and POI inference network, respectively.

G. Complexity Analysis

In WaPOIR, there are three graph-structured data, i.e., user' check-in matrix $\mathbf{C}$, social relations graph G_u, and geographical graph among POIs G_p. For each graph, the time complexity of calculating gradients and updating parameters in each iteration is $O(E \cdot (|\mathcal{N}| \cdot L + L \cdot D + D))$, where E is the number of the edges, $|\mathcal{N}|$ is the average degree of the entities, L is the size of hidden layers, and D is the dimensionality of representation. In practice, we only need to reconstruct the nonzero elements in each graph, which means the complexity of the encoder and decoder networks is $O(|\mathcal{N}| \cdot L)$. Compared to the variational recommender systems [15], [23], [24], the main overhead is to compute the Wasserstein distance between the node distributions, which only requires $O(D)$ extra complexity. As shown in the experiments, we empirically found that WaPOIR can converge to the best performance within a few iterations, e.g., 40–50 epochs dependent on the scale of the datasets.

TABLE II
STATISTICS OF DATASETS USED IN EXPERIMENTS

Dataset	#Users	#POIs	#User-POI pairs	Sparsity
Gowalla	18,737	32,510	1,278,274	99.87%
Foursquare	24,941	28,593	1,196,248	99.90%
Yelp	30,887	18,995	860,888	99.86%

IV. EXPERIMENTS

Since the main goal of our work is to develop a novel VAE-based learning method for personalized POI recommendation, we conducted a series of quantitative and qualitative evaluations that can demonstrate the benefits of our contribution. Specifically, we try to answer the following questions.

1) *Q1:* Can WaPOIR improve the performance compared with the state-of-the-art recommendation models?
2) *Q2:* Does the uncertainty-aware embedding method capture meaningful user and POI representations?
3) *Q3:* Do the auxiliary information-specific priors benefit the recommendation performance?
4) *Q4:* How do the key hyper-parameters affect the recommendation performance?
5) *Q5:* Is the proposed model efficient?

A. Experimental Settings

We conducted our experiments on three publicly available LBSNs datasets, including Gowalla,[1] Foursquare,[2] and Yelp.[3] Following [38], we filter out those users with fewer than 20 check-in POIs and those POIs with fewer than 20 visitors for the Gowalla dataset. In Foursquare and Yelp, the users whose check-in are fewer than 10 and the POIs with fewer than ten visitors are discarded.

The datasets after preprocessing are described in Table II. We partition each dataset into training, tuning, and testing sets. For each user, the earlier 70% check-ins are used as the training data, the recent 20% data for testing, and the remaining 10% for model validation.

1) Baselines: We empirically evaluate the proposed model for POI recommendation by comparing our method with state-of-the-art approaches.

1) *USG [28]:* Is a CF-based method combining user preference, social influence, and geographical influence.
2) *IRenMF [39]:* Incorporates the features of neighboring POIs into a weighted matrix factorization model for POI recommendation.
3) *GeoMF [4]:* Leverages user's interest-based preference and the geographical preference, and augments latent

[1]http://snap.stanford.edu/data/loc-gowalla.html
[2]https://sites.google.com/site/yangdingqi/home/foursquare-dataset
[3]https://www.yelp.com/dataset

TABLE III
SUMMARY OF SIDE INFORMATION IN OUR EVALUATION

Information	SAE-NAD	USG	APOIR	IRenMF	GeoMF	RankGeoFM	PACE	CVAE	CLVAE	Mult-VAE	WaPOIR
Geographical	✓	✓	✓	✓	✓	✓	✓	✓	×	×	✓
Social	×	✓	✓	×	×	×	✓	×	✓	×	✓
User Preference	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

factors with user activity areas and location influential areas, respectively.

4) *RankGeoFM [6]:* Is a geographical factorization method considering spatial-temporal factors and users' preference rankings for POIs.
5) *PACE [40]:* Is an embedding-based method that jointly learns the representations of users and POIs.
6) *Self-Attentive Encoder (SAE)-Neighbor-Aware Decoder (NAD) [21]:* Consists of an SAE and an NAD with geographical influence, which adaptively differentiates the user preference degrees in multiple aspects, by adopting a multidimensional attention mechanism.
7) *APOIR [17]:* Is a state-of-the-art POI recommendation model, which consists of two parts, a recommender and a discriminator, which are jointly trained for learning user preference by playing a minimax game considering geographical influence and social relations as rewards in a reinforcement learning manner.
8) *GPR [41]:* Is a graph neural network-based POI recommendation model, which learns the latent representations with graph convolutional networks and estimates user preferences with the influences of POIs.

In addition, we compared our approach against the following VAE-based recommendation models.

1) *CVAE [15]:* Is the first collaborative VAE-based item recommendation method that maximizes the lower bound of the conditional log-likelihood of target variables incorporating content information from items into MF.
2) *CLVAE [23]:* Is a conditional ladder VAE [42] based recommendation method which extends the CVAE with hierarchical VAE structure and uses a generative adversarial network to extract the low-dimensional representations influenced by the social links.
3) *Mult-VAE [24]:* Is very similar to CVAE except that it uses multinomial conditional likelihood as the prior and that it only incorporates the user preference information for recommendation.

We summarize the comparison of the side information used in each baseline in Table III, where "✓" means that a dataset contains social or geographical information, and "×" denotes the opposite. Since there are no social relationships in the original Foursquare data, we constructed one based on the cosine similarity between users. According to [43], we treat user u_i and user u_j as friends when $U_{ij} > 0.5$, [cf. (5)]. For those general models, we use POIs' geographical information to represent content information of items.

2) Metrics: For validity purposes, we selected four common metrics to evaluate the performance comparison for POI recommendation, including Pre@K (Precision), Rec@K (Recall), nDCG@K (normalized discounted cumulative gain), and MAP@K (mean average precision) following previous works [38], [40]. They show different perspectives of the performance evaluation: while precision and recall measure the number of correct recommendations, nDCG and MAP consider the rank of recommended POIs.

3) Implementation Details: We implemented all methods with TensorFlow on a machine with one NVIDIA GeForce GTX 1080Ti, Intel Xeon CPU E5-2650 with 2.30 GHz, and 128-GB RAM. For a fair comparison, we set the latent factors $\mathbf{z}_{u_i}$ and $\mathbf{z}_{v_j}$ for each user and each POI in the low-dimensional space of the same dimension—128 (unless otherwise specified). All deep learning models are trained with the Adam optimizer. The regularization coefficients λ_u and λ_v are set to 6 and 8, respectively; while the mini-batch size and learning rate were tuned according to performance in validation sets. According to the statistics of [38], we realize that 50% of users' transition distances are less than 2 km in Gowalla, 3 km in Foursquare, and 6 km in Yelp. Therefore, the geographical distance parameter d_t is accordingly set to 2, 3, and 6 km for Gowalla, Foursquare, and Yelp, respectively. For the social influence, we keep the similarity when the value is above 0.1 for all three datasets. We set $\gamma = 0.2$ and $\beta = 0.5$ on both Gowalla and Yelp. Since there is no social relation in Foursquare data, we do not need to explicitly calculate F_{ij}, and then set the $\alpha = 0$. In our experiments, we adjust parameter ρ to be in the range [0, 1]. In addition, the geographical correlation level is set to 60.

B. Overall Performance (Q1)

Table IV summarizes the performance of WaPOIR in comparison to the existing state-of-the-art models for top-K POI recommendation on three datasets, where the best performance is shown in bold and a paired t-test was performed for statistical significance of the results ($p < 0.005$). Based on the results, we have the following important observations.

(O1): Our proposed model WaPOIR performs the best on the three datasets across all evaluation metrics. Overall, our model outperforms the baselines by a large margin. Taking the Yelp data for example, WaPOIR achieves 5.13% on Precision@10, 4.39% on Recall@10, 5.24% on nDCG@10, 5.75% on MAP@10 over the second best model GPR, which confirms the superiority of our model. Among baselines, GPR usually outperforms other non-VAE-based methods (e.g., PACE, APOIR, SAE-NAD, USG, etc.), which indicates its effectiveness of capturing user preference through modeling POI interactions with graphs and estimating geographical and social influences using graph neural networks.

(O2): The MF-based methods (e.g., IRenMF, GeoMF, and RankGeoFM) exhibit better recommendation quality

TABLE IV
RECOMMENDATION PERFORMANCE COMPARISONS AMONG DIFFERENT ALGORITHMS ON THREE LBSN DATASETS. A PAIRED t-TEST IS PERFORMED AND * INDICATES A STATISTICAL SIGNIFICANCE $p < 0.005$ AS COMPARED TO THE BEST BASELINE METHOD

Dataset	Model	Pre@5	Pre@10	Pre@20	Rec@5	Rec@10	Rec@20	nDCG@5	nDCG@10	nDCG@20	MAP@5	MAP@10	MAP@20
Gowalla	USG	0.0615	0.0487	0.0383	0.0551	0.0665	0.1152	0.0623	0.0517	0.0443	0.0357	0.0329	0.0345
	IRenMF	0.0712	0.0583	0.0425	0.0599	0.0745	0.1295	0.0734	0.0592	0.0511	0.0453	0.0372	0.0387
	GeoMF	0.0691	0.0521	0.0403	0.0573	0.0683	0.1195	0.0702	0.0573	0.0503	0.0412	0.0334	0.0375
	RankGeoFM	0.0734	0.0597	0.0443	0.0601	0.0753	0.1373	0.0758	0.0612	0.0527	0.0462	0.0385	0.0423
	SAE-NAD	0.0592	0.0493	0.0392	0.0412	0.0667	0.1203	0.0594	0.0503	0.0437	0.0338	0.0315	0.0325
	CVAE	0.0751	0.0612	0.0487	0.0645	0.0895	0.1412	0.0794	0.0635	0.0589	0.0487	0.0423	0.0456
	CLVAE	0.0683	0.0573	0.0412	0.0603	0.0762	0.1383	0.0713	0.0574	0.0504	0.0392	0.0363	0.0372
	MultiVAE	0.0613	0.0502	0.0397	0.0507	0.0732	0.1365	0.0634	0.0519	0.0479	0.0383	0.0332	0.0367
	PACE	0.0753	0.0679	0.0573	0.0631	0.0872	0.1431	0.0831	0.0741	0.0647	0.0495	0.0475	0.0482
	APOIR	0.0765	0.0683	0.0582	0.0645	0.0901	0.1445	0.0852	0.0782	0.0703	0.0501	0.0493	0.0496
	GPR	0.0788	0.0687	0.0594	0.0673	0.0913	0.1452	0.0878	0.0801	0.0706	0.0516	0.0502	0.0511
	WaPOIR	**0.0805***	**0.0701***	**0.0603***	**0.0684***	**0.0927***	**0.1486***	**0.0898***	**0.0813***	**0.0712***	**0.0537***	**0.0529***	**0.0534***
Foursquare	IRenMF	0.0602	0.0498	0.0301	0.0601	0.0813	0.1352	0.0612	0.0564	0.0413	0.0382	0.0365	0.0391
	GeoMF	0.0591	0.0475	0.0295	0.0593	0.0809	0.1234	0.0603	0.0552	0.0425	0.0367	0.0358	0.0372
	RankGeoFM	0.0625	0.0511	0.0315	0.0608	0.0825	0.1387	0.0698	0.0582	0.0436	0.0453	0.0384	0.0404
	SAE-NAD	0.0543	0.0412	0.0288	0.0496	0.0807	0.1278	0.0553	0.0482	0.0358	0.0345	0.0322	0.0366
	CVAE	0.0631	0.0521	0.0326	0.0625	0.0854	0.1389	0.0673	0.0592	0.0475	0.0465	0.0361	0.0471
	CLVAE	0.0595	0.0479	0.0294	0.0582	0.0768	0.1277	0.0598	0.0549	0.0437	0.0363	0.0359	0.0371
	MultiVAE	0.0512	0.0399	0.0283	0.0492	0.0803	0.1235	0.0543	0.0477	0.0363	0.0344	0.0319	0.0355
	PACE	0.0656	0.0532	0.0375	0.0665	0.0913	0.1391	0.0703	0.0594	0.0473	0.0459	0.0393	0.0502
	APOIR	0.0675	0.0535	0.0381	0.0680	0.0927	0.1392	0.0707	0.0603	0.0479	0.0462	0.0404	0.0509
	GPR	0.0679	0.0538	0.0387	0.0692	0.0929	0.1396	0.0712	0.0606	0.0483	0.0465	0.0419	0.0512
	WaPOIR	**0.0698***	**0.0541***	**0.0402***	**0.0704***	**0.1002***	**0.1422***	**0.0728***	**0.0612***	**0.0501***	**0.0497***	**0.0471***	**0.0523***
Yelp	USG	0.0241	0.0223	0.0166	0.0312	0.0453	0.0712	0.0275	0.0241	0.0217	0.0163	0.0169	0.0189
	IRenMF	0.0268	0.0249	0.0211	0.0357	0.0481	0.0763	0.0301	0.0248	0.0223	0.0192	0.0201	0.0223
	GeoMF	0.0301	0.0253	0.0223	0.0369	0.0483	0.0784	0.0312	0.0275	0.0241	0.0201	0.0213	0.0321
	RankGeoFM	0.0312	0.0271	0.0241	0.0372	0.0489	0.0813	0.0337	0.0312	0.0261	0.0225	0.0237	0.0258
	SAE-NAD	0.0302	0.0261	0.0225	0.0373	0.0491	0.0793	0.0325	0.0301	0.0258	0.0226	0.0235	0.0252
	CVAE	0.0334	0.0293	0.0254	0.0383	0.0517	0.0825	0.0352	0.0321	0.0296	0.0253	0.0257	0.0277
	CLVAE	0.0321	0.0277	0.0251	0.0379	0.0512	0.0819	0.0341	0.0312	0.0301	0.0241	0.0252	0.0273
	MultiVAE	0.0275	0.0251	0.0218	0.0356	0.0485	0.0765	0.0308	0.0271	0.0242	0.0196	0.0203	0.0225
	PACE	0.0342	0.0295	0.0272	0.0388	0.0545	0.0843	0.0362	0.0347	0.0308	0.0247	0.0258	0.0285
	APOIR	0.0372	0.0298	0.0281	0.0392	0.0556	0.0886	0.0383	0.0357	0.0312	0.0267	0.0274	0.0291
	GPR	0.0386	0.0312	0.0286	0.0398	0.0570	0.0903	0.0388	0.0365	0.0326	0.0272	0.0286	0.0299
	WaPOIR	**0.0398***	**0.0328***	**0.0297***	**0.0406***	**0.0595***	**0.0927***	**0.0417***	**0.0386***	**0.0342***	**0.0286***	**0.0301***	**0.0315***

than USG. One possible reason is that although USG models user preference, geographical, and social influence simultaneously for POI recommendation, it still cannot model user's check-ins as implicit feedback and learn latent features underpinning the complex interactions between the users and the POIs. However, MF-based models only learn linear interactions among users and POIs, which cannot capture the nonlinear relations that are important for estimating user preference.

(O3): One of the latest methods, SAE-NAD, does not perform very well. The main reason is that SAE-NAD is an autoencoder-based recommendation model with an attention mechanism, which cannot simulate data in chronological order. In addition, the recommendation performance is constrained by the POI distance as it only considers close POIs. Besides, SAE-NAD ignores the social influence and high-order interactions between users and POIs, while failing to capture the uncertainty of the users' preference and POI representations, which has been proved to be effective in recommending user-interested items.

(O4): PACE is another baseline performing well, which indicates its effectiveness in learning user-POI interactions through modeling the implicit feedback data and context knowledge by neural embedding. However, PACE relies on sparse historical data to learn user and POI representations, which may not capture high-order interactions. APOIR, in contrast, considers geographical and social influences in an adversarial learning manner, which allows it to capture nonlinear and high-order interactions through estimating the awards with reinforcement learning. Nevertheless, APOIR requires geographically closed POIs and social friends as positive labels, limiting its ability to capture meaningful signals from implicit and negative feedback. As WaPOIR, GPR is a graph-based POI recommendation model, generally showing the best performance among baseline approaches. However, it only exploits the POI-POI graph and user-POI interactions, failing to consider the social influence. Besides, GPR relies on graph neural networks to capture the relations, which is known to suffer from the oversmoothing problem when stacking deeper layers. In other words, it can only model shallow interactions and relations when aggregating information from neighbor nodes. In contrast, our WaPOIR learns the embeddings in the Wasserstein space while being able to model the uncertainty of representations for both users and POIs.

(O5): Finally, VAE-based models, such as CVAE, CLVAE, and Mult-VAE perform relatively well, although they are not specifically tailored for POI recommendations. These results demonstrate the potential effectiveness in modeling the nonlinearity and uncertainty of user-POI interactions with a Bayesian framework. Among these methods, Mult-VAE is a nonlinear

TABLE V
PARAMETER SETTINGS FOR CONSIDERING DIFFERENT AUXILIARY INFORMATION

	γ	β	ρ
POIR-Geo	1	-	0
POIR-Soc	0.2	0.5	1
WaPOIR	0.2	0.5	0.5

probabilistic model and applies a deep neural structure for implicit feedback, but it still does not achieve better results than RankGeoFM, primarily because it does not leverage the auxiliary information important for POI recommendation. For other Bayesian models, CLVAE benefits from learning richer flexibility of variational distributions with hierarchical VAE, and it also incorporates the social information to improve the performance. CVAE is a strong baseline that performs well in most cases. The reason is that CVAE uses both geographical influence and collaborative information concerning the users and POIs.

C. Visualization Analysis (Q2)

Given that our goal is to learn better latent representations for users and POIs that could improve the recommendation performance, an important/direct indicator to show the validity of our model is visualization. Moreover, disentangled representation is generally considered to contain interpretable semantic information and reflect separate factors of variation in the data. Therefore, we visualize the learned latent representations of users and POIs. Specifically, we first extract the latent low-dimensional (128 dimensions) representations of each user and each POI obtained from the model, and then map them in 2-D using the t-SNE algorithm [44].

The visualization of the latent space is plotted in Fig. 4, where we can clearly observe the clustering phenomena learned by WaPOIR. The results also demonstrate that the uncertainty-aware embedding and Wasserstein distance measure used in our model can extract meaningful representations that could help discriminate the user preference over POIs. On the other hand, we can observe that the entities are better clustered in Gowalla and Foursquare while being more entangled in Yelp data. This result suggests that the better the clustering effect, the more accurate recommendation performance the model can achieve, e.g., all models perform better on Gowalla and Foursquare than on Yelp (cf. Table IV).

D. Ablation Study (Q3)

The social and geographical information show a significant impact on the effectiveness of POI recommendation. To verify the performance of each aspect, we set the values of three hyper-parameters γ, β, and ρ as shown in Table V (for Yelp, the ρ is set to 0.7 in WaPOIR), which indicates that we can vary the parameters to understand the roles of social influence and geographical influence played in POI recommendation. Here, we denote the model with user preference and geographical information as *POIR-Geo*, the model with user preference and social information as *POIR-Soc*, *WaPOIR* is our model considering all of the information contexts.

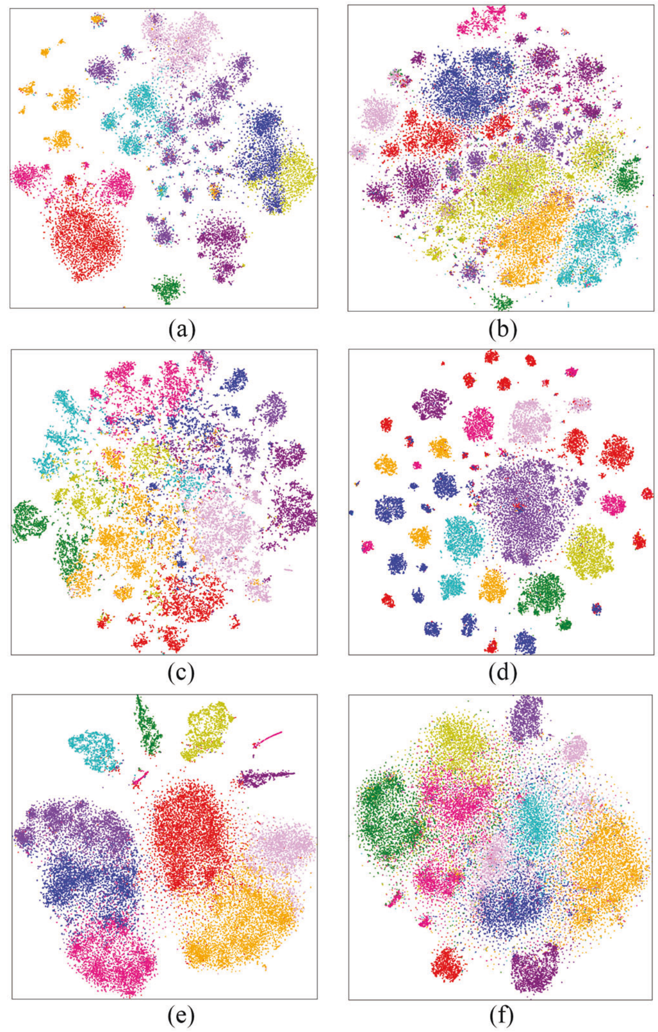


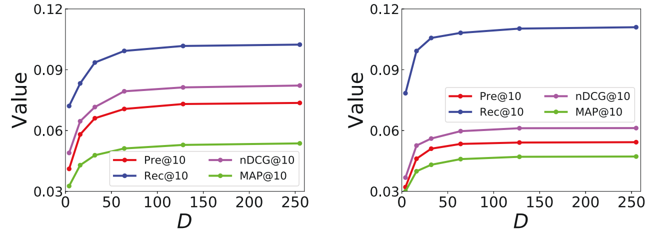
Fig. 4. Visualization of the user and POI representations. (a) User on Gowalla. (b) POI on Gowalla. (c) User on Foursquare. (d) POI on Foursquare. (e) User on Yelp. (f) POI on Yelp.

TABLE VI
PERFORMANCE COMPARISON AMONG DIFFERENT AUXILIARY INFORMATION FOR POI RECOMMENDATION ON GOWALLA, FOURSQUARE, AND YELP

Gowalla	**Pre@10**	**Rec@10**	**MAP@10**	**nDCG@10**
POIR-Soc	0.0682	0.0904	0.0495	0.0772
POIR-Geo	0.0696	0.0912	0.0504	0.0796
WaPOIR	0.0701	0.0927	0.0529	0.0813
Foursquare	**P@10**	**R@10**	**MAP@10**	**nDCG@10**
POIR-Soc	0.0524	0.0925	0.0436	0.0569
POIR-Geo	0.0537	0.0936	0.0458	0.0586
WaPOIR	0.0541	0.1002	0.0471	0.0612
Yelp	**P@10**	**R@10**	**MAP@10**	**nDCG@10**
POIR-Soc	0.0302	0.0569	0.0285	0.0371
POIR-Geo	0.0314	0.0582	0.0267	0.0359
WaPOIR	0.0328	0.0595	0.0301	0.0386

The results shown in Table VI exhibit the effectiveness of different side information. We can observe that.

(O1): *WaPOIR* achieves better performance than *POIR-Geo* and *POIR-Soc* which means that combining two important information for POI recommendation can effectively improve the performance of the model.



(a) (b)

Fig. 5. Impact of D in three datasets. (a) Gowalla. (b) Foursquare.

(O2): *POIR-Geo* shows the second best on Gowalla and Foursquare which indicates that geographical influence has a more significant impact than social influence on the effectiveness of POI recommendations.

(O3): *POIR-Geo* and *POIR-Soc* demonstrate the opposite performance on the Yelp. One explanation could be that geographical influence has less effectiveness than social influence for POI recommendation, since the distance between the Yelp users' check-in POIs is relatively larger making it difficult for modeling users' geographical preferences.

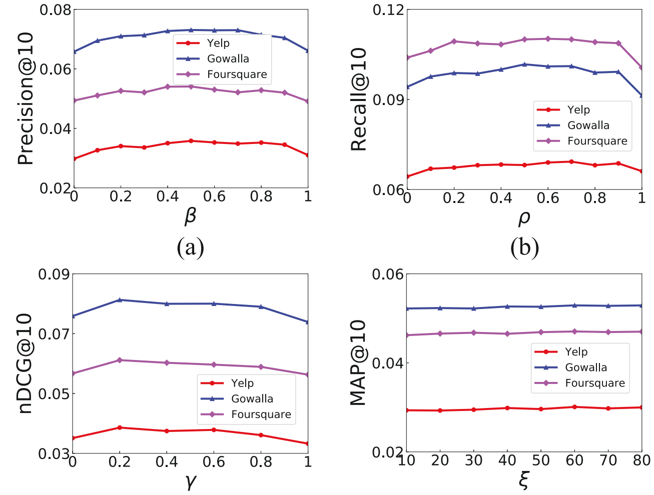
E. Sensitivity of Parameters (Q4)

There are several important hyper-parameters that need to be carefully tuned for WaPOIR. We optimized our model by varying these parameters—the dimension of latent representations D, the hyper-parameter for balancing the weight of social and geographical influence ρ, the user similarity parameters γ and β, and the geographical correlation level ξ.

As shown in Fig. 5, the dimensionality of embedding D is tuned with six different values: $\{4, 16, 32, 64, 128, 256\}$. The performance increases initially as the value of D on four metrics in three datasets. However, as it continues to increase ($D \geq 128$), the performance tends to be stable which means that most of the useful information about users and POIs have already been encoded into the latent vectors. A larger dimension consumes more computing resources, and it has a less improvement on performance. Therefore, we fix the embedding size to 128 for all datasets.

Fig. 6 shows the effect of the four different parameters on the results of the three data sets. The variation of ξ is shown in Fig. 6(d). From the figure, we can see that when $\xi > 10$ the model is insensitive to this hyper-parameter. In our experiments, we use $\xi = 60$. Fig. 6(a) and (c) illustrate the impact of user similarity parameters on nDCG and Precision for the three datasets. As can be seen, the best performances are achieved when $\gamma = 0.2$ and $\beta = 0.5$. Fig. 6(b) illustrates the impact of ρ on Recall for the three datasets.

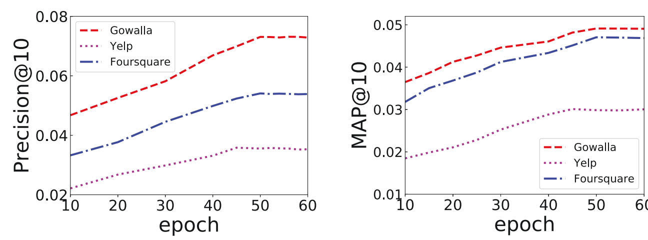
We note that the results are consistent with Table VI—i.e., the performance of *POIR-Geo* ($\rho = 0$) and *POIR-Soc* ($\rho = 1$) are worse than WaPOIR with other parameters fixed. It demonstrates that both social influence and geographical influence are important for POI recommendation. More specifically, we observe that our method is not sensitive to the choice of ρ. For the best performance, we choose $\rho = 0.5$, $\rho = 0.5$, and $\rho = 0.7$ for Gowalla, Foursquare, and Yelp, respectively.



(a) (b)

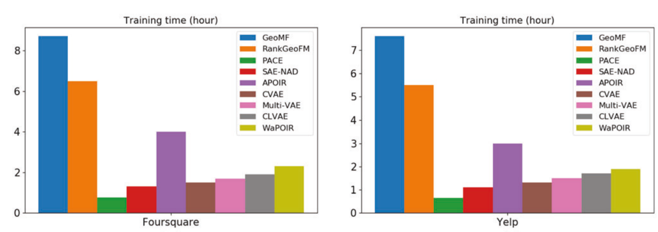
(c) (d)

Fig. 6. Results of parameter sensitivity in three datasets. (a) Precision. (b) Recall. (c) nDCG. (d) MAP.



(a) (b)

Fig. 7. Convergence of the training process. (a) Precision. (b) MAP.



(a) (b)

Fig. 8. Training time comparison. (a) Foursquare. (b) Yelp.

F. Training Process and Efficiency Comparison (Q5)

The advantages of WaPOIR stem from its capability to efficiently learn users' and POIs' latent representations. Fig. 7 plots the training process on Gowalla, Foursquare, and Yelp. We observe that WaPOIR can converge within a few epochs on three datasets, e.g., around 50 epochs on Gowalla and Foursquare datasets and 45 epochs on Yelp dataset, which demonstrates the efficiency of our pairwise similarity measure and ranking loss optimization (cf. Section III-D).

Fig. 8 compares the training time of the POI recommendation algorithms. Deep learning-based approaches are significantly faster than MF-based approaches (e.g., GeoMF and RankGeoFM) which usually need more time to estimate user preference. APOIR is computationally intensive among deep learning methods because GAN-style representation learning requires more time to explore the feature space. The training time of WaPOIR is very close to the previous VAE-based approaches (i.e., SAE-NAD, CVAE, and Multi-VAE), except that our model requires slightly more time to estimate the Wasserstein distances. In other words, WaPOIR achieves a good balance between expensive GAN-based and VAE-based approaches while consistently outperforming previous generative recommendation models. These results demonstrate that our method can be scaled to large-scale LBSN datasets.

V. Related Work

A. POI Recommendation

With the expeditious growth of the geographical data, personalized POI recommendation has been widely studied and various machine learning techniques have been introduced into POI recommendation systems. Many proposed methods incorporate different side information related to users and POIs, such as check-ins, comments, and social correlations, into POI recommendation to enhance the recommendation accuracy in a CF setting [4], [6], [45], [46], [47]. For example, early works [28], [39], [48] assumed that people share their check-in activities among friends and therefore recommended POIs by exploring the preferences of social friends. Recently, some efforts have been conducted in the community to address the reliability [49], [50], cold start, and data sparsity issues [51] that could benefit the POI recommender systems.

B. Deep Generative Recommendation

Deep recommendation models have shown great success in capturing meaningful data representations and achieving advanced recommendation performance. However, desirable, accuracy improvement may be insufficient for a modern recommender system [12]. Therefore, researchers begin to focus on providing holistic recommendation experience to users, such as diversity, fairness, and interpretability. On the other hand, deep generative models have shown promising improvements in recommendation systems by learning effective representations of network nodes. The representative deep generative models are GANs [37], VAEs [18], and flow-based generative networks [52].

GAN has been widely used in information fusion and recommendation systems to improve the model robustness via perturbation with adversarial samples [53] or optimizing model parameter inference [54] with adversarial learning. IRGAN [29] is a representative work, which leverages a game theoretical minimax game to iteratively optimize discriminator and generator, simultaneously. A most recent work APOIR [17] leverages different side information (e.g., geographical and social influence) into the rewards in a reinforcement learning manner and adopts a generative framework for training. VAE is a nonlinear probabilistic model of traditional neural autoencoders. CVAE [27] is one of the first Bayesian generative models that considered both user ratings and auxiliary features. It jointly performs deep representation learning for the side information and CF for the rating (feedback) matrix. Later, Lee et al. [23] proposed a VAE-based model that augmented CF with ladder network [42], and then leveraged GAN to exact the low-dimensional representations influenced by the auxiliary information. Meantime, Liang et al. [24] introduced a multinomial conditional likelihood-based VAE framework. It can address the problem of under-fitting when modeling large, sparse, high-dimensional data [55]. Recently, there is a growing interest in improving variational CF with flow-based models [56], [57], which could alleviate the inaccurate posterior estimation problem in previous VAE-based recommender systems.

Though our work is inspired by the recent progress of deep generative models in the fields of recommendation systems, we have several key differences compared to existing approaches. First, we combine several available side information as input features to reduce the data sparsity issue and use a *sampling trick* described in Section III-F to address the time complexity of calculating gradients and updating parameters. Second, compared to conventional CF-based POI recommender systems, we model the problem within a probabilistic recommendation setting which allows our model for Bayesian inference and capturing nonlinear and complicated user-POI interactions. Third, in contrast to the traditional VAE-based methods [23], [24], [27] which use the KL divergence as the distribution measure, our approach learns representations in the Wasserstein space, enabling our model to better reflect the uncertainties of POI and user representation, and preserve the transitivity of representations among users and POIs.

VI. Conclusion

In this article, we proposed a method WaPOIR. It can model the uncertainties of user and POI representation and address the issues related to biased inference, data sparsity, and cold start by leveraging the Bayesian inference for probabilistic recommendation and learning the latent representations as Gaussian distributions in the Wasserstein Space. WaPOIR is among the first method that combines collaborative VAE and Wasserstein distance for POI recommendation. Further, WaPOIR is a generative model capable of alleviating the agnostic posterior estimation problem inherent in existing Bayesian recommendation methods.

One of our future directions is to find a suitable way of learning priors for better capturing the real distributions in the CF settings. Meanwhile, the auxiliary information explored in this work can provide more interaction signals by extracting the user-item interactions using graph neural networks. Therefore, incorporating the uncertain user/item embedding into the graph neural networks-based recommender system is worth investigating. In this work, we embedded users/items into Gaussians, which may not fully reflect the real distributions. In this vein, how to extend the present work to model the tractable data distribution using more expressive generative models such as normalizing flows is an interesting direction for improving the performance of Bayesian recommendation.

References

[1] N. Lim, B. Hooi, S. Ng, Y. L. Goh, R. Weng, and R. Tan, "Hierarchical multi-task graph recurrent network for next POI recommendation," in *Proc. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2022, pp. 1133–1143.

[2] X. Wang, G. Sun, X. Fang, J. Yang, and S. Wang, "Modeling spatio-temporal neighbourhood for personalized point-of-interest recommendation," in *Proc. Int. Joint Conf. Artif. Intell.*, 2022, pp. 3530–3536.

[3] H. Zhang, W. Ni, X. Li, and Y. Yang, "Modeling the heterogeneous duration of user interest in time-dependent recommendation: A hidden semi-Markov approach," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 48, no. 2, pp. 177–194, Feb. 2018.

[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 *Proc. ACM SIGKDD Int. Conf. Knowl. Disc. Data Min.*, 2014, pp. 831–840.

[5] C. Chudzicki, D. E. Pritchard, and Z. Chen, "GeoSoCa: Exploiting geographical, social and categorical correlations for point-of-interest recommendations," in *Proc. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2015, pp. 443–452.

[6] 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 *Proc. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2015, pp. 433–442.

[7] Y. Zhang, C. Yin, Q. Wu, Q. He, and H. Zhu, "Location-aware deep collaborative filtering for service recommendation," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 51, no. 6, pp. 3796–3807, Jun. 2021.

[8] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," *J. Comput.*, vol. 48, no. 8, pp. 30–37, 2009.

[9] L. Wu, P. Sun, R. Hong, Y. Ge, and M. Wang, "Collaborative neural social recommendation," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 51, no. 1, pp. 464–476, Jan. 2021.

[10] D. Wu, X. Luo, M. Shang, Y. He, G. Wang, and M. Zhou, "A deep latent factor model for high-dimensional and sparse matrices in recommender systems," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 51, no. 7, pp. 4285–4296, Jul. 2021.

[11] J. He, X. Li, and L. Liao, "Category-aware next point-of-interest recommendation via listwise Bayesian personalized ranking," in *Proc. Int. Joint Conf. Artif. Intell.*, 2017, pp. 1837–1843.

[12] S. Zhang, L. Yao, A. Sun, and Y. Tay, "Deep learning based recommender system: A survey and new perspectives," *ACM Comput. Surveys*, vol. 52, no. 1, pp. 1–38, 2019.

[13] T. Qian, B. Liu, Q. V. H. Nguyen, and H. Yin, "Spatiotemporal representation learning for translation-based POI recommendation," *ACM Trans. Inf. Syst.*, vol. 37, no. 2, pp. 1–24, Mar. 2019.

[14] B. Hu, C. Shi, W. X. Zhao, and P. S. Yu, "Leveraging meta-path based context for top-N recommendation with a neural co-attention model," in *Proc. ACM SIGKDD Int. Conf. Knowl. Disc. Data Min.*, 2018, pp. 1531–1540.

[15] X. Li and J. She, "Collaborative variational autoencoder for recommender systems," in *Proc. ACM SIGKDD Int. Conf. Knowl. Disc. Data Min.*, 2017, pp. 305–314.

[16] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian approximation: Representing model uncertainty in deep learning," in *Proc. Int. Conf. Mach. Learn.*, 2016, pp. 1050–1059.

[17] F. Zhou, R. Yin, K. Zhang, G. Trajcevski, T. Zhong, and J. Wu, "Adversarial point-of-interest recommendation," in *Proc. World Wide Web Conf.*, 2019, pp. 3462–3468.

[18] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," in *Proc. Int. Conf. Learn. Represent.*, 2014, pp. 1–14.

[19] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme, "BPR: Bayesian personalized ranking from implicit feedback," in *Proc. Conf. Uncertainty Artif. Intell.*, 2009, pp. 452–461.

[20] P. Clement and W. Desch, "An elementary proof of the triangle inequality for the Wasserstein metric," in *Proc. Amer. Math. Soc.*, 2008, pp. 333–339.

[21] C. Ma, Y. Zhang, Q. Wang, and X. Liu, "Point-of-interest recommendation: Exploiting self-attentive autoencoders with neighbor-aware influence," in *Proc. Conf. Inf. Knowl. Manag.*, 2018, pp. 697–706.

[22] M. Xie, H. Yin, H. Wang, F. Xu, W. Chen, and S. Wang, "Learning graph-based POI embedding for location-based recommendation," in *Proc. Conf. Inf. Knowl. Manag.*, 2016, pp. 15–24.

[23] W. Lee, K. Song, and I.-C. Moon, "Augmented variational autoencoders for collaborative filtering with auxiliary information," in *Proc. Conf. Inf. Knowl. Manag.*, 2017, pp. 1139–1148.

[24] D. Liang, R. G. Krishnan, M. D. Hoffman, and T. Jebara, "Variational autoencoders for collaborative filtering," in *Proc. World Wide Web Conf.*, 2018, pp. 689–698.

[25] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 214–223.

[26] I. Tolstikhin, O. Bousquet, S. Gelly, and B. Schoelkopf, "Wasserstein auto-encoders," in *Proc. Int. Conf. Learn. Represent.*, 2018, pp. 1–20.

[27] H. Li, Y. Ge, D. Lian, and H. Liu, "Learning user's intrinsic and extrinsic interests for point-of-interest recommendation: A unified approach," in *Proc. Int. Joint Conf. Artif. Intell.*, 2017, pp. 2117–2123.

[28] M. Ye, P. Yin, W.-C. Lee, and D.-L. Lee, "Exploiting geographical influence for collaborative point-of-interest recommendation," in *Proc. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2011, pp. 325–334.

[29] J. Wang et al., "IRGAN: A minimax game for unifying generative and discriminative information retrieval models," in *Proc. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2017, pp. 515–524.

[30] D. Zhu, P. Cui, D. Wang, and W. Zhu, "Deep variational network embedding in Wasserstein space," in *Proc. ACM SIGKDD Int. Conf. Knowl. Disc. Data Min.*, 2018, pp. 2594–2603.

[31] A. Mnih and R. R. Salakhutdinov, "Probabilistic matrix factorization," in *Proc. Adv. Neural Inf. Process. Syst.*, 2008, pp. 1257–1264.

[32] D. J. Rezende, S. Mohamed, and D. Wierstra, "Stochastic backpropagation and approximate inference in deep generative models," in *Proc. Int. Conf. Mach. Learn.*, 2014, pp. 1–14.

[33] D. M. Blei, A. Kucukelbir, and J. D. McAuliffe, "Variational inference: A review for statisticians," *J. Amer. Stat. Assoc.*, vol. 112, no. 518, pp. 859–877, 2017.

[34] L. Mescheder, S. Nowozin, and A. Geiger, "Adversarial variational Bayes: Unifying variational autoencoders and generative adversarial networks," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 2391–2400.

[35] D. Liben-Nowell and J. Kleinberg, "The link-prediction problem for social networks," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 58, no. 7, pp. 1019–1031, 2007.

[36] J. Tang, M. Qu, M. Wang, M. Zhang, J. Yan, and Q. Mei, "LINE: Large-scale information network embedding," in *Proc. Int. World Wide Web Conf.*, 2015, pp. 1067–1077.

[37] I. Goodfellow et al., "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.

[38] Y. Liu, T.-A. N. Pham, G. Cong, and Q. Yuan, "An experimental evaluation of point-of-interest recommendation in location-based social networks," *Proc. VLDB Endowm.*, vol. 10, no. 10, pp. 1010–1021, 2017.

[39] Y. Liu, W. Wei, A. Sun, and C. Miao, "Exploiting geographical neighborhood characteristics for location recommendation," in *Proc. Conf. Inf. Knowl. Manag.*, 2014, pp. 739–748.

[40] C. Yang, L. Bai, C. Zhang, Q. Yuan, and J. Han, "Bridging collaborative filtering and semi-supervised learning: A neural approach for POI recommendation," in *Proc. ACM SIGKDD Int. Conf. Knowl. Disc. Data Min.*, 2017, pp. 1245–1254.

[41] B. Chang, G. Jang, S. Kim, and J. Kang, "Learning graph-based geographical latent representation for point-of-interest recommendation," in *Proc. Conf. Inf. Knowl. Manag.*, 2020, pp. 135–144.

[42] C. K. Sønderby, T. Raiko, L. Maaløe, S. K. Sønderby, and O. Winther, "Ladder variational autoencoders," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 3738–3746.

[43] H. Ma, "On measuring social friend interest similarities in recommender systems," in *Proc. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2014, pp. 465–474.

[44] L. V. D. Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 11, pp. 2579–2605, 2008.

[45] S. Yang, J. Liu, and K. Zhao, "GETNext: Trajectory flow map enhanced transformer for next POI recommendation," in *Proc. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2022, pp. 1144–1153.

[46] Y. Li, T. Chen, Y. Luo, H. Yin, and Z. Huang, "Discovering collaborative signals for next POI recommendation with iterative Seq2Graph augmentation," in *Proc. Int. Joint Conf. Artif. Intell.*, 2021, pp. 1491–1497.

[47] D. Li and Z. Gong, "A deep neural network for crossing-city POI recommendations," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 8, pp. 3536–3548, Aug. 2022.

[48] H. A. Rahmani, M. Aliannejadi, S. Ahmadian, M. Baratchi, M. Afsharchi, and F. Crestani, "LGLMF: Local geographical based logistic matrix factorization model for POI recommendation," in *Proc. Asia Inf. Retrieval Soc. Conf.*, 2019, pp. 66–78.

[49] S. Ahmadian, M. Meghdadi, and M. Afsharchi, "Incorporating reliable virtual ratings into social recommendation systems," *Appl. Intell.*, vol. 48, no. 11, pp. 4448–4469, 2018.

[50] S. Ahmadian, N. Joorabloo, M. Jalili, Y. Ren, M. Meghdadi, and M. Afsharchi, "A social recommender system based on reliable implicit relationships," *Knowl.-Based Syst.*, vol. 192, Mar. 2020, Art. no. 105371.

[51] F. Tahmasebi, M. Meghdadi, S. Ahmadian, and K. Valiallahi, "A hybrid recommendation system based on profile expansion technique to alleviate cold start problem," *Multimedia Tools Appl.*, vol. 80, no. 2, pp. 2339–2354, 2021.

[52] D. Rezende and S. Mohamed, “Variational inference with normalizing flows,” in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 1530–1538.
[53] X. He, Z. He, X. Du, and T.-S. Chua, “Adversarial personalized ranking for recommendation,” in *Proc. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2018, pp. 355–364.
[54] Q. Wang, H. Yin, Z. Hu, D. Lian, H. Wang, and Z. Huang, “Neural memory streaming recommender networks with adversarial training,” in *Proc. ACM SIGKDD Int. Conf. Knowl. Disc. Data Min.*, 2018, pp. 2467–2475.
[55] R. Krishnan, D. Liang, and M. Hoffman, “On the challenges of learning with inference networks on sparse, high-dimensional data,” in *Proc. Int. Conf. Artif. Intell. Stat.*, 2018, pp. 143–151.
[56] T. Zhong, Z. Wen, F. Zhou, G. Trajcevski, and K. Zhang, “Session-based recommendation via flow-based deep generative networks and Bayesian inference,” *Neurocomputing*, vol. 391, pp. 129–141, May 2020.
[57] F. Zhou, Y. Mo, G. Trajcevski, K. Zhang, J. Wu, and T. Zhong, “Recommendation via collaborative autoregressive flows,” *Neural Netw.*, vol. 126, pp. 52–64, Jun. 2020.

Fan Zhou (Member, IEEE) received the B.S. degree in computer science from Sichuan University, Chengdu, China, in 2003, and the M.S. and Ph.D. degrees in computer science from the University of Electronic Science and Technology of China, Chengdu, in 2006 and 2012, respectively.

He is currently a Professor with the School of Information and Software Engineering, University of Electronic Science and Technology of China. His research interests include machine learning, neural networks, spatio-temporal data management, graph learning, recommender systems, and social network data mining.

Tangjiang Qian received the B.S. degree in software engineering from the University of Electronic Science and Technology of China, Chengdu, China, in 2021, where he is currently pursuing the M.S. degree.

His research interests include social computing, multimodal learning, and deep generative learning.

Yuhua Mo received the B.S. degree in software engineering from Southwest Minzu University, Chengdu, China, in 2017, and the M.S. degree in software engineering from the University of Electronic Science and Technology of China, Chengdu, in 2020.

Her research interests include recommendation systems, spatio-temporal data mining, and deep generative learning.

Zhangtao Cheng received the B.S. degree in electronic and information engineering from Southwest Jiaotong University, Chengdu, China, in 2020. He is currently pursuing the M.S. degree in software engineering with the University of Electronic Science and Technology of China, Chengdu.

His current research interests include social network data mining and knowledge discovery, spatio-temporal data mining, and recommender systems.

Chunjing Xiao received the Ph.D. degree in computer software and theory from the University of Electronic Science and Technology of China, Chengdu, China, in 2013.

He is currently an Associate Professor with the School of Computer and Information Engineering, Henan University, Kaifeng, China. He was a Visiting Scholar with the Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL, USA. His current research interests include recommender systems, anomaly detection, and Internet of Things.

Jin Wu received the B.S. degree in automatic control and the M.S. and Ph.D. degrees in computer application technology from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 1993, 1996, and 2004, respectively.

She is currently an Associate Professor with UESTC. Her research interests include machine learning, knowledge mapping, software development techniques, and process technology.

Goce Trajcevski (Member, IEEE) received the B.Sc. degree in informatics and automation from the University of Sts. Kiril i Metodij, Skopje, North Macedonia, in 1989, and the M.S. and Ph.D. degrees in computer science from the University of Illinois at Chicago, Chicago, IL, USA, in 1995 and 2002, respectively.

He is currently an Associate Professor with the Department of Electrical and Computer Engineering, Iowa State University, Ames, IA, USA. His research has been funded by the NSF, ONR, BEA, and Northrop Grumman Corporation. In addition to a book chapter and three encyclopedia chapters, he has coauthored over 140 publications in refereed conferences and journals. His main research interests are in the areas of spatio-temporal data management, uncertainty and reactive behavior management in different application settings, and incorporating multiple contexts.

Dr. Trajcevski was the General Co-Chair of the IEEE International Conference on Data Engineering 2014 and ACM SIGSPATIAL 2019, the PC Co-Chair of the ADBIS 2018 and ACM SIGSPATIAL 2016 and 2017, and has served in various roles in organizing committees in numerous conferences and workshops. He is an Associate Editor of the *ACM Transactions on Spatial Algorithms and Systems* and the *Geoinformatica* Journals.