2021 22nd IEEE International Conference on Mobile Data Management (MDM) | 978-1-6654-2845-3/20/$31.00 ©2021 IEEE | DOI: 10.1109/MDM52706.2021.00023

Meta-learning Enhanced Neural ODE for Citywide Next POI Recommendation

Haining Tan[1,2] Di Yao[1,2] Tao Huang[1,2] Baoli Wang[1,2] Quanliang Jing[1,2] Jingping Bi[1,2]
[1] *Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China*
[2]*University of Chinese Academy of Sciences, Beijing, China*
{tanhaining, yaodi, huangtao01, wangbaoli, jingquanliang, bjp}@ict.ac.cn

***Abstract*—Recommending citywide POIs where users would visit in the next future benefits many location-based businesses and individuals. To train a decent recommendation model, adequate historical data is usually a prerequisite. However, historical check-ins are usually distributed unevenly, which leads to many cities suffering from data scarcity. To make matters worse, transferring knowledge from data sufficient cities is challenging due to the varying distribution of POIs and city structures. Most of existing next POI recommendation methods assume that the training data is adequate and can not solve these problems. In this paper, we propose a novel meta-learning enhanced neural ordinary differential equation (ODE) method, namely METAODE, which models city-invariant information and city-specified information separately to achieve accurate citywide next POI recommendation. For transferring knowledge from data sufficient cities, METAODE learns city-invariant information including the representation of POIs categories and user groups to extract user preference. Basing on that, METAODE employs a GRU-ODE-Bayes model for city-specified information modeling. It can not only capture the sequential relationships within the historical check-ins but also model the irregular-sampled timestamp in the continuous timeline. Moreover, METAODE leverages meta-learning mechanism to optimize the parameters on various data sufficient cities and train a well-generalized initialization, which can be effectively adapted to data insufficient cities to enhance recommendation performance. Extensive experiments on real-world datasets demonstrate the effectiveness of METAODE. Comparing with the state-of-the-art baselines, METAODE achieves 6.21% and 14.77% improvements on HR and NDCG, respectively.**

***Index Terms*—Meta-learning, Neural ODEs, Memory Network, Recommender Systems, Next POI Recommendation.**

I. INTRODUCTION

Next POI recommendation, which recommends locations with high visiting probabilities for users at a specific future timestamp, has been considered as a critical task for many location-based social networks (LBSNs), such as Foursquare, Yelp and *etc.* In practice, a branch of well-trained recommendation models for various cities have been proved to be vital for accurately evaluating how likely a user would visit a POI in the near future [7], [10], [15]. Training these citywide recommendation models requires a slew of historical data that covers most of POIs in the target cities. Nevertheless, as illustrated in Figure 1, the historical data is unevenly distributed in geographical space so that many cities have no sufficient data for model training. Thus, it is virtually impossible to obtain decent models for these data insufficient cities only utilizing the historical data of themselves. Citywide next POI recommendation, which transfers knowledge from data sufficient source cities to improve the recommendation performance of insufficient ones, is designed to solve this problem and can be widely used in many circumstances.

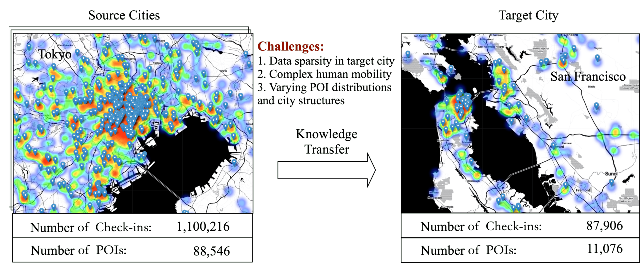


Fig. 1. Challenges in citywide next POI recommendation.

However, as shown in Figure 1, citywide POI recommendation is not trivial due to the following three challenges: (1) Due to the limited training data, user preferences in the data insufficient cities are difficult to be captured. As proved in many works [13], [35], this factor is extremely important to recommend next POIs precisely. (2) Human mobilities are complex, while the obversions (check-in records) of a specific user are sparse and irregular in continuous time. The recommendation method should model this irregularity to generate proper POIs at a specified time. This situation could be even worse in data insufficient cities. (3) Urban structures and POIs distributions are varying in different cities. Models trained on data sufficient cities cannot directly adapt to data insufficient ones.

Despite that a great deal of works have been conducted for POI recommendation, none of them can handle all the three challenges systematically. To solve the first challenge, a lot of time-aware next POI recommendation methods are proposed. Continuous time information is either cut into discrete time bins [18] or discounted the influence of time intervals linearly by point processes [5] for recommendation. We argue that the time information is better to be naturally modeled within the recommendation model, yet these methods fail in that way. For the last two challenges, although many out-of-town POI recommendation methods, such as JIM [35] and CTLM [13], are proposed to recommend POIs in new target city for previously served users, the problem definitions in their

978-1-6654-2845-3/21/$31.00 ©2021 IEEE
DOI 10.1109/MDM52706.2021.00023

works are quite different with ours. Moreover, there exist some spatial transfer learning methods [33] that are proposed to predict macrography city patterns for data insufficient city, such as the traffic flow. Nevertheless, in citywide next POI recommendation, we aim to model the micrography pattern of users and these methods would fail in this task.

With the emergence of meta-learning [8] and neural ODEs [4], these challenges are promising to be solved systematically. By modeling invariant relationships across different cities, meta-learning methods are suitable for transferring knowledge from data sufficient cities and solving the data sparsity problem. Moreover, neural ODEs which naturally model the irregular-sampled time series can be used to capture the complex mobility patterns of users. In this paper, we propose a novel meta-learning enhanced neural ODE method, namely METAODE, for citywide next POI recommendation. It has two attractive characters: (1) Robust to data insufficient cities. METAODE borrows city-invariant knowledge from multiple cities via the common POI categories and user groups to improve the recommendation performance of data insufficient cities. (2) Time-specified recommendation results. The continuous temporal information of user check-ins is naturally modeled in METAODE. Taking the historical data of a user and a future timestamp as inputs, METAODE generates POIs that are highly related to the specified timestamp. In order to have the up-mentioned characters, METAODE models city-invariant information and city-specified information with two parallel processes and optimizes the parameters with meta training.

To transfer knowledge from the data sufficient cities, METAODE shares city-invariant information, *i.e.* categories of POIs and user groups, spanning across all cities to model the user preferences. Instead of employing embedding vectors for users [6], METAODE utilizes a memory network and learns representations of POI categories and user groups from the historical data. For a specific user, METAODE first utilizes a weighted sum of category embeddings that the user previously visited, and then conducts an attention-based query on the user group representations to obtain the final user preference vector.

For modeling the city-specified information, METAODE employs GeoHash to initialize the POI embeddings of the target city and adopts GRU-ODE-Bayes [1] to capture the complex human mobilities. Building upon a continuous-time version of GRU, METAODE not only captures the sequential influence of user check-in but also embeds various time intervals in the continuous hidden states for recommendation. Moreover, by combining the Bayesian updating mechanism upon the representations, METAODE can better integrate the current information with the previously processed data and further enhance the power of these representations. After that, the recommended POIs results can be generated by combining the output of GRU-ODE-Bayes and the user preference vector.

To learn the parameters in METAODE, a meta-learning paradigm is conducted for model optimization. Specifically, METAODE treats POI recommendation in different cities as different meta tasks and learns an intermediate model that performs not too bad on all cities. For the target city, METAODE fine-tunes the intermediate model and yields the recommendation results. In this way, METAODE can achieve robust performance on data insufficient cities even though the training data is limited.

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

- To the best of our knowledge, this is the first work that introduces Neural Ordinary Equation to POI recommendation. The intrinsic characters of neural ODEs are perfectly suitable to model the continuous time intervals of check-in data.
- We propose METAODE, a meta-learning enhanced model for citywide next POI recommendation. It first learns city-specified knowledge and city-invariant knowledge with GRU-ODE-Bayes and memory network respectively. Then, the city-invariant knowledge is transferred to enhance the recommendation performance of data insufficient target cities.
- We conduct extensive experiments on real-world datasets. The results show the superiority of METAODE comparing with all baselines.

II. RELATED WORK

In this section, we summarize the related works of METAODE. These works can be classified into three categories: POI recommendation, neural ODEs and meta-learning.

A. POI recommendation

Existing POI recommendation methods related to this paper can be grouped into two subcategories: next POI recommendation methods and cross-city POI recommendation methods.

For next POI recommendation, a slew of studies [29], [30], [36]–[38] model the sequential influence in historical data and achieving good performance. Additionally, other mobility prediction methods [6], [18] can also be adopted to solve the next POI recommendation problem. Nevertheless, these works are designed for recommend POIs in one specific area and hard to transfer knowledge from other cities/areas.

Recently, many cross-city POI recommendation methods [13], [16], [35] have been proposed to recommend POIs when users visit another city. However, the problem setting of cross-city POI recommendation is quite different from our problem and these methods cannot be used in citywide next POI recommendation.

B. Neural Ordinary Differential Equation

Neural ODE(Neural Ordinary Differential Equation) is recently proposed by [4] which outperforms other works on modeling real-world sporadically time series, *e.g.* standard RNN [3], [17] and RNN-decay [2], [3], [19], [22]. Neural ODE is a continuous version of neural networks that overcomes the limits imposed by discrete-time recurrent neural networks. It proposes a natural way of generating irregularly sampled data coupled with a variational auto-encoder architecture [11]. [26] lately propose a convincing new VAE architecture that uses a Neural-ODE architecture for both encoding and decoding the data. Besides, [1] introduce a continuous-time version of the Gated Recurrent Unit, GRU-ODE, building upon Neural ODEs. Furthermore, [1] leverages a Bayesian update network, GRU-bayes, to process sporadic series, which allows feeding sporadic observations into a continuous ODE dynamics describing the evolution of the probability distribution of the data. As we all know, check-in activities of users in LBSNs can be seen as an irregularly-sampled time series. Therefore, due to the great advantages of neural ODE in modeling time series, in this study, we use neural ODE as a building block for our citywide next POI recommendation method.

C. Meta-learning

Meta-learning, allowing machines to learn new skills or adapt to new environments rapidly with only a small number of training samples [8], has demonstrated success in both supervised learning and reinforcement learning settings. There are four common approaches [34]: 1) use a recurrent neural network equipped with either external or internal memory storing and querying meta-knowledge [20], [21]; 2) learn a meta-optimizer which can quickly optimize the model parameters [14], [23]; 3) learn an effective distance metric between examples [27], [28]; 4) learn an appropriate initialization from which the model parameters can be updated within a few gradient steps [8], [12], [33]. Nevertheless, only a few attempts have been made on space. [33] leverage the similarity of regions between a source city and a target city to construct the similarity regularization for knowledge transfer and utilize a meta-learning paradigm to learn a well-generalized initialization of the spatial-temporal network for traffic prediction that falls into the fourth aforementioned category.

However, these methods work on multimodal features instead of spatial-temporal POI sequences we focus on. Therefore, they cannot be applied directly to solve the problem. Compared with these methods, METAODE, falls into the fourth aforementioned category, explicitly trains the parameters for citywide next POI recommendation on the given task distribution, allowing for extremely efficient adaptation for problems such as a target task learning and rapid adaption for spatio-temporal learning.

III. PRELIMINARIES

In this section, we first summarize the notations and define the citywide POI recommendation problem. Based on the notations, we overview the proposed method METAODE.

A. Problem Formulation

POI in LBSN is defined as a location attached with its category information *i.e.* $v = (l, c)$. Location l is usually represented with the GPS coordinates $l = (x, y)$. Specifically, a check-in record can be organized into a triplet $r^u = (u, v, t)$ which indicates user u visiting POI v at time point t. For a specific user, its historical check-ins form a sequence, *i.e.* $T_u = [r_1, r_2, ..., r_K]$.

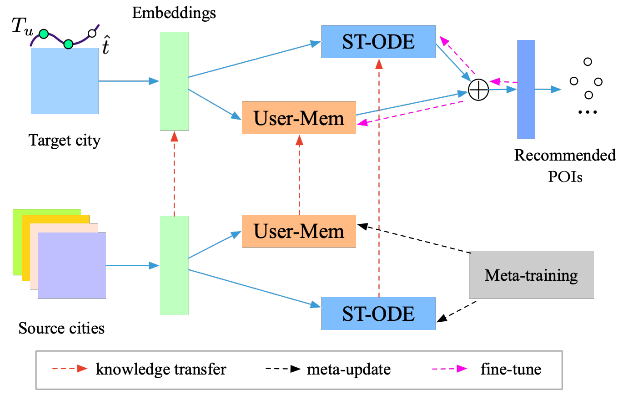


Fig. 2. The architecture of METAODE.

Given the historical dataset of P cities $\mathcal{D} = \{\mathcal{D}^p | p \in 1, 2, ..., P\}$, the citywide next POI recommendation problem aims to predict POIs that users most likely to visit in the near future. For a specific user u in city p, denoting $\mathcal{V} = v_1, ..., v_D$ as the set of all POIs in p, the citywide next POI recommendation problem can be formally defined as follows:

Definition 1: **Citywide Next POI Recommendation**. The goal of citywide next POI recommendation is to predict the most likely POI $\hat{v}$ that u will visit at a specific time point $\hat{t} = t_k + \varDelta t$ by utilizing all information in $\mathcal{D}$. *i.e.*,

$$\hat{v} = \arg\max_{v} \mathbb{P}(v \in \mathcal{V}; \hat{t}, u, T_u, \mathcal{D}, \Theta)$$

where $T_u = [r_1, r_2, ..., r_K]$ represents the historical check-ins and Θ denotes all parameters in the recommendation model.

B. Overview of METAODE

METAODE is a meta-learning framework. It utilizes the recommendation tasks in different cities for meta training and enables the recommendation model to borrow knowledge from multiple data sufficient cities to address the data sparsity problem. The overall architecture is illustrated in Figure 2.

For the convenience of knowledge transfer, METAODE separates the recommendation model into two parallel procedures, *i.e.* USER-MEM for user preference modeling and ST-ODE for spatial-temporal sequence modeling. In USER-MEM, the city-invariant information, such as POI categories and user groups, are employed to represent the preference of the specific user u. In ST-ODE, the city-specified information that contains POIs, locations and continuous timestamps are captured by a GRU-ODE-Bayes model to generate the context representation at $\hat{t}$. Then, METAODE combines the outputs of USER-MEM and ST-ODE to obtain the final recommended POIs.

In parameter learning, both USER-MEM and ST-ODE are jointly optimized in a meta-learning framework. In meta training, for different source cities, METAODE utilizes different POI embeddings in ST-ODE to model the city-specified information, and share one USER-MEM to learn the city-invariant information. In testing, we fine-tune the parameters

TABLE I
NOTATIONS USED IN THIS PAPER

Notations	Descriptions
u, v, l, t	User, POI, location and check-in timestamp.
$\mathcal{U}, \mathcal{V}$	User set and POIs set.
$r^u = (u, v, t)$	A record indicates u visit v at t.
T_u	Historical check-ins of user u.
$\mathcal{D}^p$	Historical data of users in city p.
$\mathcal{D}$	Historical data of all given cities.
h_k	Hidden status of user u at t_k.
$\mathbf{M}$	Embedding matrix of USER-MEM.
d	Embedding dimension.
Θ	All parameters in METAODE.

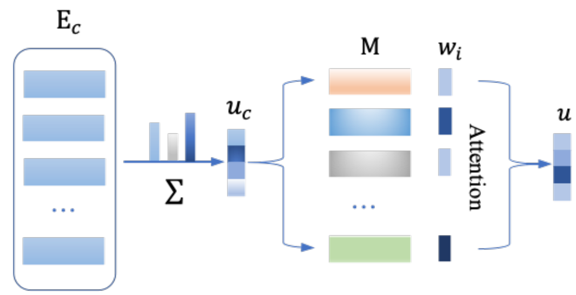


Fig. 3. The architecture of the USER-MEM.

in METAODE on the historical data of p to adopt METAODE to the target city. With this architecture, the city-invariant information lying in data sufficient cities can be automatically transferred to enhance the recommendation performances of data insufficient cities.

IV. METHODOLOGY

METAODE consists of three key modules, *i.e.* USER-MEM, ST-ODE, and meta-learning enhanced models optimization. Next, we specify the details of these modules, respectively.

A. USER-MEM: *Memory Network for User Preference Modeling*

For transferring knowledge from data sufficient cities to insufficient ones, city-invariant information should be used to bridge the transformation. In this paper, we consider two kinds of city-invariant information, *i.e.* categories of POIs and user groups, and integrate them in USER-MEM for modeling the user preference. The structure of USER-MEM is illustrated in Figure 3. Taking the historical check-ins of a user u as input, USER-MEM first embeds the category information of previously visited POIs into vectors $\mathbf{u_c}$. Then, it employs an attentive memory network to model the user groups and yield the final user preference vector $\mathbf{u}$ of u. We will specify the generation of these two vectors below.

Generation of $\mathbf{u_c}$. Assuming that there are N categories of POIs and S user groups spanning across all cities, we denote the embedding vectors of them as $\mathbf{E}_c \in \mathbb{R}^{N\times d_c}$ and $\mathbf{M} \in \mathbb{R}^{S\times d_m}$, respectively. Given the historical check-ins $T_u = [r_1, r_2, ..., r_K]$ of user u, USER-MEM extracts the categories of POIs and generates the previously visited category sequence $[c_1, c_2, ..., c_K]$. By leveraging a lookup operation on $\mathbf{E}_c$, we can obtain the category embeddings $[\mathbf{e_1}, \mathbf{e_2}, ..., \mathbf{e_K}]$ of them and generate $\mathbf{u_c}$ as follows:

$$\mathbf{u_c} = \sum_{k=1}^{K} w_k^c \cdot \mathbf{e_k} \tag{1}$$

where w_k^c is the weight of each category and subject to $\sum_{k=1}^{K} w_k = 1$. After that, the user category embedding is fed to a memory network for the next processing step.

Generation of $\mathbf{u}$. Taking $\mathbf{u_c}$ as the input, USER-MEM employs a memory network for generating the representation of user groups of u. As illustrated in Figure 3, the matrix $\mathbf{M} \in \mathbb{R}^{S\times d}$, represents the embeddings of all user groups shared in all cities. Leveraging $\mathbf{u_c}$ as the query vector and $\mathbf{M}$ as the context matrix, we adopt an attention mechanism to generate the user group representation $\mathbf{u_g}$ as follows:

$$\begin{aligned} w_i^g &= \mathrm{softmax}(\mathbf{u}_c^T \mathbf{m}_i) \\ \mathbf{u} &= \sum_{i\in 1...S} w_i^g \mathbf{m}_i \end{aligned} \tag{2}$$

where $\mathbf{m}_i$ denotes the *i*-th slot in the memory $\mathbf{M}$; w_i^g denotes the weight which represents the extent of u similar with the *i*-th user group. In this way, the preference of the specified user u is transformed to the representation of different user groups. Intuitively, USER-MEM is robust to model users having few check-ins and convenience to knowledge transfer.

Objective function of USER-MEM. For optimizing $\mathbf{E}$, $\mathbf{M}$ and other weights in METAODE, we propose a pair-wise optimization mechanism. Specifically, we first adopt a scoring layer for any user-category pair to calculate the visiting score, and then employ negative sampling to design the pair-wise objective function. Given a pair $< u, c >$, we calculate the joint embedding between $\mathbf{u}$ and the category embedding $\mathbf{e}$ through element-wise dot. *i.e.* $\mathbf{u_{<u,c>}} = \mathbf{u} \odot \mathbf{e}$. Then, $\mathbf{u_{<u,c>}}$ is fed into a fully connected layer to obtain a score:

$$s_{<u,c>} = f(\mathbf{u_{<u,c>}}) \tag{3}$$

where f represents the fully connected layer that transforms the vector into a numeric score.

For the given historical check-ins of user u, we extract its previously visited user-POI pairs $\Delta_p = < u, c_1 >, ... < u, c_K >$ from T_u and treat them as positive pairs. Accordingly, we sample a set of categories that u have not visited as the negative samples $\Delta_n = < u, c'_1 >, ... < u, c'_K >$. Based on the Bayesian personalized ranking (BPR) [24], $s_{<u,c'>}$ is supposed to less than $s_{<u,c>}$. Therefore, the objective function of USER-MEM can be formulated as:

$$\mathcal{L}_{mem}(u) = \sum_{<u,c>\in\Delta_p} \sum_{<u,c'>\in\Delta_n} \max(0, s_{<u,c'>} - s_{<u,c>} + \eta) \tag{4}$$

where η is the decision margin that separates positive and

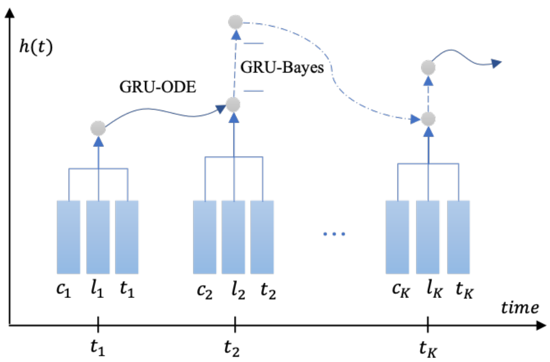


Fig. 4. The architecture of the ST-ODE.

negative examples. By this pair-wise optimization scheme, we can sample negative samples depending on the category distribution of different cities.

B. ST-ODE*: Spatial-Temporal Neural Ordinary Differential Equations*

Besides of the city-invariant information modeled by USER-MEM, city-specific information only related to one specific city is also critical for our task, such as POIs and location distributions. In this subsection, we propose ST-ODE to capture these pieces of information by a neural ODE model [1]. ST-ODE systemically models the influences of three kinds of city-specified information, *i.e.* POIs, geographical locations, and check-in time. Given historical check-ins of u at target city p, *i.e.* $T_u = [r_1, r_2, ..., r_K]$, ST-ODE first embeds each record into a vector and employs GRU-ODE-Bayes to model the spatial-temporal influence. Next, we first describe the generation of check-in record embeddings. After that, the architecture of ST-ODE is specified.

Record embedding. In ST-ODE, we firstly utilize an embedding layer to obtain the record embeddings. Taking the record $r_k^u = (u, v_k, t_k)$ as input, we only focus on three features $[c_k, l_k, t_k]$ which represent the category of POI v_k, the GPS coordinates of v_k and the check-in timestamp t_k respectively to generate the record embedding.

- For the category embedding, we look up the category embedding matrix $\mathbf{E}_c$ (specified in Section IV-A) to generate the embedding of c_k.
- For the POI location embedding, we embed all POIs into vectors by a city-specific embedding matrix $\mathbf{V} \in \mathbb{R}^{D\times d_l}$, where D is the total number of POIs in p. This matrix is initialized with the GeoHash encoding of the GPS coordinates of POIs.
- For timestamp embedding, we discretize one week into 168 slots(7 days $\times$ 24 hours) following [32] and learn an embedding matrix $\mathbf{T} \in \mathbb{R}^{168\times d_l}$. Thus, for any input timestamp t_k, we can map it into a d_t-dimensional vector.

By concatenating these three embeddings, we obtain an initial d-dimensional vector $\mathbf{r}_k$ for record r_k^u, where $d = d_c + d_l + d_t$. After embedding all records in T_u, we feed them into a GRU-ODE-Bayes model for next processing.

GRU-ODE-Bayes for spatial-temporal influence modeling. Enlightened by [1], we use GRU-ODE-Bayes to exploit the spatial-temporal preference of users for citywide POI recommendation. Here, we employ GRU-ODE to control the influence of historical records and conduct a Bayesian update operation in GRU-Bayes to control the influence of the current observed record. In this way, for each input record, we first utilize GRU-ODE to generate the hidden states propagated from the last check-in record and then switch the ST-ODE from propagation to update the hidden states incorporating with the current record.

As illustrated in Figure 4, for the coming check-in record $\mathbf{r}_k$, we firstly employ GRU-ODE to propagate the previous hidden state $\mathbf{h}_{k-1}$ from t_{k-1} to t_k as follows:

$$\mathbf{h}_k^{'} = \text{GRU-ODE}(\mathbf{h}_{k-1}, t_{k-1}, t_k) \tag{5}$$

Then, we utilize GRU-Bayes to process the sporadically incoming check-in record $\mathbf{r}_k$ to update the hidden representations of users which can be formulated as:

$$\mathbf{h}_k = \text{GRU-Bayes}(\mathbf{h}_k^{'}, \mathbf{r}_k) \tag{6}$$

here $\mathbf{h}_k^{'}$ and $\mathbf{h}_k$ denote the hidden representation before and after the update according to the check-in record r_k.

Objective function of ST-ODE. Two losses are employed to optimize the parameters in ST-ODE. The first one is $\mathcal{L}_{ode}$ which drives from GRU-ODE before updating. It represents the negative log-likelihood of the perviously observed check-ins. For a check-in record r_k, $\mathcal{L}_{ode}$ is defined as follows:

$$\begin{aligned} \mathbf{p}_{pre} &= \text{softmax}(f(\mathbf{h}_{k,j}^{'})) \\ \mathcal{L}_{ode}(r_k) &= -\sum_{i=1}^{D} \mathbf{v}_k[i] \cdot \log\, \mathbf{p}_{pre}[i] \end{aligned} \tag{7}$$

where f is a fully connected layer to transform $\mathbf{h}_{k,j}^{'}$ into a D-dimensional vector; $\mathbf{p}_{pre}$ represents the probability distribution on all POIs in city p; $\mathbf{v}_k \in \mathbb{R}^{D\times 1}$ is an one-hot embedding of v_k.

For the second loss, denoting the prior probability of currently observed POI v_k as $\mathbf{p}_{obs}$, we first compute the analogue of the Bayesian update:

$$\mathbf{p}_{post} \propto \mathbf{p}_{pre} \cdot \mathbf{p}_{obs} \tag{8}$$

Let p_{post} denote the predicted probability distribution(from h_k) after applying GRU-Bayes with $\mathbf{r}_k$. The $\mathcal{L}_{bayes}(r_k)$ can be defined as:

$$\mathcal{L}_{bayes}(r_k) = \mathbf{D}_{KL}(\mathbf{p}_{bayes}||\mathbf{p}_{post}) \tag{9}$$

Intuitively, this loss regularizes the difference between the pre-update distribution and the post-update distribution not too huge. Finally, for a user has K records, the total loss of the ST-ODE can be obtained by adding both losses with a balance parameter λ:

$$\mathcal{L}_{st}(u) = \sum_{k=1}^{K} \mathcal{L}_{ode}(r_k) + \lambda \mathcal{L}_{bayes}(r_k) \tag{10}$$

C. Meta-learning Enhanced Model Optimization

For transferring the knowledge from data sufficient cities and enhancing the recommendation performance on insufficient ones, in this section, we propose to use the meta-learning mechanism to optimize the parameters in METAODE. Without losing the generality, we choose one city p as the target city, and other $P-1$ cities denoted with $\mathcal{S}$ can be treated as source cities. The optimization is composed to two procedures: meta training on source cities and fine-tuning on target city for the recommendation.

Meta training on source cities. As described in Section IV-A and IV-B, the loss of METAODE contains two parts, *i.e.* $\mathcal{L}_{mem}$ for learning the city-invariant embeddings and $\mathcal{L}_{st}$ for learning the city-specified spatio-temporal influences. For a given user u in source city s, the objective function of METAODE integrates the two parts.

$$\mathcal{L}^s(u) = \mathcal{L}^s_{st}(u) + \gamma \mathcal{L}^s_{mem}(u) \tag{11}$$

where γ is a trade-off hyper-parameter and is used to balance the effect of each part.

For citywide POI recommendation, the best parameter sets in different cities are not the same. In meta training, we aim to learn an initial parameter set Θ_0 which has the best generalization ability and can be adapted to the target city with limited fine-tuning. To achieve this, model-agnostic meta-learning [8] is employed to learn Θ_0. Specifically, we take the recommendation in different source cities as different tasks. In each training batch, we first sample users from source cities $\mathcal{S}$, and calculate the gradient update for each source city as follows:

$$\begin{aligned} \mathcal{L}^s &= \sum_{\text{Sampled } u \text{ in } \mathcal{U}^s} \mathcal{L}^s(u) \\ \Theta_0^s &= \Theta_0 - \alpha \nabla_{\Theta_0} \mathcal{L}^s \end{aligned} \tag{12}$$

The model parameters are trained by minimizing the loss $\mathcal{L}^s$ with respect to Θ_0 across all users sampled from the source cities. Formally, the meta-objective is as follows:

$$\min_{\Theta_0} \sum_{s \in \mathcal{S}} \mathcal{L}^s | \Theta_0^s \tag{13}$$

where $\mathcal{L}^s|\Theta_0^s$ is the updated loss for sampled users in s with respect to Θ_0^s. Then, the model parameters can be optimized utilizing stochastic gradient descent (SGD), such that the parameters Θ_0 are optimized as follows:

$$\Theta_0 \leftarrow \Theta_0 - \beta \nabla_{\Theta_0} \sum_{s \in \mathcal{S}} \mathcal{L}^s | \Theta_0^s \tag{14}$$

Thus, by optimizing Equation 14, we can obtain an initialization set of parameters Θ_0 which can generalize well on different source cities.

Fine-tuning on target city for recommendation. When recommending POIs for target city p, METAODE takes Θ_0 to initialize the parameters and fine-tunes them with the historical data of p. Specifically, the outputs of ST-ODE and USER-MEM are fused together to generate the final visiting probabilities of all POI for user u at time $\hat{t}$.

Algorithm 1 Optimization of METAODE.

Input:
Historical in source cities $\mathcal{S}$;
Historical in target city p;
Output:
Recommended POIs for users in target city p
Sample users from source cities.
while METAODE not convergence **do**
  for each city $s \in \mathcal{S}$ **do**
    Evaluate $\mathcal{L}_{st}$, $\mathcal{L}_{mem}$ according to Eq. (10) and Eq. (4)
    Calculate the gradient update of Θ_s by Eq. (12)
  end for
  Update Θ_0 with SGD by Eq. (14);
end while
# Fine-tune METAODE on target city p
while METAODE not convergence **do**
  Update parameter Θ_p with gradient descent by Eq. (17)
end while

$$\begin{aligned} \mathbf{h} &= \text{concat}(\mathbf{h}_{\hat{t}}, \mathbf{u}) \\ \mathbf{O} &= \text{softmax}(\mathbf{W}_r \mathbf{h} + b_r) \end{aligned} \tag{15}$$

where $\hat{t}$ is a continuous time point and the latent representation $\mathbf{h}$ contains both spatio-temporal context and user preference information; $\mathbf{O} \in \mathbb{R}^{D \times 1}$ indicates the visiting probabilities of all POIs in city p at $\hat{t}$.

Due to the large amount of POIs in one city, directly optimize the probability distribution over all POIs would be intractable and inefficient. Denoting the ground truth POI of u at $\hat{t}$ as v and y_v as the probability of related item in $\mathbf{O}$, we leverage negative sampling strategy to sample a set of negative POIs $\mathcal{V}^-$ for v. For all check-ins in p, we first calculate all the probabilities of all ground truth POIs $\mathcal{V}$ and the loss function can be formulated as follows:

$$\mathcal{L}_{rec}(u) = -\sum_{v \in \mathcal{V}} \sum_{v^- \in \mathcal{V}^-} ln(y_{v^-} - y_v) \tag{16}$$

For the target city p, we first initialize Θ_p with Θ_0 and optimize as follows:

$$\Theta_p \leftarrow \Theta_p - \alpha \nabla_{\Theta_p} \sum_{u \in \mathcal{U}^p} \mathcal{L}_{rec}(u) \tag{17}$$

where $\mathcal{U}^p$ is the user set of city p. The optimization of METAODE is shown in Algorithm 1. In usage, POIs with high probabilities in $\mathbf{O}$ are selected as the recommended POIs.

V. EXPERIMENT

In this section, we conduct extensive experiments to demonstrate the effectiveness of METAODE. Our experimental evaluation is designed to answer several research questions(**RQs**).

- **RQ1**: Does METAODE outperform other state-of-the-art methods for citywide next POI recommendation?
- **RQ2**: What is the capability of the proposed ST-ODE and USER-MEM?
- **RQ3**: What are the benefits of the meta-learning mechanism?
- **RQ4**: What are the influences of different hyper-parameter settings?
- **RQ5**: What is learned in METAODE to transfer from data sufficient cities?

Next, we introduce the experimental settings, experimental results, ablations studies and hyper-parameter studies respectively. Finally, we conduct some case studies to illustrate what is learned in METAODE.

A. Experimental settings

We first introduce the datasets, compared baseline, evaluation metrics and parameter settings of our experiments. Then we evaluate METAODE against other state-of-the-art algorithms.

1) Datasets: In our experiments, we use open-source check-in datasets, Foursquare [31], to evaluate the performance of METAODE. This dataset includes about 18 months (from April 2012 to September 2013) global-scale check-in data containing 33,278,683 check-ins by 266,909 users on 3,680,126 venues spanning across 415 cities. We overview the statistics of cleaned datasets in Table II. As shown, the average check-in records of users are fairly different between cities. Here, New York (NY), Chicago (CHI) and Washington DC (DC) are used as the source cities, and Los Angeles (LA) and San Francisco (SF) are used as the target cities. For the convenience of evaluation, we select sub-datasets of the target cities that only contain one week and one month check-ins to simulate the data scarcity scenarios.

TABLE II
STATISTICS OF THE SELECTED DATASETS

Cities	Users	POIs	Check-ins
New York	17,385	66,660	581,544
Los Angeles	8,969	21,556	162,658
San Francisco	5,455	11,076	87,906
Chicago	6,870	21,880	184,435
Washington	6,420	13,828	131,444
Tokyo	12,613	88,546	1,100,216

2) Evaluation Metrics: We use two different metrics for performance evaluation, Hit Ration (HR@K)and Normalized Discounted Cumulative Gain (NDCG@K). HR@K measures whether the test POI shows within the top K in the ranked list while the NDCG@K takes the position of the test POI into account and penalizes the score if it is ranked lower in the list.

3) Compared Baselines: Here, we roughly divide 10 compared baselines into three groups: general recommendation methods, next POI recommendation methods, and ablations of METAODE.

General recommendation methods.

- **MF-BPR** [25]: A Bayesian personalized ranking optimized MF model with a pairwise ranking loss. It is tailored to recommendations with implicit feedback data.
- **CML** [9]: A recently incepted algorithm that minimizes the distance between each user-POI interaction in Euclidean space.

Next POI prediction methods.

- **PRME** [7]: A metric embedding based next POI recommendation method to avoid drawbacks of the MF. Specifically, it embeds users and POIs into the same latent space to capture the user transition patterns.
- **STGCN** [37]: Recent next POI recommendation method, which implements time gates and distance gates into LSTM to capture the spatio-temporal relation between successive check-ins for POI recommendation.
- **ST-RNN** [18]: A RNN-based location prediction method, which considers time interval influences. It can be adopted to next POI recommendation directly.
- **DeepMove** [6]: An effectively attentional recurrent network for mobility prediction from lengthy and sparse trajectories, which utilizes the periodicity nature to augment the RNN for mobility prediction.

Ablations of METAODE.

- **METARNN**: Replace the ST-ODE module in METAODE to standard LSTM and train the model with the meta-learning mechanism.
- **METAODE** $_{\text{meta-}}$: Train METAODE in all source city directly without the meta-learning mechanism and fine-tune the model on the target city.
- **METAODE** $_{\text{mem-}}$: A simplified version of METAODE without the memory module. The user preference is modeled with an embedding matrix and concatenated with the record embedding in ST-ODE.

Note that these compared methods are not for citywide next POI recommendation, we utilize historical data in all cities to train these models and use the same testing data to evaluate the performances.

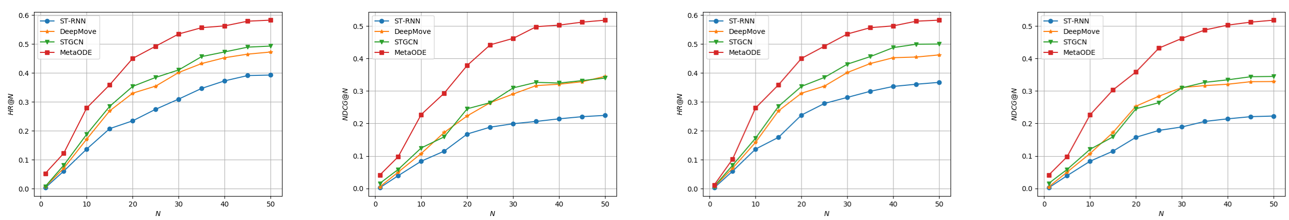
4) Parameter Settings: For USER-MEM, we set the dimensions of each user group as 10, the dimensions of each category as 10. For ST-ODE, the dimension of hidden state of $\mathbf{h}$ is set as 50, and d_t and d_l are set as 48 and 40, respectively. In the training process, the trade-off hyper-parameter γ is set as 10^{-1} and the number of updates for each task is set as 6. Moreover, all the models are trained by Adam. The training batch size for each meta-iteration is set as 128, and the maximum of iteration of meta-learning is set as 1000.

B. Experimental Results

The comparison experimental results of METAODE are shown in Table III. For better understanding, we also illustrate the performance changing of METAODE and other selected baselines in Figure 5 with the increasing of the recommended number of POIs. Next, we analyze the results and answer the research questions **RQ1**.

TABLE III
PERFORMANCE COMPARISON OF DIFFERENT METHODS IN NEXT POI RECOMMENDATION SCENARIOS IN TERMS OF HR@5, NDCG@5, HR@10 AND NDCG@10. BEST PERFORMANCE IS IN BOLDFACE.

Type	Method	City	HR@5	NDCG@5	HR@10	NDCG@10
General	MF-BPR	SF	0.03456	0.02304	0.06283	0.04409
		LA	0.03312	0.02801	0.06103	0.04334
	CML	SF	0.05872	0.04912	0.14871	0.10594
		LA	0.05541	0.04647	0.14276	0.10012
Context-aware	PRME	SF	0.04076	0.03247	0.10189	0.06719
		LA	0.03972	0.03032	0.09535	0.06197
	ST-RNN	SF	0.05006	0.03903	0.13649	0.08359
		LA	0.04718	0.03827	0.12902	0.08143
	DeepMove	SF	0.06012	0.04997	0.15096	0.10889
		LA	0.05982	0.04921	0.14872	0.10727
	STGCN	SF	0.07043	0.05819	0.16762	0.12440
		LA	0.06801	0.05654	0.16426	0.12026
Ablations	METARNN	SF	0.08122	0.07031	0.20872	0.16252
		LA	0.08086	0.06769	0.20224	0.16157
	METAODE $_{\text{meta}-}$	SF	0.09092	0.07568	0.23125	0.17798
		LA	0.08923	0.07221	0.22767	0.17503
	METAODE $_{\text{mem}-}$	SF	0.10522	0.08301	0.25125	0.18799
		LA	0.09886	0.07969	0.24872	0.18503
Our	METAODE	SF	**0.11175(+6.21%)**	**0.09740(+14.77%)**	**0.27863(+10.89%)**	**0.22732(+20.92%)**
		LA	**0.10782(+9.06%)**	**0.09554(+16.58%)**	**0.26882(+9.87%)**	**0.22650(+22.41%)**



(a) SF-HR@N (b) SF-NDCG@N (c) LA-HR@N (d) LA-NDCG@N

Fig. 5. Evaluation of Top-N POI recommendation where N ranges from {1, 5, 10, 15, ..., 45, 50} on San Francisco(SF) and Los Angeles(LA)

- As shown, the performance of METAODE beats all compared baselines on all evaluation metrics consistently. This result shows the superiority of the proposed modules and answers the **RQ1**.
- Among all the compared methods, **STGCN** is the strongest baseline. Although it models the sequential influences and spatial-temporal correlations, it is also inferior to METAODE. The reason is that **STGCN** only captures the information lying in the target city and can not borrow knowledge from other cities.
- For different groups of baselines, the next POI recommendation methods outperform the general ones, which shows that the sequential influences of historical check-ins is critical in this task.

C. Ablation Studies

To answer the **RQ2** and **RQ3**, we compared the performance of METAODE with its ablations.

Contribution of USER-MEM. Comparing METAODE with METAODE $_{\text{mem}-}$, the averaged relative improvement is above 10%. This is because our model can not only capture the city-invariant category information for knowledge transfer but also learn a better initialization of METAODE. Moreover, the long-term user-category pattern memory helps learn a further enhanced initialization.

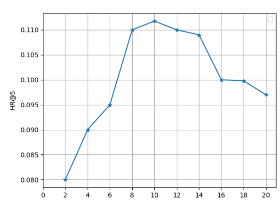
Contribution of ST-ODE. Comparing METAODE with METARNN, METAODE achieves better performance. The results indicate that neural ODE can handle the sporadic data more naturally and can more finely model the dynamics and correlations between the check-in records, which results in higher performance than other methods for both datasets.

Contribution of Meta-learning. Comparing METAODE with METAODE $_{\text{meta}-}$, METAODE performs better in most cases, suggesting the superiority of our meta-learning mechanism. The potential reason is that METAODE $_{\text{meta}-}$ cannot distinguish the city-invariant information, and thereby largely decay the performance. By incorporating meta-training, METAODE not only learns the well-generalized initialization of parameters but also achieves the better performance in target cities.

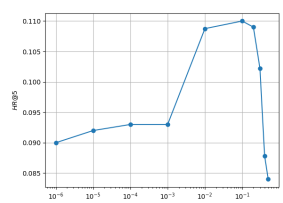
D. Hyper-parameter Studies

To answer the **RQ4**, in this subsection, we evaluate the influences of hyper-parameter settings. Specifically, we analyze the impacts of two key parameters of METAODE, *i.e.*, the dimension d of memory representation and the trade-off factor γ of two losses in the joint objective.

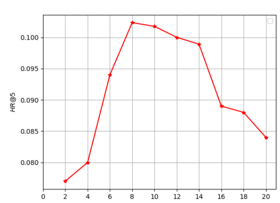
For the dimension d, we change the d from 2 to 20 in USER-MEM. The performance of POI recommendation on San Francisco and Los Angeles are shown in Figure 6(a) and Figure 6(c), respectively. With the increase of d, we find that the performance first increases and then decreases. One potential reason is that the memory provides too little information when the d is too small, while it can include too much irrelevant information when it's too large. Both of the scenarios hurt the performance. For γ, we search γ from 10^{-6} to 0.5. The results are shown in Figure 6(b) and Figure 6(d). The changing pattern is quite similar to d, which indicates both USER-MEM and ST-ODE are critical for our task.



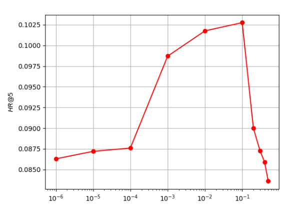
(a)



(b)



(c)



(d)

Fig. 6. (a) (c) HR@5 with respect to the memory dimension on San Francisco/Los Angeles; (b) (d) HR@5 with respect to the value of γ on San Francisco/Los Angeles.

E. Case Studies

To answer the **RQ5**, we sample some representative users in different cities and analyze what is learned in METAODE. Specifically, we visualize the attention weights of users on the city-invariant user group embeddings and compared the favorite categories of different users. Here, we randomly select 10 users from San Francisco (SF) and Los Angeles (LA), and show the top-10 favorite categories of them in Table IV. Note that, $\{u_0 - u_4\}$ are sampled from SF and $\{u_5 - u_9\}$ are sampled from LA. The heat-map of attention weights on the user group embeddings are shown in Figure 7, where the color scale represents the strength of the attention weights, and each row represents the attention score vector for each representative user.

As shown, compared attention weights with the semantic meaning for each user as presented in Table IV, we can find that similar user preferences have similar attention-styles. User u_2, u_7 and u_9 have similar attention vectors and their top-10 categories all includes "Bar", "Train station" and "Coffee shop". In contrast, attention weights of user u_1 is distinctive with others, demonstrating that the preference of u_1 is different from other users. Moreover, although users are sampled from different cities, find that users having similar favorite categories also have close attention style, *e.g.* u_2 and u_9. u_3 and u_6. It indicate that user groups are relatively similar across different cities. These results indicate that the user group embeddings can be used to capture city-invariant information.

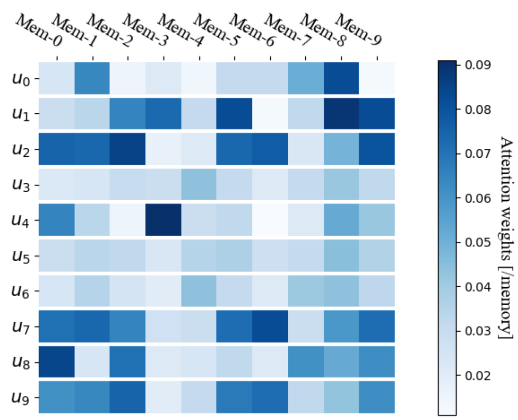


Fig. 7. Attention weights of sampled users.

VI. CONCLUSION

In this paper, we study the problem citywide POI recommendation. A meta-learning enhanced neural ODEs model, namely METAODE, is proposed to leverage knowledge from data sufficient cities to enhance the recommendation performance of insufficient ones. Specifically, METAODE utilizes the memory network to model the user preference with the city-invariant POI category information and user groups, and integrates a neural ODE model for capturing the city-specific spatio-temporal influence. Moreover, we propose to learn a well-generalized initialization of parameters in METAODE through meta-learning mechanism to adapt METAODE to data insufficient target cities. Comprehensive experiments demonstrate the proposed method significant outperforms all compared baselines. In the future, we plan to extend METAODE on two directions:(1) We plan to incorporate the content information such as user's profiles and POIs' attributes to further improve the performance; (2) We plan to consider graph structure(*e.g.*, road network) to enhance the spatial influence modeling.

VII. ACKNOWLEDGMENT

This work has been supported by the National Natural Science Foundation of China under Grant No.: 62077044, 61702470, 62002343.

TABLE IV
FAVORITE CATEGORIES OF SELECTED USERS

UserID	Top-10 categories
u_0	['Sculpture Garden', 'General Travel', 'Clothing Store', 'Gourmet Shop', 'Tattoo Parlor', 'Drugstore / Pharmacy', 'Gift Shop', 'Residential Building (Apartment / Condo)', 'Coffee Shop', "Doctor's Office"]
u_1	['Auditorium', 'Bed & Breakfast', 'Indian Restaurant', 'Parking', 'Assisted Living', 'Taxi', 'Bank', 'Government Building', 'Plaza', 'Fast Food Restaurant']
u_2	['Art Gallery', 'Pier', 'Chinese Restaurant', 'Theater', 'Park', 'Pizza Place', 'Building', 'Bar', 'Train Station', 'Coffee Shop']
u_3	['Rest Area', 'Video Store', 'Light Rail', 'City', 'Hot Spring', 'Neighborhood', 'Skate Park', 'Road', 'Outdoors & Recreation', 'Optical Shop']
u_4	['Donut Shop', 'Grocery Store', 'Building', 'Department Store', 'African Restaurant', 'Miscellaneous Shop', 'General Entertainment', 'Bowling Alley', 'Paper / Office Supplies Store', 'Fried Chicken Joint']
u_5	['Lighthouse', 'Travel & Transport', 'Video Store', 'Scenic Lookout', 'Dog Run', 'General Travel', 'Convenience Store', 'Park', 'Bridge', 'Road']
u_6	['Gym', 'Paella Restaurant', 'Baseball Field', 'Bakery', 'Latin American Restaurant', 'Sporting Goods Shop', 'Fish Market', 'Building', 'Furniture / Home Store', 'Juice Bar']
u_7	['City', 'Greek Restaurant', 'Baseball Stadium', 'Burger Joint', 'Office', 'Bar', 'Train Station', 'Coffee Shop', 'Gym', 'Grocery Store']
u_8	['Auditorium', 'Bed & Breakfast', 'Indian Restaurant', 'Parking', 'Assisted Living', 'Taxi', 'Bank', 'Government Building', 'Plaza', 'Fast Food Restaurant']
u_9	['Art Gallery', 'Bookstore', 'Music Venue', 'Airport', 'Office', 'Bar', 'Train Station', 'Coffee Shop', 'Gym', 'Grocery Store']

REFERENCES

[1] Edward De Brouwer, Jaak Simm, Adam Arany, and Yves Moreau. Gru-ode-bayes: Continuous modeling of sporadically-observed time series. In *NeurIPS*, pages 7377–7388, 2019.

[2] Wei Cao, Dong Wang, Jian Li, Hao Zhou, Lei Li, and Yitan Li. Brits: Bidirectional recurrent imputation for time series. In *Advances in Neural Information Processing Systems*, pages 6775–6785, 2018.

[3] Zhengping Che, Sanjay Purushotham, Kyunghyun Cho, David Sontag, and Yan Liu. Recurrent neural networks for multivariate time series with missing values. *Scientific reports*, 8(1):1–12, 2018.

[4] Tian Qi Chen, Yulia Rubanova, Jesse Bettencourt, and David Duvenaud. Neural ordinary differential equations. In *NeurIPS*, pages 6572–6583, 2018.

[5] Nan Du, Yichen Wang, Niao He, Jimeng Sun, and Le Song. Time-sensitive recommendation from recurrent user activities. In *NeurIPS*, pages 3492–3500, 2015.

[6] Jie Feng, Yong Li, Chao Zhang, Funing Sun, Fanchao Meng, Ang Guo, and Depeng Jin. Deepmove: Predicting human mobility with attentional recurrent networks. In *WWW*, pages 1459–1468. ACM, 2018.

[7] Shanshan Feng, Xutao Li, Yifeng Zeng, Gao Cong, Yeow Meng Chee, and Quan Yuan. Personalized ranking metric embedding for next new POI recommendation. In *IJCAI*, pages 2069–2075. ijcai.org, 2015.

[8] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *ICML*, pages 1126–1135. PMLR, 2017.

[9] Cheng-Kang Hsieh, Longqi Yang, Yin Cui, Tsung-Yi Lin, Serge J. Belongie, and Deborah Estrin. Collaborative metric learning. In *WWW*, pages 193–201. ACM, 2017.

[10] Liwei Huang, Yutao Ma, Shibo Wang, and Yanbo Liu. An attention-based spatiotemporal lstm network for next poi recommendation. *IEEE Transactions on Services Computing*, 2019.

[11] Diederik P. Kingma and Max Welling. Auto-encoding variational bayes. In Yoshua Bengio and Yann LeCun, editors, *ICLR*, 2014.

[12] Yoonho Lee and Seungjin Choi. Gradient-based meta-learning with learned layerwise metric and subspace. In Jennifer G. Dy and Andreas Krause, editors, *ICML*, volume 80, pages 2933–2942. PMLR, 2018.

[13] Dichao Li, Zhiguo Gong, and Defu Zhang. A common topic transfer learning model for crossing city POI recommendations. *IEEE Trans. Cybernetics*, 49(12):4282–4295, 2019.

[14] Ke Li and Jitendra Malik. Learning to optimize. *CoRR*, 2016.

[15] 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 *SIGKDD*, pages 831–840. ACM, 2014.

[16] Xiao Lin, Min Zhang, and Yongfeng Zhang. Joint factorizational topic models for cross-city recommendation. In *APWeb-WAIM*, pages 591–609. Springer, 2017.

[17] Zachary C Lipton, David Kale, and Randall Wetzel. Directly modeling missing data in sequences with rnns: Improved classification of clinical time series. In *Machine Learning for Healthcare Conference*, pages 253–270, 2016.

[18] Qiang Liu, Shu Wu, Liang Wang, and Tieniu Tan. Predicting the next location: A recurrent model with spatial and temporal contexts. In *AAAI*, pages 194–200. AAAI Press, 2016.

[19] Hongyuan Mei and Jason Eisner. The neural hawkes process: A neurally self-modulating multivariate point process. In *NeurIPS*, pages 6754–6764, 2017.

[20] Nikhil Mishra, Mostafa Rohaninejad, Xi Chen, and Pieter Abbeel. A simple neural attentive meta-learner. In *ICLR*. OpenReview.net, 2018.

[21] Tsendsuren Munkhdalai, Xingdi Yuan, Soroush Mehri, and Adam Trischler. Rapid adaptation with conditionally shifted neurons. In Jennifer G. Dy and Andreas Krause, editors, *ICML*, volume 80, pages 3661–3670. PMLR, 2018.

[22] A Rajkomar, E Oren, K Chen, Hajaj N Dai AM, PJ Liu, et al. Scalable and accurate deep learning for electronic health records. npj digit med. 2018, 1801.

[23] Sachin Ravi and Hugo Larochelle. Optimization as a model for few-shot learning. In *ICLR*, 2017.

[24] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. BPR: bayesian personalized ranking from implicit feedback. In *UAI*, pages 452–461. AUAI Press, 2009.

[25] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. BPR: bayesian personalized ranking from implicit feedback. In *UAI*, pages 452–461. AUAI Press, 2009.

[26] Yulia Rubanova, Tian Qi Chen, and David Duvenaud. Latent ordinary differential equations for irregularly-sampled time series. In *NeurIPS*, pages 5321–5331, 2019.

[27] Jake Snell, Kevin Swersky, and Richard S. Zemel. Prototypical networks for few-shot learning. In Isabelle Guyon, Ulrike von Luxburg, Samy Bengio, Hanna M. Wallach, Rob Fergus, S. V. N. Vishwanathan, and Roman Garnett, editors, *NeurIPS*, pages 4077–4087, 2017.

[28] Flood Sung, Yongxin Yang, Li Zhang, Tao Xiang, Philip H. S. Torr, and Timothy M. Hospedales. Learning to compare: Relation network for few-shot learning. In *CVPR*. IEEE Computer Society, 2018.

[29] Suhang Wang, Yilin Wang, Jiliang Tang, Kai Shu, Suhas Ranganath, and Huan Liu. What your images reveal: Exploiting visual contents for point-of-interest recommendation. In *WWW*, pages 391–400. ACM, 2017.

[30] Carl Yang, Lanxiao Bai, Chao Zhang, Quan Yuan, and Jiawei Han. Bridging collaborative filtering and semi-supervised learning: A neural approach for POI recommendation. In *SIGKDD*, pages 1245–1254. ACM, 2017.

[31] Dingqi Yang, Bingqing Qu, Jie Yang, and Philippe Cudré-Mauroux. Revisiting user mobility and social relationships in lbsns: A hypergraph embedding approach. In *WWW*, pages 2147–2157. ACM, 2019.

[32] Di Yao, Chao Zhang, Jian-Hui Huang, and Jingping Bi. SERM: A recurrent model for next location prediction in semantic trajectories. In *CIKM*, pages 2411–2414. ACM, 2017.

[33] Huaxiu Yao, Yiding Liu, Ying Wei, Xianfeng Tang, and Zhenhui Li. Learning from multiple cities: A meta-learning approach for spatial-temporal prediction. In *WWW*, pages 2181–2191. ACM, 2019.

[34] Huaxiu Yao, Ying Wei, Junzhou Huang, and Zhenhui Li. Hierarchically structured meta-learning. In *ICML*, volume 97, pages 7045–7054. PMLR, 2019.

[35] Hongzhi Yin, Bin Cui, Xiaofang Zhou, Weiqing Wang, Zi Huang, and Shazia W. Sadiq. Joint modeling of user check-in behaviors for real-time point-of-interest recommendation. *ACM Trans. Inf. Syst.*, 35(2):11:1–11:44, 2016.

[36] Fuqiang Yu, Lizhen Cui, Wei Guo, Xudong Lu, Qingzhong Li, and Hua Lu. A category-aware deep model for successive POI recommendation on sparse check-in data. In *WWW*, pages 1264–1274. ACM, 2020.

[37] Pengpeng Zhao, Haifeng Zhu, Yanchi Liu, Jiajie Xu, Zhixu Li, Fuzhen Zhuang, Victor S. Sheng, and Xiaofang Zhou. Where to go next: A spatio-temporal gated network for next POI recommendation. In *AAAI*, pages 5877–5884. AAAI Press, 2019.

[38] Xiao Zhou, Cecilia Mascolo, and Zhongxiang Zhao. Topic-enhanced memory networks for personalised point-of-interest recommendation. In *SIGKDD*, pages 3018–3028. ACM, 2019.