CAPTOR: A Crowd-Aware Pre-Travel Recommender System for Out-of-Town Users

Haoran Xin*
xinhaoran@mail.ustc.edu.cn
University of Science and Technology of China
Baidu Research

Xinjiang Lu†
luxinjiang@baidu.com
Baidu Research

Nengjun Zhu
zhu_nj@shu.edu.cn
Shanghai University

Tong Xu
tongxu@ustc.edu.cn
University of Science and Technology of China

Dejing Dou†
doudejing@baidu.com
Baidu Research

Hui Xiong†
xionghui@gmail.com
HKUST Shenzhen-Hong Kong Collaborative Innovation Research Institute
The Hong Kong University of Science and Technology (Guangzhou)

ABSTRACT

Pre-travel out-of-town recommendation aims to recommend Point-of-Interests (POIs) to the users who plan to travel out of their hometown in the near future yet have not decided where to go, i.e., their destination regions and POIs both remain unknown. It is a non-trivial task since the searching space is vast, which may lead to distinct travel experiences in different out-of-town regions and eventually confuse decision-making. Besides, users' out-of-town travel behaviors are affected not only by their personalized preferences but heavily by others' travel behaviors. To this end, we propose a Crowd-Aware Pre-Travel Out-of-town Recommendation framework (CAPTOR) consisting of two major modules: spatial-affined conditional random field (SA-CRF) and crowd behavior memory network (CBMN). Specifically, SA-CRF captures the spatial affinity among POIs while preserving the inherent information of POIs. Then, CBMN is proposed to maintain the crowd travel behaviors w.r.t. each region through three affiliated blocks reading and writing the memory adaptively. We devise the elaborated metric space with a dynamic mapping mechanism, where the users and POIs are distinguishable both inherently and geographically. Extensive experiments on two real-world nationwide datasets validate the effectiveness of CAPTOR against the pre-travel out-of-town recommendation task. The code is available at https://github.com/xhran2010/CAPTOR.

CCS CONCEPTS

• **Information systems** → **Recommender systems**; • **Computing methodologies** → **Learning latent representations**.

KEYWORDS

Pre-travel recommendation, CRF, memory network, metric learning

ACM Reference Format:
Haoran Xin, Xinjiang Lu, Nengjun Zhu, Tong Xu, Dejing Dou, and Hui Xiong. 2022. CAPTOR: A Crowd-Aware Pre-Travel Recommender System for Out-of-Town Users. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '22), July 11–15, 2022, Madrid, Spain.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3477495.3531949

1 INTRODUCTION

With the advances of location-based services (LBS), e.g., Google Maps and Yelp, Point-of-Interest (POI) recommendation has become inevitable in our daily lives. Whereas, when a user travels out of her/his hometown, traditional POI recommender systems may fail since they are dedicated to recommending POIs within a specific region. Most of the existing works [4, 6, 20, 39, 47, 51, 53] for out-of-town recommendation aim to solve issues such as data sparsity, interest drift and geographical gap.

However, a more challenging task, i.e., pre-travel out-of-town recommendation, is rarely studied. Pre-travel recommender system is devoted to recommending POIs to the users who plan to travel out of their hometowns yet have not decided the destination region. When deciding which city and/or spots to visit, a user would like to search and compare a list of cities/points of interest. The pre-travel out-of-town recommender can help the users make their decisions more efficiently. For instance, in Fig. 1, to spend a vacation, a user who resides in Chengdu is planning to travel somewhere else s/he never visited before. In such case, the pre-travel recommender should provide travel advice that fits the user's interest on potential POIs in different out-of-town regions. This is indeed challenging since the candidate POIs are very diversified when out-of-town POIs in different regions are fairly taken into account.

*This work was done when the first author was an intern at Baidu Research under the supervision of the second author.
†Corresponding authors.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
SIGIR '22, July 11–15, 2022, Madrid, Spain.
© 2022 Association for Computing Machinery.
ACM ISBN 978-1-4503-8732-3/22/07…$15.00
https://doi.org/10.1145/3477495.3531949



Figure 1: An illustrative example of pre-travel out-of-town recommendation.

In the literature, a body of research efforts has already been made on the out-of-town recommendation problem. For example, [39, 51–53] employed Latent Dirichlet Allocation (LDA) to infer user's behaviors on the designed probabilistic graph by taking the preference, interest drift, spatiotemporal context into account. [47] investigated the cross-city travel intention with the deep neural topic model. [34] exploited region-based out-of-town pattern discovery to address the data sparsity problem. [4, 20] utilized transfer learning techniques to model the interest drift of out-of-town users. Nevertheless, these methods may fail to tackle the pre-travel out-of-town recommendation problem due to two key challenges. First, as the destination region remains unknown, the candidates involve a large number of diversified POIs from all other regions. Second, the crowd behaviors, i.e., where other travelers regularly visit, tend to affect the individual's travel behavior. For example, an out-of-town user who does not favor history might still have a try at the historical spots as many other travelers did if s/he is in a city with a long history. Plus, different regions may have totally different crowd patterns. It is hard to capture and utilize the crowd behaviors in different regions fittingly.

In out-of-town recommendation, the candidate POIs that users might be interested in are highly diverse, which are attributed to not only the intrinsic characteristics of POIs but also the cities the candidates belong to. Therefore, it is wise to distinguish the candidate POIs geographically and integrate such region profiles into the POI representation for pre-travel out-of-town recommendation. Moreover, regarding conformity, the crowd behaviors in respective regions should be investigated for out-of-town users. While treating conformity without focusing on the diversity among candidates, just like the canonical collaborative filtering, appears insufficient. Thus, a more dedicated solution is required for crowd behavior maintenance and utilization.

To this end, in this work, we propose a Crowd-Aware Pre-Travel Out-of-town Recommendation framework (CAPTOR). We equip CAPTOR with two devised modules, namely *spatial-affined conditional random field* (SA-CRF) and *crowd behavior memory network* (CBMN). To be specific, in SA-CRF, a set of spatial-affined POI embeddings are viewed as random variables constrained by both pairwise energy and unary energy. The pairwise energy is responsible for capturing the spatial affinity among POIs with the help of the attention mechanism over a spatial graph. The unary energy preserves generic POI properties. For CBMN, we maintain crowd behavior memory matrices to group and summarize the user's hometown and out-of-town behaviors. Besides, we design three blocks to write, read and maintain the memory group, respectively. In particular, the write block updates the memory w.r.t. the user's out-of-town behaviors adaptively. The read block infers the user's memory embedding with a hierarchical attention network. The reconstruction block bridges the user's inferred memory and factual memory. Accordingly, we perform the recommendation in the elaborated metric space, which is divided into several subspaces w.r.t. the user's hometown through a dynamic mapping mechanism.

To summarize, the main contributions of our work are as follows:

- To the best of our knowledge, we are among the first to study the pre-travel out-of-town POI recommendation problem, which can benefit the out-of-town users who have not decided on the destination region in the upcoming vacation.
- We propose a data driven approach, namely CAPTOR, to tackling the pre-travel out-of-town recommendation problem. Several techniques, such as SA-CRF and CBMN, are proposed to fulfill the pre-travel recommendation through distinguishable representations by simultaneously capturing the spatial affinity and crowd behaviors.
- Extensive experiments conducted on two real-world datasets validate the effectiveness of the proposed approach for pre-travel recommendation.

2 RELATED WORK

POI Recommendation is generally differentiated from the traditional recommenders [19, 35, 37, 60] that are dedicated for online services, with regard to the influential factors. Recent studies have worked on incorporating various auxiliary information for POI recommendation. For instance, there have been studies focusing on the spatial clustering phenomenon of users' check-ins [24], intimate relationships among neighboring POIs [27], and spatiotemporal constraints [31] in human mobility. Besides, other types of factors have also been exploited, such as texts [2, 39], photos [41, 58] and social ties [21, 54]. On the other hand, from the point of view of technologies, recent POI recommenders employ a joint learning paradigm, where the check-in is depicted as a synchronized decision influenced by multiple factors together. For example, matrix-factorization (MF) incorporates different factors into collaborative filtering (CF) framework [16, 24, 40]. The generative model with the help of Bayesian inference is another line of research [39, 42, 52]. Particularly, reinforcement learning techniques are explored for planning problem of POIs [13, 56]. In addition, the next POI recommendation task attracts lots of attention. Many deep models have been explored, such as recurrent neural network (RNN) [26, 57], Markov chain [3, 55] and attention mechanism [5, 23, 28].

Out-of-town Recommendation suggests new POIs for out-of-town users to visit. The out-of-town recommendation is more intractable due to issues like cold start and interest drift. [6] was the first to solve the problem by studying the social influence. Existing works [39, 51–53] mainly utilized Latent Dirichlet Allocation (LDA) to model the interest drift with the help of user preferences, spatiotemporal effects and POI contents. Besides, there have been deep

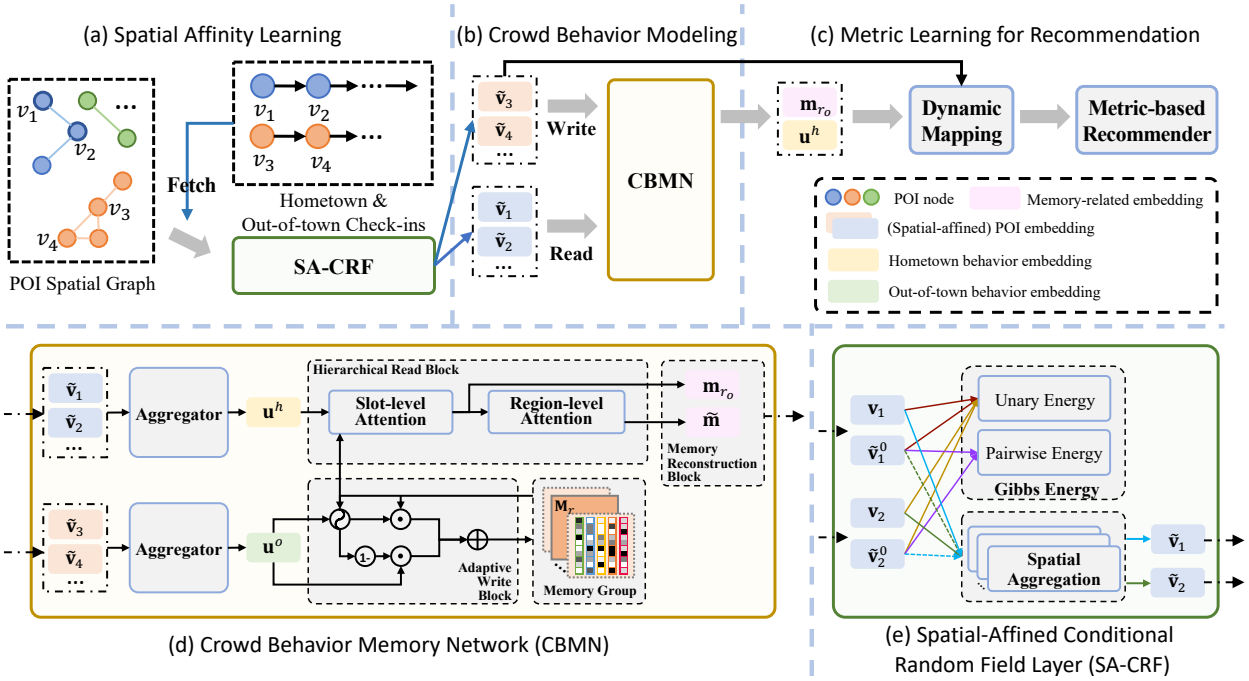


Figure 2: The illustration of CAPTOR framework. Figures (a)(b)(c) show the overview of CAPTOR, and Figures (d)(e) present the respective inner structures of CBMN and SA-CRF.

learning approaches devoted to addressing the problem. [4, 20] combined interest drift together with transfer learning, and learned the user preference from both hometown and out-of-town cities. [47] performed in-depth analysis of the user's out-of-town travel intentions with the neural topic model. [34] exploited region-based out-of-town pattern discovery to address the data sparsity issue. Nevertheless, few studies have touched the topic of pre-travel out-of-town recommendation.

Memory Network was first proposed to access long-term memory for the question answering task [44]. [17, 33] utilized differentiable operators, allowing them to access past experiences in an end-to-end manner. With the progress of deep learning, external memory mechanism has been shown to be effective in handling long term and global signals in different areas such as reinforcement learning [30], natural language processing [14, 36], computer vision [8, 46], and recommender system [35, 59]. In this work, we devise a novel crowd behavior memory network with hierarchical reading, adaptive writing and memory reconstruction to maintain diversified crowd behaviors in different regions and enhance the recommendation for out-of-town users.

Conditional Random Field is a probabilistic graphical model to leverage the pairwise relationship within a number of random variables. [18] originally came up with CRF for sequential data prediction. Recent works developed CRF models with neural networks to solve problems in various domains. For example, [7] armed graph convolutional neural networks with CRF to encourage similar nodes to have similar features. [29] also applied CRF to graph neural networks to exploit the correlation of node labels. Furthermore, [48] integrated this powerful tool to learn representations for consumer loan fraud detection. Inspired by the previous works, we incorporate CRF to learn the spatial affinity of POIs in real world, which plays an important role for pre-travel out-of-town recommendation.

3 PROBLEM DEFINITION

We start by defining several relevant data structures and concepts that help understand our problem.

Definition 1 (POI). *A POI is a spatial item related to a geographical location. We use v to represent a POI, r_v to denote its located region, and ℓ to denote its location as longitude and latitude coordinates.*

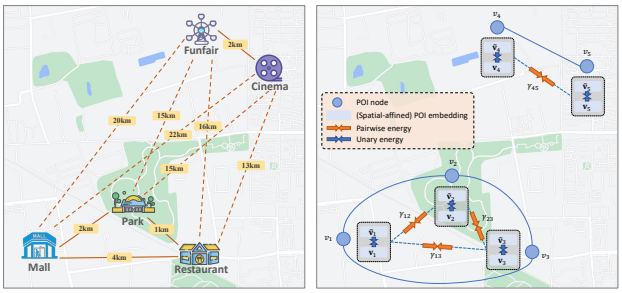
Definition 2 (Check-in). *If a user checks in a POI, it indicates that s/he has a direct interaction with the POI in the real world. Formally, we denote a user u's check-in activity on POI v at time t with a three-tuple $c = (u, t, v)$.*

Definition 3 (User's hometown). *Given a user u, we denote a region r_h as the user's hometown where the user lives or works for a period of time.* [1]

Definition 4 (Out-of-town Travel Behavior). *Given a user u and her/his hometown r_h in which s/he leaves check-in records $\vec{c}_h$, when s/he travels to the out-of-town region r_o and also leaves check-in records $\vec{c}_o$ in r_o, we call such behavior an out-of-town travel behavior, denoted as a five-tuple $\tau = (u, \vec{c}_h, \vec{c}_o, r_h, r_o)$.*

Formally, given a set of users $\mathcal{U} = \{u_i\}_{i=1}^{|\mathcal{U}|}$, a set of regions $\mathcal{R} = \{r_i\}_{i=1}^{|\mathcal{R}|}$, a set of POIs $\mathcal{V} = \{v_i\}_{i=1}^{|\mathcal{V}|}$, and the corresponding

[1] We develop an algorithm to discover a user's hometown, which is available in Appendix.



(a) (b)

Figure 3: (a) An example of the spatial affinity of POIs. (b) The explanation of two types of Gibbs energy in SA-CRF.

out-of-town travel behavior records $\mathfrak{T} = \{\tau_i\}_{i=1}^{|\mathfrak{T}|}$ generated by $\mathcal{U}$, our problem is stated as below.

Problem 1 (Pre-travel Out-of-town Recommendation). *Given the historical out-of-town travel behavior records $\mathfrak{T}$ generated by $\mathcal{U}$, learn a recommender $\mathcal{F}$. Then, a new coming user $u^* \notin \mathcal{U}$ shows up with her/his hometown check-in records $\vec{c}_h$ being observed in r_h^*, while her/his out-of-town region r_o^* and out-of-town check-ins $\vec{c}_o$ remain unknown. Next, recommend a list of out-of-town POIs $\mathcal{V}^* \subset \mathcal{V}$ which are not in r_h^* to u^* with $\mathcal{F}$.*

4 METHODOLOGY

4.1 Framework Overview

Fig. 2 presents the architecture of CAPTOR, which mainly consists of two components, i.e., *Spatial-Affined Conditional Random Field (SA-CRF)* and *Crowd Behavior Memory Network (CBMN)*:
1) SA-CRF takes the spatial-affined POI embeddings as random variables conditioned on preliminary POI embeddings, then incorporates CRF characterized by the Gibbs distribution, where the energy function is defined as the composition of unary energy and pairwise energy. **2)** CBMN consists of three blocks to maintain the group of crowd memory matrices. In particular, a write block takes the out-of-town embedding as input and updates the crowd memory matrix adaptively. A read block uses a hierarchical attention network to infer the user's crowd-aware memory. In addition, a memory reconstruction block is designed to bridge the inferred and factual memories.

Then, users' hometown behavior embeddings and memory embeddings are sent to a metric space and divide the metric space into subspaces according to users' hometown regions with a dynamic mapping mechanism. Finally, the parameters involved in CAPTOR are learned by minimizing the composite loss consisting of CRF loss, reconstruction loss and metric loss.

4.2 Spatial-Affined Conditional Random Field

Exploring the geographical properties has proved to be helpful in the studies of POIs [22–24, 31]. According to Tobler's first law of geography [38], during short-term out-of-town travel, users are inclined to pay visits to those POIs not very far from each other. Therefore, it is vital to explore the spatial affinity of POIs in the pre-travel out-of-town recommendation scenario. For example, in Fig. 3(a), with regard to the geographical distance among these POIs, it is reasonable to assume that the spatial affinity of the POIs connected with solid lines is stronger than that of those with dash lines. Even more to the point, with the spatial affinity of POIs being modeled, the affinity of regions can also be naturally captured, i.e., POIs in different regions would have relatively distinct representations. Such regional profile information could assist in learning the user's regional preference to facilitate recommendation, especially when the candidates involve the POIs from different regions. Inspired from CRF [18], a probabilistic model to infer dependencies within a number of random variables, we devise a novel Spatial-Affined CRF layer (SA-CRF) to capture the spatial affinity comprehensively.

Specifically, we first assign a preliminary embedding $\mathbf{v}_i \in \mathbb{R}^d$ to each POI v_i and then view the spatial-affined POI embeddings as random variables $\widetilde{\mathbf{V}} = \{\widetilde{\mathbf{v}}_i\}_{i=1}^{|\mathcal{V}|}$ which are conditioned on $\mathbf{V} = \{\mathbf{v}_i\}_{i=1}^{|\mathcal{V}|}$, where $\widetilde{\mathbf{v}}_i \in \mathbb{R}^d$ is the spatial-affined embedding of v_i, and d is the dimensionality. We build the CRF characterized by the Gibbs distribution as below,

$$P(\widetilde{\mathbf{V}}|\mathbf{V}) = \frac{1}{Z(\mathbf{V})}\exp(-E(\widetilde{\mathbf{V}}|\mathbf{V})), \tag{1}$$

where $Z(\cdot)$ and $E(\cdot)$ are the partition function and Gibbs energy function, respectively.

As shown in Fig. 3(b), the Gibbs energy function consists of the unary energy ψ_u and the pairwise energy ψ_p:

$$E(\widetilde{\mathbf{v}}_i|\mathbf{v}_i) = \alpha\psi_u(\mathbf{v}_i, \widetilde{\mathbf{v}}_i) + \beta\psi_p(\widetilde{\mathbf{v}}_i, \widetilde{\mathbf{v}}_j), \tag{2}$$

where α and β are two hyper-parameters controlling the respective contributions. Similar to [7], we define the unary energy as,

$$\psi_u(\mathbf{v}_i, \widetilde{\mathbf{v}}_i) = \|\mathbf{v}_i - \widetilde{\mathbf{v}}_i\|_2^2 .$$

By minimizing such unary energy, the spatial-affined POI embeddings are encouraged to preserve the inherent information from the preliminary POI embeddings. With regard to the pairwise energy, we intend to incorporate spatial information and enable the component to handle the spatial affinity. Therefore, we first construct a POI spatial graph $\mathcal{G}_s = (\mathcal{V}_s, \mathcal{E}_s)$, where $\mathcal{V}_s = \mathcal{V}$ is the set of POIs, $\mathcal{E}_s$ is the edge set indicating connectivity among POIs. Specifically, we define spatial connectivity $e_{ij}^s \in \mathcal{E}_s$ as,

$$e_{ij}^s = \begin{cases} 1, & d(v_i, v_j) \le \rho_d , \\ 0, & otherwise , \end{cases}$$

where $d(\cdot,\cdot)$ denotes the geographical distance, ρ_d is a distance threshold. Then, we can define the pairwise energy as,

$$\psi_p(\widetilde{\mathbf{v}}_i, \widetilde{\mathbf{v}}_j) = \gamma_{ij} \left\|\widetilde{\mathbf{v}}_i - \widetilde{\mathbf{v}}_j\right\|_2^2 , \tag{3}$$

where γ_{ij} is the spatial strength between a pair of POIs. In order to adaptively weigh the spatial strength among the POIs, we further adopt an attention mechanism based on the POI spatial graph as,

$$\xi_{ij} = (\mathbf{a}^s)^\mathsf{T}\sigma\left(\mathbf{W}^s\left[\widetilde{\mathbf{v}}_i \oplus \widetilde{\mathbf{v}}_j\right]\right),$$
$$\gamma_{ij} = \frac{\exp(\xi_{ij})}{\sum_{z\in\mathcal{N}_i}\exp(\xi_{iz})},$$

where $\mathbf{a}^s \in \mathbb{R}^d$ and $\mathbf{W}^s \in \mathbb{R}^{d\times 2d}$ are two trainable parameters, $\sigma(\cdot)$ is the activation function, $\oplus$ is the concatenation operator, and $\mathcal{N}_i$

denotes the set of adjacent POIs of v_i in $\mathcal{G}_s$. Through the constraint in Eq. (3), the POIs that have stronger spatial affinity are forced to be closer in the metric space as well.

In order to make it feasible to compute the posterior probability in Eq. (1) and update the spatial-affined embeddings, we resort to the mean-field approximation widely used in variational inference. Specifically, the distribution $Q(\widetilde{\mathbf{V}})$ is derived by the independent marginal distributions as,

$$Q(\widetilde{\mathbf{V}}) = \prod_{i=1}^{|\mathcal{V}|} Q_i(\widetilde{\mathbf{v}}_i).$$

And we use $Q(\widetilde{\mathbf{V}})$ to approximate $P(\widetilde{\mathbf{V}}|\mathbf{V})$ by minimizing

$$\mathrm{KL}\left(Q(\widetilde{\mathbf{V}}) \| P(\widetilde{\mathbf{V}}|\mathbf{V})\right),$$

where $\mathrm{KL}(\cdot\|\cdot)$ is the Kullback-Leibler divergence. From Eqs. (1) and (2), we can find that

$$Q_i^*(\widetilde{v}_i) \sim \exp\left(-\left(\alpha \left\|\mathbf{v}_i - \widetilde{\mathbf{v}}_i\right\|_2^2 + \beta \sum_{j \in \mathcal{N}_i} \gamma_{ij} \left\|\widetilde{\mathbf{v}}_i - \widetilde{\mathbf{v}}_j\right\|_2^2\right)\right), \quad (4)$$

where Q_i^* is the optimal distribution of Q_i. Then we acquire the maximum probability of $P(\widetilde{\mathbf{V}}|\mathbf{V})$ by performing the expectation of Q_i^*, since Q_i^* is proved to be a Gaussian function in Eq. (4). Hence, we define the spatial aggregation of the affined embeddings as,

$$\widetilde{\mathbf{v}}_i^{l+1} = \frac{\alpha \mathbf{v}_i + \beta \sum_{j \in \mathcal{N}_i} \gamma_{ij} \widetilde{\mathbf{v}}_j^l}{\alpha + \beta \sum_{j \in \mathcal{N}_i} \gamma_{ij}}, \quad l = 0, \cdots, L, \quad (5)$$

where $\widetilde{\mathbf{v}}_j^0 = \mathbf{v}_j$. After L iterations following Eq. (5), we let the outcome spatial-affined POI embeddings $\widetilde{\mathbf{v}}_i = \widetilde{\mathbf{v}}_i^L$. In this way, the spatial-affined POI embeddings can incorporate the spatial information among POIs while retaining the inherent properties at the same time.

Finally, we can define the CRF loss by minimizing the Gibbs energy function in Eq. (2) as follows,

$$\mathcal{L}_C = \sum_{i=1}^{|\mathcal{V}|} \left(\alpha \left\|\mathbf{v}_i - \widetilde{\mathbf{v}}_i\right\|_2^2 + \beta \sum_{j \in \mathcal{N}_i} \gamma_{ij} \left\|\widetilde{\mathbf{v}}_i - \widetilde{\mathbf{v}}_j\right\|_2^2\right). \quad (6)$$

4.3 Crowd Behavior Memory Network

As aforementioned, a user's pre-travel decision is quite a complicated process. On the one hand, when determining the destination region, a user usually takes the general characteristics of the region into account. For example, if s/he prefers scenery, then s/he is less likely to visit a well-developed metropolitan city where malls and skyscrapers are more popular. On the other hand, a user's POI check-in activities in an out-of-town region are also entangled with the behaviors of other travelers. To meet these, we propose a Crowd Behavior Memory Network (CBMN). Please refer to Fig. 2(d) for the framework overview of CBMN.

Specifically, we first employ two aggregators to summarize a user's hometown and out-of-town check-in behaviors separately. Given a user u's travel behavior record $\tau = (u, \vec{c}_h, \vec{c}_o, r_h, r_o)$, her/his hometown and out-of-town behavior embeddings are derived based on her/his check-in records as,

$$\begin{aligned} \mathbf{u}^h &= \mathrm{AGG}\left([\widetilde{\mathbf{v}}_i]_{i=1}^{|\vec{c}_h|}\right), \\ \mathbf{u}^o &= \mathrm{AGG}\left([\widetilde{\mathbf{v}}_i]_{i=1}^{|\vec{c}_o|}\right), \end{aligned} \quad (7)$$

where $\mathrm{AGG}(\cdot)$ is an arbitrary aggregation operation, such as average pooling, max pooling or attention pooling, and $\widetilde{\mathbf{v}}_i$ is the spatial-affined embedding generated from SA-CRF in Section 4.2. Then, in order to store the crowd behaviors in different regions, we maintain a memory for each region and build an external memory matrix group $\mathbf{M} \in \mathbb{R}^{|\mathcal{R}| \times m \times d}$ where m is the number of memory slots. To be specific, there are three blocks in CBMN, i.e., adaptive write block, hierarchical read block and memory reconstruction block. These three blocks are introduced to maintain the built memory matrix group.

4.3.1 ***Adaptive Write Block.*** The write block is devised to automatically overwrite the memory according to the current user's out-of-town behaviors. We develop a gate-controlled forgetting mechanism to determine how much behavioral information should be passed into the memory.

To be specific, the gate for the i-th slot in the memory matrix of region r_o is calculated by

$$g_i = \sigma\left((\mathbf{a}^g)^\top \left[\mathbf{u}^o \oplus \mathbf{M}_{r_o,i}\right]\right),$$

where $\mathbf{a}^g \in \mathbb{R}^{2d}$ is a trainable parameter, and $\mathbf{M}_{r_o,i} \in \mathbb{R}^d$ denotes the i-th slot vector of memory matrix of region r_o. Then the i-th slot of the memory matrix of region r_o is updated as

$$\mathbf{M}'_{r_o,i} = g_i \mathbf{M}_{r_o,i} + (1 - g_i)\mathbf{u}^o.$$

The updated memory matrix $\mathbf{M}'_{r_o} \in \mathbb{R}^{d \times m}$ of region r_o is,

$$\mathbf{M}'_{r_o} = \left[\mathbf{M}'_{r_o,1}, \mathbf{M}'_{r_o,2}, \cdots, \mathbf{M}'_{r_o,m}\right].$$

After these propagations, the new crowd memory matrix of region r_o preserves part of the old crowd behavior information in r_o while absorbing the current user's behavior knowledge traveling to r_o.

4.3.2 ***Hierarchical Read Block.*** When a user is planning to travel out-of-town, the crowd behaviors could affect her/his out-of-town check-in behaviors. Thus, we propose a hierarchical read block that utilizes a hierarchical attention network to derive the crowd memory for each out-of-town user.

Firstly, we define a slot-level attention network for a user to interact with the memory as follows,

$$\begin{aligned} \mu_{i,r}^s &= (\mathbf{a}^p)^\top \sigma\left(\mathbf{W}^p \left[\mathbf{u}^h \oplus \mathbf{M}_{r,i}\right]\right), \\ \pi_{i,r}^s &= \frac{\exp(\mu_{i,r}^s)}{\sum_{j=1}^m \exp(\mu_{j,r}^s)}, \end{aligned} \quad (8)$$

where $\mathbf{a}^p \in \mathbb{R}^d$ and $\mathbf{W}^p \in \mathbb{R}^{d \times 2d}$ are parameters, $\mathbf{M}_r \in \mathbb{R}^{d \times m}$ is the memory matrix of region r, and $\mathbf{M}_{r,i} \in \mathbb{R}^d$ denotes its i-th column. Note that, as the user's hometown behaviors are relatively long-term and could profile the user's inherent preference, we take the hometown behavior embedding here as the query.

Given the user u and region r, we query the crowd memory matrix of region r via the slot-level attention network in Eq. (8) and obtain user-specific memory embedding on region r as,

$$\mathbf{m}_r = \sum_{i=1}^{m} \pi_{i,r}^{s} \cdot \mathbf{M}_{r,i} \,. \tag{9}$$

Similarly, then we query user-specific memory of each region with hometown embedding $\mathbf{u}^h$ by another region-level attention network,

$$\mu_i^r = (\mathbf{a}^r)^\top \sigma\left(\mathbf{W}^r \left[\mathbf{u}^h \oplus \mathbf{m}_i\right]\right),$$
$$\pi_i^r = \frac{\exp(\mu_i^r)}{\sum_{j=1}^{|\mathcal{R}|} \exp(\mu_j^r)},$$

where $\mathbf{a}^r \in \mathbb{R}^d$ and $\mathbf{W}^r \in \mathbb{R}^{d\times 2d}$ are trainable parameters. π_i^r indicates the correlation between the user's inherent preference and the i-th region's user-specific memory embedding. Along this line, we can infer the user-specific crowd memory as a mixture of crowd memories of all regions. Formally, the inferred user-specific crowd memory embedding is given by

$$\widetilde{\mathbf{m}} = \sum_{i=1}^{|\mathcal{R}|} \pi_i^r \mathbf{m}_i. \tag{10}$$

4.3.3 ***Memory Reconstruction Block***. Given the user's ground truth out-of-town region r_o and the corresponding crowd memory matrix $\mathbf{M}_{r_o}$, we can obtain her/his factual crowd memory embedding $\mathbf{m}_{r_o}$ by Eq. (9). We devise a supervised reconstruction loss to bridge the inferred memory embedding and the factual memory embedding as follows,

$$\mathcal{L}_R = \sum_{u\in\mathcal{U}} \left\|\widetilde{\mathbf{m}} - \mathbf{m}_{r_o}\right\|_2^2 . \tag{11}$$

4.4 Metric Learning for Recommendation

Inspired by the great success of metric learning in knowledge graph embedding [1, 12, 25] and recommendation [11, 15, 31, 59], we implement a metric-based out-of-town recommender. In general, the triangle inequality in the embedding space can alleviate the severe data scarcity issues in out-of-town recommendation.

We first partition the embedding space by users' hometowns with a dynamic mapping mechanism. In particular, given an out-of-town user u, her/his related embeddings are encoded as

$$\begin{aligned} \mathbf{u}^{hr} &= \mathbf{T}_h^u \mathbf{u}^h, \\ \widetilde{\mathbf{v}}_i^r &= \mathbf{T}_{v_i}^u \widetilde{\mathbf{v}}_i \,, \end{aligned} \tag{12}$$

where $\mathbf{T}_h^u, \mathbf{T}_{v_i}^u \in \mathbb{R}^{d\times d}$ are two dynamic transition matrices generated for each user:

$$\mathbf{T}_h^u = \mathbf{R}_{r_h}\left(\mathbf{W}^d \mathbf{u}^h + \mathbf{b}^d\right)^\top,$$
$$\mathbf{T}_{v_i}^u = \mathbf{R}_{r_h}\left(\mathbf{W}^d \widetilde{\mathbf{v}}_i + \mathbf{b}^d\right)^\top,$$

where $\mathbf{W}^d \in \mathbb{R}^{d\times d}$ and $\mathbf{b}^d \in \mathbb{R}^d$ are trainable parameters, $\mathbf{R} \in \mathbb{R}^{d\times|\mathcal{R}|}$ is a learnable partition matrix, and $\mathbf{R}_{r_h}$ denotes the column vector in $\mathbf{R}$ w.r.t. region r_h. With such dynamic mapping operations, the embedding space is divided into several subspaces, making the metric more distinguishable.

Taking the factual crowd memory into account, we integrate the user's memory into her/his preference, and calculate the distance between such enriched preference and a POI v_i as follows,

$$e(u, v_i) = \left\|\left[\mathbf{u}^{hr} \oplus \mathbf{m}_{r_o}\right] - \widetilde{\mathbf{v}}_i^r\right\|_2^2 . \tag{13}$$

Then we learn the metric-based recommender under the Bayesian personalized ranking (BPR) criterion. Given a positive out-of-town POI $v_i \in \vec{c}_o$ visited by u, we sample a corrupted POI $v_j \notin \vec{c}_o$. Based on pairwise comparisons, the metric loss for pre-travel out-of-town recommendation is defined as,

$$\mathcal{L}_M = \sum_{u\in\mathcal{U}} \sum_{i\in\vec{c}_o} \sum_{j\notin\vec{c}_o} \left[\delta + e(u, v_i) - e(u, v_j)\right]_+ , \tag{14}$$

where $[x]_+ = max(x, 0)$ denotes the standard hinge loss and δ is the margin size.

4.5 Optimization and Recommendation

With Eqs. (6), (11) and (14) we can jointly optimize the composite loss function in an end-to-end fashion as below,

$$\mathcal{L} = \lambda_1 \mathcal{L}_C + \lambda_2 \mathcal{L}_R + \lambda_3 \mathcal{L}_M \,, \tag{15}$$

where λ_1, λ_2 and λ_3 are three hyper-parameters to balance the effects of different components in CAPTOR.

When carrying out the pre-travel out-of-town recommendation, we aim to recommend a list of out-of-town POIs to a new coming user u^*. Specifically, we first derive her/his mapped hometown behavior embedding $\mathbf{u}_*^{hr}$ (i.e., the inherent preference) and inferred memory embedding $\widetilde{\mathbf{m}}_*$ following Eqs. (7), (10) and (12). Then we select top-k out-of-town POIs whose mapped spatial-affined embeddings $\widetilde{\mathbf{v}}_i^r$ are closest to the concatenation of $\mathbf{u}_*^{hr}$ and $\widetilde{\mathbf{m}}_*$ as the recommendation list, where the distance is formulated as,

$$e(u^*, v_i) = \left\|\left[\mathbf{u}_*^{hr} \oplus \widetilde{\mathbf{m}}_*\right] - \widetilde{\mathbf{v}}_i^r\right\|_2^2 . \tag{16}$$

5 EXPERIMENTS

We conduct experiments on two real-world nationwide datasets to answer the following research questions:

- **Q1:** How does CAPTOR perform on the pre-travel out-of-town POI recommendation task compared to the baseline recommenders?
- **Q2:** How much do different components in CAPTOR contribute to the recommendation performance?
- **Q3:** How do the settings of parameters in CAPTOR affect the recommendation performance?
- **Q4:** How efficient is CAPTOR with regards to the computational cost?

5.1 Experimental Setups

5.1.1 ***Dataset***. We evaluate CAPTOR framework on two real-world nationwide travel behavior datasets: *CN* and *US*. The statistics

Table 1: Statistics of our datasets.

Statistics	Values	
	CN	*US*
# Users	53,330	3,889
# Regions	14	21
# POIs	14,813	24,389
# Check-ins	1,049,475	167,580
# Hometown check-ins	666,505	147,357
# Out-of-town check-ins	382,970	20,223

of these two datasets are given in Table 1. In *CN*, the travel behavior records were collected between 07/01/2019 and 12/31/2019 in China from an online map service, where the user check-ins were recorded. In *US*, the records between Apr. 2012 and Sep. 2013 in the United States were generated from a global-scale Foursquare check-in dataset published by Yang et al. [49, 50]. Afterward, for these two datasets, we chose the users who left check-ins in both hometown and out-of-town regions and reformed the check-in sequence of each user as an out-of-town travel record, i.e., $\tau = (u, \vec{c}_h, \vec{c}_o, r_h, r_o)$ (please refer to Definition 4).

To assure the data quality, we filtered out the POIs visited less than 5 times in *CN* and 2 times in *US*. Then, we removed regions where the amount of POIs is less than 500 in *CN* and 100 in *US*. Besides, we eliminated a user if her/his travel behavior $\tau = (u, \vec{c}_h, \vec{c}_o, r_h, r_o)$ reaches one of the following three criteria: (1) $|\vec{c}_h| < 5$ in *CN* and $|\vec{c}_h| < 4$ in *US*; (2) $|\vec{c}_o| < 4$ in *CN* and $|\vec{c}_o| < 2$ in *US*; (3) the frequency of (r_h, r_o) in the whole dataset is less than 90 in *CN* and 10 in *US*. In both two datasets, we randomly split users for each available hometown and out-of-town region pair, following the proportions: 80%, 10% and 10%, and eventually form a training set, a test set and a validation set with the same proportions. Note that all of the users in the two datasets are treated anonymized, guaranteeing the fairness of the evaluations.

5.1.2 ***Baselines***. We compare our CAPTOR with the following nine baselines:

1) A vanilla method:

- **PopRec** ranks POIs based on their popularity judged by the number of check-ins and recommends the top-K most popular out-of-town POIs.

2) Four general recommendation methods:

- **BPR** [32] employs Bayesian personalized ranking (BPR) to optimize the matrix factorization model with a pairwise ranking loss.
- **STA** [31] embeds the users and POIs into a transition space and takes spatiotemporal contexts as translation vectors.
- **TEMN** [59] integrates the topic model and memory network to learn the latent patterns and neighborhood-based features. A geographical module is further introduced to enhance the recommendation.
- **ST-LDA** [53] devises a probabilistic graphical model to learn the region-dependent personalized and crowd's preference for out-of-town POI recommendations.

3) Four next item recommendation methods:

- **SR-GNN** [45] incorporates graph models to consider the complex structure and transitions between items in the sequences for session-based recommendation.
- **GRU4Rec** [10] uses RNNs to model the sequential interactions for session-based recommendation.
- **TRAINOR** [47] applied neural topic model to discover the user's complex travel intention for out-of-town recommendation.
- **STAN** [28] learns the explicit spatiotemporal correlations within the POI check-in trajectory by a bi-attention system.

Particularly, STA, TEMN and STAN are proposed for POI recommendation, while ST-LDA and TRAINOR are tailored for out-of-town recommendation.

5.1.3 ***Implementation Details***. The dimensionality of all latent representations was set to 128. In SA-CRF, the coefficients α and β of the two types of energy functions were respectively set to 1 and 3 for *CN*, and both set to 1 for *US*. The number of iterations for updating the POI embedding (i.e., L) was set to 2. Moreover, the distance threshold ρ_d was set to 3 km for *CN* and 1 km for *US*. In CBMN, we set the number of memory slots m to 3 for *CN* and 10 for *US*. In the metric-based recommender, we set the margin size δ to 1. In the optimization stage, we set $\lambda_1 = \lambda_2 = \lambda_3 = 1$ in Eq. (15) and utilized Adam optimizer to train CAPTOR with an initial learning rate as 0.001 and an L2 regularization with weight 10^{-5}. Note that the evaluation was repeated three times using different initial parameters, after which the average was reported.

5.1.4 ***Evaluation Metrics***. We employ three widely-used metrics for recommendation to evaluate the effectiveness of the pre-travel out-of-town recommendation task, i.e., Hit Rate, Precision, and NDCG [43] at a cutoff k, denoted as *HR@k*, *Pre@k*, and *NDCG@k*, respectively. Particularly, following the common strategy in previous works [9], we treat a fixed number of negative POIs randomly sampled from the candidate set together with the positive out-of-town POIs as the final candidates for each user. In our experiments, we set the number of negative POIs as 100.

5.2 Overall Performance (Q1)

Table 2 shows the overall performance results of our proposed CAPTOR compared to all the baselines on *CN* and *US* datasets. As can be seen, CAPTOR achieves the best performance w.r.t *HR@k* and *Pre@k* on two datasets. The improvements of *HR@10* are 9.90% and 4.53% on *CN* and *US*, respectively, over the best baselines. In addition, CAPTOR beats all baselines w.r.t *NDCG@10* and *NDCG@15* on *CN* dataset with 6.88% and 5.90% improvements, respectively. Besides, we have the following observations.

PopRec recommends POIs to out-of-town users by merely considering the check-in popularity, which is not personalized but achieves better performances over some individualized baselines. The possible reason lies that most out-of-town travelers tend to pay visits to the landmarks with high prevalence (such as the Great Wall and the Statue of Liberty).

Among the personalized methods, we can clearly observe that next item recommendation methods exhibit considerable improvements over the general recommendation methods on most metrics,

Table 2: The overall recommendation performance comparison of CAPTOR and baselines. The best results are marked with bold face, and the second best results are underlined. (*significantly outperform at the 0.01 level, paired t-test)

Methods	*CN*						*US*					
	k=10			k=15			k=10			k=15		
	HR@k	*Pre@k*	*NDCG@k*	*HR@k*	*Pre@k*	*NDCG@k*	*HR@k*	*Pre@k*	*NDCG@k*	*HR@k*	*Pre@k*	*NDCG@k*
PopRec	0.5763	0.2670	0.3122	0.6752	0.2089	0.2629	0.3429	0.1617	0.2022	0.4102	0.1321	0.1732
BPR [32]	0.3162	0.1448	0.1651	0.3864	0.1181	0.1426	0.1226	0.0596	0.0662	0.1784	0.0583	0.0639
STA [31]	0.2846	0.1313	0.1272	0.4144	0.1278	0.1259	0.1366	0.0699	0.0718	0.1909	0.0646	0.0678
TEMN [59]	0.4892	0.2262	0.2449	0.5974	0.1842	0.2125	0.2479	0.1039	0.1299	0.3242	0.0905	0.1151
ST-LDA [53]	0.5349	0.2473	0.3471	0.6398	0.1976	0.3047	0.2400	0.1090	0.1526	0.3008	0.0943	0.1347
SR-GNN [45]	0.5994	0.2730	0.3290	0.7002	0.2125	0.2758	0.4003	0.1833	0.2364	0.4786	0.1472	0.2002
GRU4Rec [10]	0.6495	0.2942	0.3600	0.7312	0.2216	0.2965	0.4175	0.1915	**0.2462**	0.4971	0.1539	**0.2086**
TRAINOR [47]	0.6579	0.3038	0.3634	0.7553	0.2328	0.3022	0.3864	0.1714	0.2003	0.5141	0.1554	0.1829
STAN [28]	0.6676	0.3105	0.3723	0.7609	0.2363	0.3085	0.4136	0.1848	0.2269	0.5064	0.1527	0.1959
CAPTOR	**0.7337***	**0.3349***	**0.3979***	**0.8202***	**0.2501***	**0.3267***	**0.4364***	**0.1942***	0.2286	**0.5268**	**0.1606**	0.1980

whether the spatiotemporal features are included or not. One possible reason is that the former methods infer the user's embedding from the hometown behaviors, while the latter ones learn a sole representation for each user from both hometown and out-of-town check-ins. Such entangled representations may cause more confusion in pre-travel recommendation. This also demonstrates the necessity of building a separate and distinguishable feature space.

Within the general recommendation methods, nearly all POI recommendation methods that model spatiotemporal contexts outperform BPR, suggesting that spatial and temporal signals fuel the pre-travel recommendation. In particular, ST-LDA shows relatively better performance than other POI recommendation methods most of the time since interest and geographical drift, and conformity that occur in out-of-town scenario are considered. However, it models the crowd's preference implicitly from a probabilistic perspective, and the spatial affinity among POIs is not well captured. Regarding the next item recommendation methods, since STAN can learn the effect of time and distance intervals, it achieves the best performance among all the baselines on *CN* dataset. However, in pre-travel scenario, the distance interval varies significantly, which is difficult to capture for STAN, as well as TRAINOR. CAPTOR designs SA-CRF to model the relative spatial affinity among POIs. Therefore, the performance of CAPTOR is improved with satisfactory margins. Moreover, on *US* dataset, GRU4Rec even slightly outperforms CAPTOR w.r.t *NDCG@k*. One potential reason might be that the data sparsity issue of *US* dataset is more severe, such that the difficulty to fit this dataset climbs up for the large model, e.g., CAPTOR, in terms of the ranking metrics. But still, CAPTOR defeats GRU4Rec with satisfactory margins in terms of *HR@k* and *Pre@k*, which validates the effectiveness of modeling spatial affinity and crowd behavior patterns.

5.3 Ablation Study (Q2)

To evaluate the effectiveness of different modules, we compare CAPTOR with its variants:

1) CAPTOR-M removes the CBMN, such that the ability to store the crowd behaviors is deprived. **2) CAPTOR-C** removes the SA-CRF. As a result, the spatial affinity among POIs is neglected. **3) CAPTOR-D** uses the metric-based recommender without dynamic mapping mechanism, and the embeddings are performed in a universal metric space.

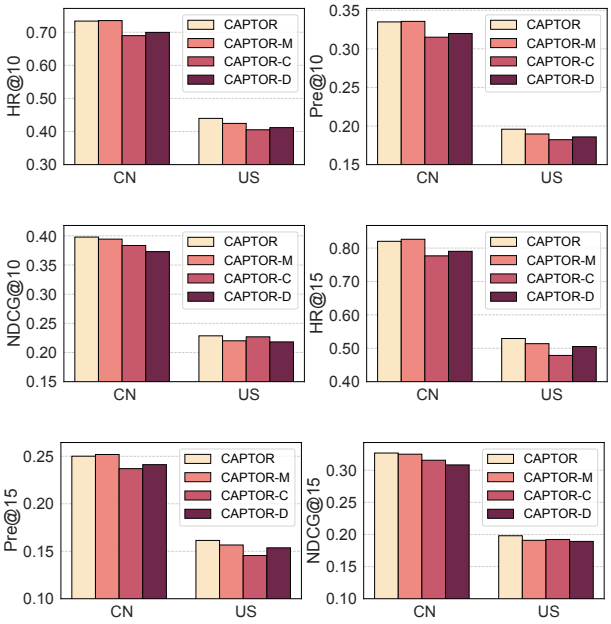
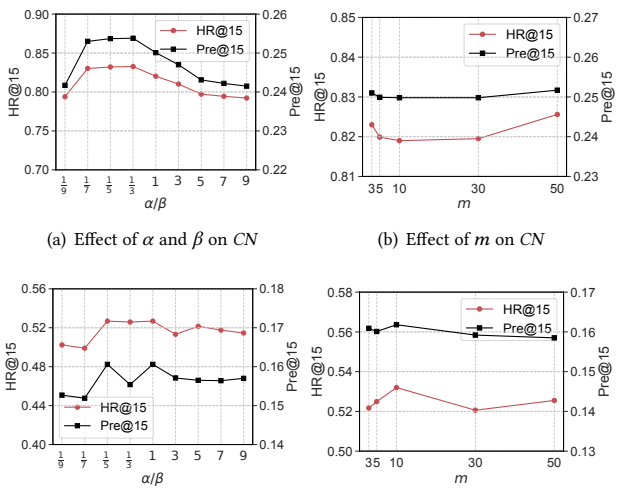


Figure 4: Contributions of different components in CAPTOR .

Fig. 4 shows the comparison results over the evaluation metrics. As can be seen, CAPTOR outperforms CAPTOR-C and CAPTOR-D on both datasets w.r.t all metrics, proving the critical roles of POI spatial affinity learning and separating metric spaces. Furthermore, CAPTOR beats CAPTOR-M consistently on *US* while having a slight fall on *CN* w.r.t *HR@k* and *Pre@k*. The reason could be that abundant travel behavior records in *CN* overfit the memory network.



(a) Effect of α and β on *CN* (b) Effect of m on *CN*

(c) Effect of α and β on *US* (d) Effect of m on *US*

Figure 5: Parameter sensitivity analysis on two datasets.

But still, the crowd behaviors are beneficial to the recommendation, especially when there are no sufficient records.

5.4 Parameter Sensitivity Analysis (Q3)

Fig. 5 presents the parameter sensitivity analysis of CAPTOR on two datasets.

The respective contributions of the unary and pairwise energy α and β. We analyze how the respective contributions of the unary energy and pairwise energy in Eq. (2) affect the recommendation performance by varying one parameter from 1 to 9 while fixing the other parameter to 1. For concise demonstration, we show the performance variation with the ratio of α and β. As shown in Fig. 5(a), whether α or β gets too large, the metrics decrease on both datasets. If α is much larger than β, the spatial affinity of POIs would not be well captured. Conversely, if β excessively dominates α, the inherent information of POIs could be neglected. Both of the above situations are not expected. Therefore, it is important to find a trade-off between α and β.

The number of memory slots m. To further investigate the impact of the number of memory slots in CBMN, we vary m from 3 to 50 on both datasets. It turns out that the performances peak when m is 50 on *CN* and 10 on *US*. Nevertheless, the metrics are relatively stable as m increases. The reason behind is that m mainly affects the dispersion of the crowd behavior information, whose impact is weakened after the aggregation by the hierarchical read block in CBMN.

5.5 Complexity Analysis (Q4)

We further analyze the computational complexity of CAPTOR. Generally, the computational cost mainly comes from SA-CRF, CBMN and the dynamic mapping mechanism. Given a user u, the complexity of SA-CRF is:

$$\mathcal{T}_{SA-CRF} = O\left(|\mathcal{V}|\, d + L\,|\mathcal{E}_s|\, d\right),$$

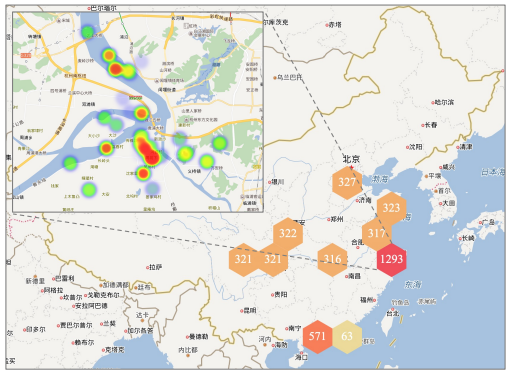


Figure 6: An illustration of the "coarse-to-fine" prototype strategy. The deeper the color, the higher the score.

where $|\mathcal{V}|$, L and $|\mathcal{E}_s|$ are the numbers of POIs, iterations of the spatial aggregation, edges in the POI spatial graph, respectively, and d is the dimensionality. Then, the complexity of CBMN is:

$$\mathcal{T}_{CBMN} = O\left((m + \kappa_u)d + m\,|\mathcal{R}|\, d^2\right),$$

where m and $|\mathcal{R}|$ are the numbers of memory slots and regions, respectively, and $\kappa_u = |\vec{c}_h| + |\vec{c}_o|$ denotes the length of the user's interacted POIs in both hometown and out-of-town. Note that in the recommendation stage, CBMN would have higher efficiency without the cost of the adaptive write operation. Afterward, we compute the complexity of the dynamic mapping mechanism as:

$$\mathcal{T}_{DM} = O\left(|\mathcal{V}|\, d^2\right).$$

Overall, given a user u, the total complexity of CAPTOR is the combination of $\mathcal{T}_{SA-CRF}$, $\mathcal{T}_{CBMN}$ and $\mathcal{T}_{DM}$ as:

$$\mathcal{T}_{CAPTOR} = O\left((|\mathcal{V}| + m\,|\mathcal{R}|)\, d^2 + (L\,|\mathcal{E}_s| + m + \kappa_u)\, d\right) \quad (17)$$

5.6 Discussion on Applicability

In practice, the pre-travel recommender could get overwhelmed due to the large searching space where a huge number of POIs act as candidates in different out-of-town regions. Essentially, the efficiency issues should be addressed when deploying the pre-travel out-of-town recommender system. Fortunately, the distinguishable embeddings learned by CAPTOR can benefit the searching space reduction for recommendation. As illustrated in Fig. 6, we could reduce the candidates with a "coarse-to-fine" prototype strategy.

To be specific, in SA-CRF, because POIs in different regions would have relatively distinct representations, i.e., the spatial affinity of regions, we could develop a heuristic algorithm to directly predict the probability of being a candidate region by aggregating the scores of sampled POIs in one region. In other words, the sampled POIs in a region can provide us with a "sketch" describing a user's preference towards this region. At the same time, the crowd behaviors in respective regions have been adaptively stored through CBMN. So we could incorporate the user embedding with the specific region crowd memory when recommending POIs in the corresponding region. We will work on implementing such

"coarse-to-fine" prototype for pre-travel recommendation in our future work.

6 CONCLUSION

In this paper, we studied the pre-travel recommendation problem for out-of-town users. We proposed a data driven framework CAPTOR to perform metric-based recommendation with distinguishable representations by capturing spatial affinity and crowd behavior simultaneously. Along this line, we devised a spatial-affined conditional random field to model the spatial affinity among POIs on a spatial graph. Then, the crowd behavior memory network was devised to store the crowd travel behaviors adaptively and offer clues to the travel decision-making. After that, the recommendation was carried out in the metric space, where embeddings of users and POIs are both distinguishable thanks to a dynamic mapping mechanism. Extensive experiments on two real-world datasets demonstrated the effectiveness of CAPTOR on pre-travel recommendation.

ACKNOWLEDGMENTS

This work was supported in part by the Project of Hetao Shenzhen-Hong Kong Science and Technology Innovation Cooperation Zone (HZQB-KCZYB-2020083), Natural Science Foundation of China (Grant No. 62072423), and Shanghai Youth Science and Technology Talents Sailing Program (No. 22YF1413700).

REFERENCES

[1] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. *Advances in neural information processing systems* 26 (2013).

[2] Buru Chang, Yonggyu Park, Donghyeon Park, Seongsoon Kim, and Jaewoo Kang. 2018. Content-Aware Hierarchical Point-of-Interest Embedding Model for Successive POI Recommendation.. In *IJCAI*. 3301–3307.

[3] Chen Cheng, Haiqin Yang, Michael R Lyu, and Irwin King. 2013. Where you like to go next: Successive point-of-interest recommendation. In *Twenty-Third international joint conference on Artificial Intelligence*.

[4] Jingtao Ding, Guanghui Yu, Yong Li, Depeng Jin, and Hui Gao. 2019. Learning from hometown and current city: Cross-city POI recommendation via interest drift and transfer learning. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 4 (2019), 1–28.

[5] Jie Feng, Yong Li, Chao Zhang, Funing Sun, Fanchao Meng, Ang Guo, and Depeng Jin. 2018. Deepmove: Predicting human mobility with attentional recurrent networks. In *Proceedings of the 2018 world wide web conference*. 1459–1468.

[6] Gregory Ference, Mao Ye, and Wang-Chien Lee. 2013. Location recommendation for out-of-town users in location-based social networks. In *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*. 721–726.

[7] Hongchang Gao, Jian Pei, and Heng Huang. 2019. Conditional random field enhanced graph convolutional neural networks. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 276–284.

[8] Dan Guo, Yang Wang, Peipei Song, and Meng Wang. 2020. Recurrent relational memory network for unsupervised image captioning. *arXiv preprint arXiv:2006.13611* (2020).

[9] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. 2017. Neural collaborative filtering. In *Proceedings of the 26th international conference on world wide web*. 173–182.

[10] Balázs Hidasi, Alexandros Karatzoglou, Linas Baltrunas, and Domonkos Tikk. 2015. Session-based recommendations with recurrent neural networks. *arXiv preprint arXiv:1511.06939* (2015).

[11] Cheng-Kang Hsieh, Longqi Yang, Yin Cui, Tsung-Yi Lin, Serge Belongie, and Deborah Estrin. 2017. Collaborative metric learning. In *Proceedings of the 26th international conference on world wide web*. 193–201.

[12] Guoliang Ji, Shizhu He, Liheng Xu, Kang Liu, and Jun Zhao. 2015. Knowledge graph embedding via dynamic mapping matrix. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. 687–696.

[13] Linlang Jiang, Jingbo Zhou, Tong Xu, Yanyan Li, Hao Chen, Jizhou Huang, and Hui Xiong. 2021. Adversarial Neural Trip Recommendation. *arXiv preprint arXiv:2109.11731* (2021).

[14] Wenxiang Jiao, Michael Lyu, and Irwin King. 2020. Real-time emotion recognition via attention gated hierarchical memory network. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 8002–8009.

[15] SeongKu Kang, Junyoung Hwang, Dongha Lee, and Hwanjo Yu. 2019. Semi-supervised learning for cross-domain recommendation to cold-start users. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*. 1563–1572.

[16] Yehuda Koren, Robert Bell, and Chris Volinsky. 2009. Matrix factorization techniques for recommender systems. *Computer* 42, 8 (2009), 30–37.

[17] Ankit Kumar, Ozan Irsoy, Peter Ondruska, Mohit Iyyer, James Bradbury, Ishaan Gulrajani, Victor Zhong, Romain Paulus, and Richard Socher. 2016. Ask me anything: Dynamic memory networks for natural language processing. In *International conference on machine learning*. PMLR, 1378–1387.

[18] John Lafferty, Andrew McCallum, and Fernando CN Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. (2001).

[19] Chao Li, Zhiyuan Liu, Mengmeng Wu, Yuchi Xu, Huan Zhao, Pipei Huang, Guoliang Kang, Qiwei Chen, Wei Li, and Dik Lun Lee. 2019. Multi-interest network with dynamic routing for recommendation at Tmall. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*. 2615–2623.

[20] Dichao Li and Zhiguo Gong. 2020. A Deep Neural Network for Crossing-City POI Recommendations. *IEEE Transactions on Knowledge and Data Engineering* (2020).

[21] Huayu Li, Yong Ge, Richang Hong, and Hengshu Zhu. 2016. Point-of-interest recommendations: Learning potential check-ins from friends. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. 975–984.

[22] Shuangli Li, Jingbo Zhou, Tong Xu, Hao Liu, Xinjiang Lu, and Hui Xiong. 2020. Competitive analysis for points of interest. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 1265–1274.

[23] Defu Lian, Yongji Wu, Yong Ge, Xing Xie, and Enhong Chen. 2020. Geography-aware sequential location recommendation. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*. 2009–2019.

[24] Defu Lian, Cong Zhao, Xing Xie, Guangzhong Sun, Enhong Chen, and Yong Rui. 2014. GeoMF: joint geographical modeling and matrix factorization for point-of-interest recommendation. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. 831–840.

[25] Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. 2015. Learning entity and relation embeddings for knowledge graph completion. In *Twenty-ninth AAAI conference on artificial intelligence*.

[26] Qiang Liu, Shu Wu, Liang Wang, and Tieniu Tan. 2016. Predicting the next location: A recurrent model with spatial and temporal contexts. In *Thirtieth AAAI conference on artificial intelligence*.

[27] Yong Liu, Wei Wei, Aixin Sun, and Chunyan Miao. 2014. Exploiting geographical neighborhood characteristics for location recommendation. In *Proceedings of the 23rd ACM international conference on conference on information and knowledge management*. 739–748.

[28] Yingtao Luo, Qiang Liu, and Zhaocheng Liu. 2021. STAN: Spatio-Temporal Attention Network for Next Location Recommendation. In *Proceedings of the Web Conference 2021*. 2177–2185.

[29] Tengfei Ma, Cao Xiao, Junyuan Shang, and Jimeng Sun. 2018. CGNF: Conditional Graph Neural Fields. (2018).

[30] Alexander Pritzel, Benigno Uria, Sriram Srinivasan, Adria Puigdomenech Badia, Oriol Vinyals, Demis Hassabis, Daan Wierstra, and Charles Blundell. 2017. Neural episodic control. In *International Conference on Machine Learning*. PMLR, 2827–2836.

[31] Tieyun Qian, Bei Liu, Quoc Viet Hung Nguyen, and Hongzhi Yin. 2019. Spatiotemporal representation learning for translation-based POI recommendation. *ACM Transactions on Information Systems (TOIS)* 37, 2 (2019), 1–24.

[32] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2012. BPR: Bayesian personalized ranking from implicit feedback. *arXiv preprint arXiv:1205.2618* (2012).

[33] Sainbayar Sukhbaatar, Arthur Szlam, Jason Weston, and Rob Fergus. 2015. End-to-end memory networks. *arXiv preprint arXiv:1503.08895* (2015).

[34] Huimin Sun, Jiajie Xu, Rui Zhou, Wei Chen, Lei Zhao, and Chengfei Liu. 2021. HOPE: a hybrid deep neural model for out-of-town next POI recommendation. *World Wide Web* (2021), 1–20.

[35] Qiaoyu Tan, Jianwei Zhang, Ninghao Liu, Xiao Huang, Hongxia Yang, Jingren Zhou, Xia Hu, et al. 2021. Dynamic memory based attention network for sequential recommendation. *arXiv preprint arXiv:2102.09269* (2021).

[36] Duyu Tang, Bing Qin, and Ting Liu. 2016. Aspect level sentiment classification with deep memory network. *arXiv preprint arXiv:1605.08900* (2016).

[37] Md Mehrab Tanjim, Congzhe Su, Ethan Benjamin, Diane Hu, Liangjie Hong, and Julian McAuley. 2020. Attentive sequential models of latent intent for next item recommendation. In *Proceedings of The Web Conference 2020*. 2528–2534.
[38] Waldo R Tobler. 1970. A computer movie simulating urban growth in the Detroit region. *Economic geography* 46, sup1 (1970), 234–240.
[39] Hao Wang, Yanmei Fu, Qinyong Wang, Hongzhi Yin, Changying Du, and Hui Xiong. 2017. A location-sentiment-aware recommender system for both home-town and out-of-town users. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*. 1135–1143.
[40] Hao Wang, Huawei Shen, Wentao Ouyang, and Xueqi Cheng. 2018. Exploiting POI-Specific Geographical Influence for Point-of-Interest Recommendation.. In *IJCAI*. 3877–3883.
[41] Suhang Wang, Yilin Wang, Jiliang Tang, Kai Shu, Suhas Ranganath, and Huan Liu. 2017. What your images reveal: Exploiting visual contents for point-of-interest recommendation. In *Proceedings of the 26th international conference on world wide web*. 391–400.
[42] Weiqing Wang, Hongzhi Yin, Ling Chen, Yizhou Sun, Shazia Sadiq, and Xiaofang Zhou. 2017. ST-SAGE: A spatial-temporal sparse additive generative model for spatial item recommendation. *ACM Transactions on Intelligent Systems and Technology (TIST)* 8, 3 (2017), 1–25.
[43] Markus Weimer, Alexandros Karatzoglou, Quoc Le, and Alex Smola. 2007. Cofirank-maximum margin matrix factorization for collaborative ranking. In *Advances in Neural Information Processing Systems, 21st Annual Conference on Neural Information Processing Systems 2007*. 222–230.
[44] Jason Weston, Sumit Chopra, and Antoine Bordes. 2014. Memory networks. *arXiv preprint arXiv:1410.3916* (2014).
[45] Shu Wu, Yuyuan Tang, Yanqiao Zhu, Liang Wang, Xing Xie, and Tieniu Tan. 2019. Session-based recommendation with graph neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 346–353.
[46] Guo-Sen Xie, Huan Xiong, Jie Liu, Yazhou Yao, and Ling Shao. 2021. Few-shot semantic segmentation with cyclic memory network. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 7293–7302.
[47] Haoran Xin, Xinjiang Lu, Tong Xu, Hao Liu, Jingjing Gu, Dejing Dou, and Hui Xiong. 2021. Out-of-Town Recommendation with Travel Intention Modeling. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 4529–4536.
[48] Bingbing Xu, Huawei Shen, Bingjie Sun, Rong An, Qi Cao, and Xueqi Cheng. 2021. Towards Consumer Loan Fraud Detection: Graph Neural Networks with Role-Constrained Conditional Random Field. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 4537–4545.
[49] Dingqi Yang, Daqing Zhang, Longbiao Chen, and Bingqing Qu. 2015. Nationtelescope: Monitoring and visualizing large-scale collective behavior in lbsns. *Journal of Network and Computer Applications* 55 (2015), 170–180.
[50] Dingqi Yang, Daqing Zhang, and Bingqing Qu. 2016. Participatory cultural mapping based on collective behavior data in location-based social networks. *ACM Transactions on Intelligent Systems and Technology (TIST)* 7, 3 (2016), 1–23.
[51] Hongzhi Yin, Bin Cui, Yizhou Sun, Zhiting Hu, and Ling Chen. 2014. LCARS: A spatial item recommender system. *ACM Transactions on Information Systems (TOIS)* 32, 3 (2014), 1–37.
[52] Hongzhi Yin, Bin Cui, Xiaofang Zhou, Weiqing Wang, Zi Huang, and Shazia Sadiq. 2016. Joint modeling of user check-in behaviors for real-time point-of-interest recommendation. *ACM Transactions on Information Systems (TOIS)* 35, 2 (2016), 1–44.
[53] Hongzhi Yin, Xiaofang Zhou, Bin Cui, Hao Wang, Kai Zheng, and Quoc Viet Hung Nguyen. 2016. Adapting to user interest drift for poi recommendation. *IEEE Transactions on Knowledge and Data Engineering* 28, 10 (2016), 2566–2581.
[54] Jia-Dong Zhang and Chi-Yin Chow. 2015. Geosoca: Exploiting geographical, social and categorical correlations for point-of-interest recommendations. In *Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval*. 443–452.
[55] Jia-Dong Zhang, Chi-Yin Chow, and Yanhua Li. 2014. Lore: Exploiting sequential influence for location recommendations. In *Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. 103–112.
[56] Weijia Zhang, Hao Liu, Fan Wang, Tong Xu, Haoran Xin, Dejing Dou, and Hui Xiong. 2021. Intelligent electric vehicle charging recommendation based on multi-agent reinforcement learning. In *Proceedings of the Web Conference 2021*. 1856–1867.
[57] Pengpeng Zhao, Anjing Luo, Yanchi Liu, Fuzhen Zhuang, Jiajie Xu, Zhixu Li, Victor S Sheng, and Xiaofang Zhou. 2020. Where to go next: A spatio-temporal gated network for next poi recommendation. *IEEE Transactions on Knowledge and Data Engineering* (2020).
[58] Pengpeng Zhao, Chengfeng Xu, Yanchi Liu, Victor S Sheng, Kai Zheng, Hui Xiong, and Xiaofang Zhou. 2019. Photo2Trip: Exploiting visual contents in geo-tagged photos for personalized tour recommendation. *IEEE Transactions on Knowledge and Data Engineering* (2019).
[59] Xiao Zhou, Cecilia Mascolo, and Zhongxiang Zhao. 2019. Topic-enhanced memory networks for personalised point-of-interest recommendation. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 3018–3028.
[60] Nengjun Zhu, Jian Cao, Xinjiang Lu, and Hui Xiong. 2021. Learning a Hierarchical Intent Model for Next-Item Recommendation. *ACM Transactions on Information Systems (TOIS)* 40, 2 (2021), 1–28.

A APPENDIX

A.1 Discovery of User's Hometown

Basically, it is not easy to directly know a user's hometown. Some previous works [4] simply choose the region where a user u has the largest number of check-ins as her/his hometown. Such straightforward method may lead to inaccurate hometown decisions. For example, if u left a number of check-ins in one region a long time ago, we can hardly say that the region is her/his hometown since this evidence have low activeness and confidence. Therefore, we design a dedicated way to decide u's hometown. In particular, given u's check-in sequence in region r chronologically ordered:

$$\vec{c}_r = [(u, t_1, v_1), (u, t_2, v_2), \cdots, (u, t_n, v_n)],$$

we view these check-ins as the hometown evidence to support $r_h = r$ (i.e. region r is u's hometown). Intuitively, u's recent check-ins as hometown evidence should be paid more attention than the records observed long ago. Thus, given a check-in c_i in $\vec{c}_r$, we use an exponential decay function to determine the importance:

$$\begin{aligned} \omega(c_i) &= \Theta^{\frac{\Delta t_i}{T}}, \\ \Delta t_i &= |t_0 - t_i|, \end{aligned} \tag{18}$$

where t_0 is the observation time for determining u's hometown, T is the width of the time window for collecting hometown evidence, and Θ is the base of exponential function.

Furthermore, earlier check-ins are as evidential as current ones in some cases, especially when the observed check-ins are limited. Inspired by this, we set the base Θ in Eq. (18) as follows:

$$\Theta = \Theta(\Delta t_n) = \begin{cases} 1, & 0 \le \Delta t_n < \Delta t', \\ \frac{\Delta t'' - \Delta t_n}{\Delta t'' - \Delta t'}, & \Delta t' \le \Delta t_n \le \Delta t'', \end{cases}$$

where $\Delta t_n = |t_0 - t_n|$, and $\Delta t'$ together with $\Delta t''$ are two scheduled thresholds. In other words, given a user, if her/his latest check-in activity in region r happens very recently, say, just three days ago, then her/his other earlier check-ins in r are treated as important as the latest one for determining her/his hometown.

Afterward, the supporting value for region r being considered as u's hometown can be calculated as below:

$$\phi(u, \vec{c}_r) = \sum_{i=1}^{n} \frac{\omega(c_i)}{N + \epsilon},$$

where N is the number of all check-ins generated by u, and ϵ is a constant to avoid the sum of weights being divided by zero. Finally, a user's hometown can be determined by selecting the region with the highest supporting value ϕ.