GeoMamba: Towards Multi-granular POI Recommendation with Geographical State Space Model

Yifang Qin[1], Jiaxuan Xie[2], Zhiping Xiao[3*], Ming Zhang [1*]

[1]State Key Laboratory for Multimedia Information Processing, School of Computer Science, PKU-Anker LLM Lab, Peking University
[2]School of Earth and Space Sciences, Peking University
[3]Paul G. Allen School of Computer Science and Engineering, University of Washington
qinyifang@pku.edu.cn, 2100012427@stu.pku.edu.cn, patxiao@uw.edu, mzhang_cs@pku.edu.cn

Abstract

Point-of-Interest (POI) recommendation plays an important role in a wide range of location-based social network applications, aiming to accurately predicting users' next visits based on their historical check-in records. Previous efforts have primarily focused on the modifications of existing sequential models, neglecting the fact that POI visiting sequences typically involve continuous state transformation of geographical and intention signals. Additionally, the diverse time span between check-ins require the model to properly recognize user's multi-granular preference. While recent advances of State Space Model (SSM) have revealed their potential in handling intricate temporal signals, we propose a state-based model that is tailored for spatio-temporal POI sequences. On top of traditional SSMs that are typically limited to linear sequences like Mamba, we propose GeoMamba, which customizes the model states to accommodate the spatio-temporal sequences, especially fitting for POI recommendations. Specifically, while the approximation operator HiPPO sets the foundation of linear SSMs, we introduce a novel GaPPO operator that extends the model's state space into graph-represented geographical domains. This innovation allows us to construct locational SSM encoders that seamlessly integrate users' spatio-temporal characteristics. The sequence-aware outputs of GeoMamba are further processed to generate multi-scale behavior representations. Extensive experimental results illustrate the superiority of GeoMamba over several state-of-the-art baselines.

Introduction

The flourishing location-based services and applications have generated great demand for more precise and personalized Point-of-Interest (POI) recommendations. Since users have left massive check-in records on location-based platforms like Foursquare and Gowalla (Cho, Myers, and Leskovec 2011; Yang et al. 2014), the goal of POI recommender systems is to maximize the utility of such spatio-temporal trajectories. A well-designed recommendation model can significantly improve user experience by suggesting POIs that align with both their geographical preferences and personal interests (Cheng et al. 2013; Wang et al. 2018; Luo, Liu, and Liu 2021; Li et al. 2021).

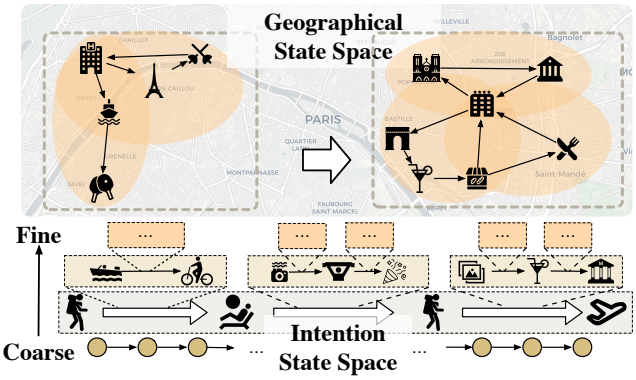


Figure 1: An example of multi-scale state transitions.

In previous efforts to develop sequential POI recommendation models, a variety of architectural approaches have been employed to handle the spatio-temporal visit sequences. Early attempts proposed to incorporate the geographical correlations of POIs into traditional recommendation techniques, such as collaborative filtering (Ye et al. 2011) and matrix factorization (Lian et al. 2014; Feng et al. 2015). Graph-embedding-based models (Wang et al. 2015; Xie et al. 2016) recognized the advantages of graph-structured data to reflect geographical influences. Other studies integrated distances and time slots into existing matching models (Wang et al. 2018) or recurrent networks (Sun et al. 2020). More recent researches have turned to utilize cutting-edge sequence-model structures, including self-attention (Luo, Liu, and Liu 2021), graph neural networks (Wang et al. 2022a), and diffusion models (Qin et al. 2023b).

Despite the advancements in different model architectures, they mostly rely on the integration of explicit correlations between check-ins, such as distance and time intervals, into existing sequence models. However, it is noteworthy that apart from these, the implicit state transitions of user's intentions are inherently neglected from these architectures. Additionally, the broad time spans between check-ins require models to effectively recognize user's preference shifts from a variety of time scales. For instance, as illustrated in Figure 1, in a user's tour during Paris Olympics, their preference evolve across different states over multiple time scales. In the early days of the trip, the visits are

*Correspondence authors.

mostly concentrated around the Seine River in western Paris to attend Olympic events. Later, their footprints shift towards the eastern part of Paris, focusing on cultural attractions such as Place de la Bastille. Throughout the trip, we can observe a clear state transformation from an Olympic spectator to a cultural explorer. From a more fine-grained perspective, the locational regions and personal states have been constantly changing over time as well. These observations suggest the natural needs for models that can extract informative state representations across multiple granularities. To address these challenges, we propose to turn to the state-space models (SSMs), which explicitly model the hidden state during time gaps, to seek a potential solution for effectively modeling POI visit sequences.

As a neural extension of the polynomial projection operator (HiPPO) for time signals, a series of SSMs have exhibited promising capability of approximating long signals with linear complexity (Gu et al. 2021). Variants of SSM have proven effective in handling complex and intricate sequences, such as natural language (Poli et al. 2023) and time series (Gu, Goel, and Re 2022). However, existing applications of SSMs are typically limited to discrete and flat sequential signals. In the context of POI recommendation, a user's visit sequences are influenced not only by sequential correlations, but also the spatial dependencies and temporal evolution across different scales. This difference raises requirement for the model to integrate locational information into the evolution within a multi-granular states space.

To address the aforementioned challenges, we propose to extend the existing SSMs and construct specific state space architectures tailored to spatio-temporal interaction data. Based on HiPPO and graph signal processing theory, we derive the Graph-autoregressive Polynomial Projection Operator (GaPPO). Compared with HiPPO, GaPPO encodes input signals in Fourier state spaces based on graph ARMA filters, enabling the processing of graph-structured temporal signals with a more expressive linear filter. GaPPO-based SSM are well-suited to capturing graph-represented geographical relationship of POIs while encoding spatio-temporal historical behavior into hidden state space. To allow model to encode intention states at varying levels of granularity, we sample multiple observation points from the state sequence and employ a Matryoshka representation learning technique to ensure multi-scale perception within the SSM. Based on these innovations, we construct a novel GeoMamba framework which jointly models spatio-temporal characteristics and continuous state transitions within the multi-granular hidden space. The main contribution can be summarized as:

- We introduce a novel GaPPO-based geographical SSM GeoMamba, by extending the sequence state space into graph signal space. GeoMamba effectively models the dynamic state transitions of users in POI recommendation.
- The proposed GeoMamba achieves multi-granular modeling of visit sequences through its fine-grained state space. Using the Matryoshka representation learning technique, the model is optimized to generate informative multi-scaled state representations of visiting sequences.
- Experimental results validate the superiority of GeoMamba in making next POI recommendations and capturing the multi-scaled spatio-temporal state of users, compared with state-of-the-art methods.

Related Work

Next POI Recommendation

The core of next POI recommendation is to precisely predict personalized POIs for each user based on their visiting history and locational information (Zhao et al. 2016; Cheng et al. 2013). To exploit the historical visits, a large branch of researches have adopted variations of existing sequence models to recurrently process sequential data. CARA (Manotumruksa, Macdonald, and Ounis 2018) and LSTPM (Sun et al. 2020) propose to integrate distance and neighborhood information into recurrent neural network (RNN) structures. ARNN (Guo et al. 2020) incorporates the geographical influence using a customized attentional RNN model. STAN (Luo, Liu, and Liu 2021) chooses to embrace the spatio-temporal transformer architecture. SGRec (Li et al. 2021) retrieves the nearby POIs as an enhancement for visit sequence graph. Approaches like STGCN (Han et al. 2020) and HMT-GRN (Lim et al. 2022) represent POI as nodes on geographical graphs. GSTN (Wang et al. 2022b) and DisenPOI (Qin et al. 2023a) process geographical and sequential features with two graph encoders respectively. However, overlooking the multi-grained state can potentially lead to suboptimal performance, calling for a model that explicitly encodes the multi-scaled hidden state of POI sequences.

State Space Sequence Models

Inspired by the projection operators in linear signal systems, State Space Models (SSMs) are initially proposed for time signal processing (Gu et al. 2020, 2021). Subsequent studies include S4 (Gu, Goel, and Re 2022), Hyena (Poli et al. 2023) and LTC (Hasani et al. 2023) further explored the possibilities of SSMs in natural language generation. Mamba (Gu and Dao 2023; Dao and Gu 2024) extended the use of SSMs to the construction of large language models. Given the demonstrated potential of SSMs across a variety of downstream tasks, we seek to leverage the capability of SSM to precisely capture state transitions in POI recommendation.

Preliminary

Differs from traditional user-item recommendation, users' visit histories of POIs involve both spatial correlations and sequential patterns. To formulate, given the POI set $\mathcal{P} = \{p_1, p_2, ..., p_{|\mathcal{P}|}\}$ and user set $\mathcal{U} = \{u_1, u_2, ..., u_{|\mathcal{U}|}\}$, a locational distance metric is defined as $\varphi : \mathcal{P} \times \mathcal{P} \to \mathbb{R}$ to represent the spatial correlation between two given POIs. Common choices of $\varphi(\cdot, \cdot)$ include step functions with distance thresholds (Wang et al. 2022a) or exponential decay functions of distance (Qin et al. 2023a). A location-based POI graph $\mathcal{G} = (\mathcal{P}, \mathcal{E}, A)$ with edge set $\mathcal{E}$ and adjacency matrix Adj is then constructed based on a cutoff threshold $d_{\max}$:

$$\mathcal{E} = \{e_{ij} | distance(p_i, p_j) < d_{\max}\}$$
$$Adj_{ij} = \begin{cases} \varphi(p_i, p_j), & e_{ij} \in \mathcal{E} \\ 0, & else \end{cases} \quad (1)$$

For each user $u \in \mathcal{U}$, its visiting sequence is represented as $\mathcal{S}_u = \{(p_1^u, \tau_1^u), (p_2^u, \tau_2^u), ..., (p_n^u, \tau_n^u)\}$ with n historical visits and timestamps. The objective of POI recommendation is to predict possible next-to-visit POIs given $\mathcal{S}_u$.

Methodology

In this section, we will elucidate the proposed GeoMamba, a specialized SSM designed to capture the fine-grained multi-scale characteristics of users' spatio-temporal visits. As illustrated in Figure 2, GeoMamba utilizes a graph state space model encoder alongside an enhanced Mamba encoder to generate informative spatio-temporal representations of input visit sequences at various scales. These multi-scale representations are further optimized using the Matryoshka Representation Learning technique, which fully exploits the benefits of the multi-scale representations.

Geographical Graph State Space Encoder

Graph-autoregressive Polynomial Projection Operator While the SSM-based models has exhibited significant potential in processing sequence data (Gu, Goel, and Re 2022), the key of SSMs' expressiveness lies in their structural connection with high-order polynomial projection operators, or HiPPOs, which learn optimal approximations of input signals. Given a temporal signal $u(t) \in \mathbb{R}^d$ at specific time step $t \in \mathbb{Z}^+$, the prediction process of the discretized HiPPO with learnable parameters is described as:

$$\begin{cases} x_t = \bar{A}x_{t-1} + \bar{B}u_t \\ y_t = \bar{C}x_t \end{cases} \tag{2}$$

where $\bar{A}, \bar{B}, \bar{C} \in \mathbb{R}^{d\times d}$ are discretized matrices parameterized by corresponding learnable network parameters. x_t represents the sequence state and y_t denotes the model output.

A common practice is to use the model output y_{t+1} from Equation 2 for downstream tasks, *i.e.* the next POI prediction in our case. However, such a straightforward solution inherently overlooks the geographical correlation between different samples of visiting signals. Moreover, the simple discretization of a continuous process causes information loss regarding the temporal effects during user's visits as well.

To address these challenges, we propose to extend HiPPO into a graph-based variant, namely Graph-autoregressive Polynomial Projection Operator (GaPPO), to incorporate POI-POI correlations. Inspired by previous works on graph ARMA filters (Isufi et al. 2016), we extend autoregressive update in Equation 2 with graph filters. Given the one-hot form of POI visiting signal $u_t \in \{0,1\}^{|\mathcal{P}|}$, the GaPPO for predicting signal y_t is formulated as:

$$\begin{cases} u_t = \text{One-hot}(p_t^u) \\ x_t = \sum_{k=1}^{K} \varphi_k \mathcal{H}_A^{(k)}(x_{t-k}) + \sum_{j=0}^{J} \psi_j \mathcal{H}_B^{(j)}(u_{t-j}) \\ y_t = \mathcal{H}_C(x_t) \end{cases} \tag{3}$$

where $x_t, y_t \in \mathbb{R}^{|\mathcal{P}|}$ are graph signals, $\mathcal{H}_{\{A,B,C\}}$ are graph filters defined on POI graph $\mathcal{G}$ with finite impulse response (FIR), φ_k, ψ_j are weight coefficients of the filter outputs. Compared with HiPPO, the proposed GaPPO is able to incorporate the geographical influence during the evolution of visiting signals by introducing $\mathcal{G}$ into the signal system. Additionally, while both HiPPO and GaPPO are formulated in infinite impulse response (IIR) form, GaPPO is able to depict the signal fluctuations over a broader historical range of input. For more detailed analyses of GaPPO and its comparison with HiPPO, please refer to the Appendix.

GaPPO-based State Space Encoder Building on the previously derived GaPPO for continuous graph temporal signals, we extend it into a Graph State Space Model (GSSM). It is noteworthy that directly applying $\mathcal{H}_A$ and $\mathcal{H}_B$ as regular polynomial filters would result in extremely large hidden states with computational complexity of $\mathcal{O}(|\mathcal{P}|^2)$ and thus makes the output intractable to compute. Therefore, we leverage Graph Fourier Transform (GFT) to accelerate the signal processing in spectral domain. Specifically, the GFT maps signals into spectral space, modeled as:

$$L_{sym} = I - D^{-1/2}AdjD^{-1/2} = U^T \Lambda U, \tag{4}$$

where L_{sym} is the Laplacian matrix of $\mathcal{G}$, D represents the degree matrix, and $\Lambda = diag\{\lambda_1, ..., \lambda_{|\mathcal{P}|}\}$ is the diagonal eigenvalue matrix that satisfies $\lambda_1 < \lambda_2 < ... < \lambda_{|\mathcal{P}|}$. As previous researches have highlighted the importance of preserving low-pass in graph-based networks (Kipf and Welling 2017), we reduce the size of the transformed signal by only retaining the smallest d components of U, *i.e.* $\hat{x}_t = U_{:d}x_t$. By transferring the GaPPO into a reduced spectral space, we are able to reduce the computational complexity:

$$\begin{cases} \hat{x}_t = \sum_{k=1}^{K} \varphi_k(t)\mathcal{H}_A^{(k)}(\hat{x}_{t-k}) + \sum_{j=0}^{J} \psi_j(t)\mathcal{H}_B^{(j)}(\hat{u}_{t-j}), \\ \hat{y}_t = \mathcal{H}_C(\hat{x}_t), \end{cases} \tag{5}$$

and the output could be recovered to graph signal via the inverse transformation $y_t = U_{:d}^T \hat{y}_t$.

We construct the graph filters $\mathcal{H}_{\{A,B,C\}}$ based on the spectral GaPPO following the zero-order hold (ZOH) discretization of Mamba (Gu and Dao 2023):

$$\mathcal{H}_A^{(k)} = \exp(\Delta A^{(k)}), \mathcal{H}_B^{(j)} = \Delta B^{(j)}, \mathcal{H}_C = C, \tag{6}$$

where $A^{(k)} \in \mathbb{R}^{d\times d}$ is trainable parameter, $B^{(j)}, C, \Delta$ are input dependent parameters. When the input signal is $\hat{u}(t)$ at time step t, the filters are parameterized by:

$$\begin{aligned} B^{(j)} = \text{Linear}_B(\hat{u}_{t-j}),\ C = \text{Linear}_C(\hat{u}_t), \\ \Delta = \text{Softplus}(\Delta_0 + \text{Linear}_\Delta(\hat{u}_t)). \end{aligned} \tag{7}$$

To model the direct influence of historical states, the weight coefficients are defined by the exponential decay of temporal gaps between successive visits:

$$\{\varphi, \psi\}_k(t) = \begin{cases} e^{-\gamma_{\{\varphi,\psi\}}(\tau_{t-k+1}^u - \tau_{t-k}^u)}, & k < t \\ 0 & \textbf{otherwise} \end{cases} \tag{8}$$

where $\gamma_\varphi, \gamma_\psi > 0$ are hyperparameters that control the decay effect. With these parameters defined, we can now obtain the model output $y_t \in \mathcal{R}^{|\mathcal{P}|}$ for arbitrary given visiting sequences and corresponding POI graphs.

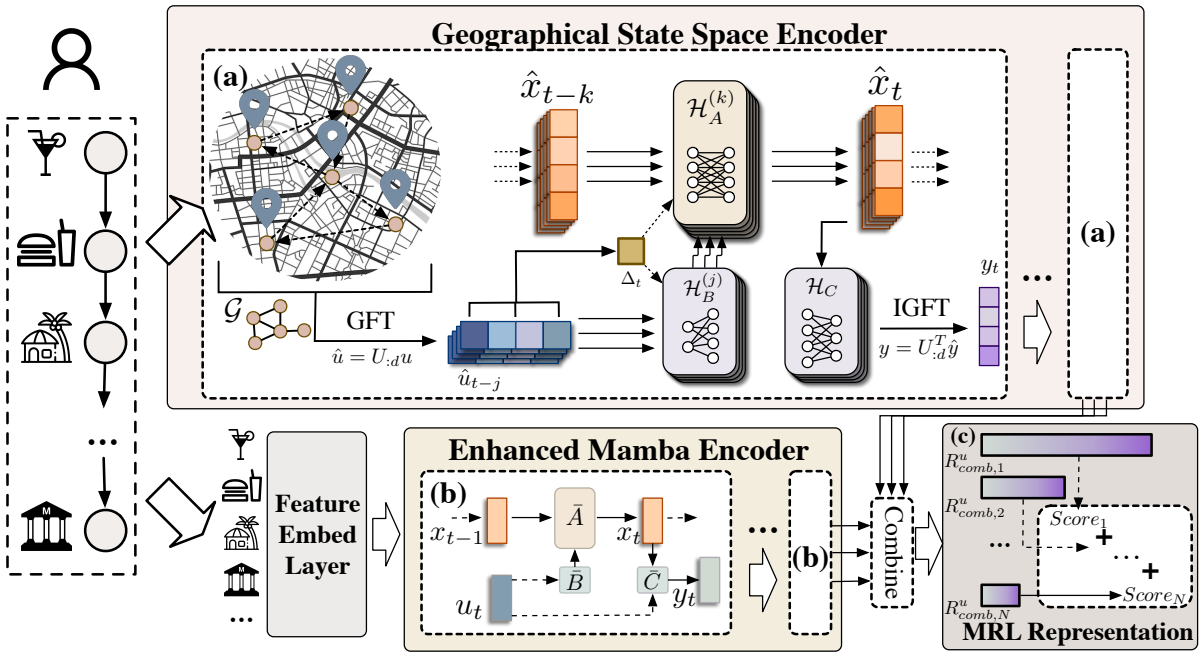


Figure 2: The overall illustration of GeoMamba. The GeoMamba model is composed of several key components include: (a) GaPPO-based geographical state space encoder model, (b) enhanced Mamba-based feature encoder model, and (c) the Matryoshka representation learning module that optimizes the model and extracts fine-grained sequence representations.

Feature-enhanced Mamba Sequence Encoder

While the GSSM encoder effectively captures the geographical correlation within visiting signals, it is constrained to the one-hot encoded graph signal space. However, POIs in real world possess more than just locational features. The rich raw features of POI, such as genre, opening hours, and price levels can not be fully described by geographical connections alone. To address this issue, we employ an additional enhanced Mamba (Gu and Dao 2023) encoder as the sequence encoder to capture the contextual correlation within POI sequences. Specifically, after extracting features, the user sequence are mapped through the embedding layer:

$$e_t^u = \text{Embed}(\text{Feature}(p_t^u)) \in \mathbb{R}^d \,. \tag{9}$$

While the vanilla Mamba layers are SSMs parameterized in the same way as in Equation 6 and 7, making all update parameter input-dependent may lead to over-fitting issue of model, especially given that the input visiting sequences are typically much shorter than those in natural language processing. Given this, we propose the enhanced Mamba layer that re-introduces input-independent parameterization into Mamba layers. To formulate, the $\bar{B}$ and $\bar{C}$ of enhanced Mamba encoder at time step t are obtained from:

$$\bar{B} = \text{Linear}'_B(e_t^u) + B_0,\ \bar{C} = \text{Linear}'_C(e_t^u) + C_0 \,, \tag{10}$$

the trainable $B_0, C_0 \in \mathbf{R}^{d\times d}$ are made data-independent to capture the universal feature throughout sequences.

Multi-scale Representation Learning

Inter-layer Propagation of GeoMamba Having defined a pair of SSMs that process visiting signals for processing visiting signals in the previous sections, we now explore stacking these layers into deeper models to fully leverage the potential of the proposed SSMs.

Based on the layer propagation mechanisms defined in previous sections, we can obtain the layer-wise representations from both SSM layers, formulated as:

$$\begin{aligned} &y_{g,t}^{(l)} = \text{G-Layer}_l(u_{g,:}^{(l)})_t,\ y_{f,t}^{(l)} = \text{F-Layer}_l(u_{f,:}^{(l)})_t \,, \\ &u_{g,t}^{(0)} = U_{:d} \cdot \text{One-hot}(p_t^u),\ u_{f,t}^{(0)} = e_t^u \,, \end{aligned} \tag{11}$$

where G-Layer and F-Layer are the GaPPO-based SSM layer and the feature-enhanced Mamba layer correspondingly. For inter-layer propagation, we adapt the gated mechanism from Mamba (Gu and Dao 2023) and implement as:

$$\begin{cases} u_{s,t}^{(l)} = \sigma(\text{Conv1D}(y_{s,:}^{(l-1)})_t) \,, \\ y_{s,t}^{(l+1)} = \text{Layer}_l(u_{s,:}^{(l)})_t \otimes \sigma(u_{s,t}^{(l)}) \,, \end{cases} \tag{12}$$

where $s \in \{g, f\}$ indicates whether is propagating in the geographical or enhanced SSM encoder. Layer_l denotes the corresponding SSM layer, σ is the activation function, and Conv1D is the causal convolution. By forwarding these hidden representations by layer, we are able to stack SSM layers to construct deep encoders for GeoMamba with L layers.

Multi-scale Representations with Observations From previous sections, we have derived the sequential output of a geographical GSSM encoder, as well as a feature-enhanced Mamba encoder. The pair of the final representations $y_{s,t}^{(L)}$ effectively captures the historical visiting signals up to the given observation time t. Since previous researches

Data	Metrics	LightGCN	HCCF	GeoIE	SGRec	DRAN	LSTPM	STAN	Diff-POI	GeoMamba	Improv.
Singapore	Recall@2	0.1606	0.2214	0.2521	0.3058	0.3086	0.2704	0.2634	0.3311	**0.3840***	16.0%
	Recall@5	0.2144	0.2889	0.2867	0.3507	0.3554	0.3253	0.2940	0.4075	**0.4265***	4.67%
	Recall@10	0.2419	0.3474	0.3162	0.3884	0.3921	0.3791	0.3280	0.4201	**0.4581***	9.05%
	NDCG@2	0.1720	0.2038	0.2428	0.2274	0.2972	0.2610	0.2831	0.3178	**0.3728***	17.3%
	NDCG@5	0.1789	0.2341	0.2583	0.2697	0.3175	0.2697	0.2995	0.3390	**0.3924***	15.8%
	NDCG@10	0.1892	0.2530	0.2679	0.2916	0.3297	0.2749	0.2892	0.3526	**0.3993***	15.4%
Tokyo	Recall@2	0.3917	0.4463	0.4975	0.5091	0.5225	0.5029	0.5105	0.5862	**0.6329***	7.97%
	Recall@5	0.4473	0.5198	0.5408	0.5488	0.5570	0.5513	0.5489	0.6265	**0.6640***	5.99%
	Recall@10	0.4936	0.5744	0.5726	0.6173	0.6210	0.5795	0.6167	0.6555	**0.6874***	4.87%
	NDCG@2	0.4207	0.4231	0.4842	0.4715	0.5089	0.4724	0.4990	0.5748	**0.6224***	7.61%
	NDCG@5	0.4354	0.4562	0.5037	0.4905	0.5390	0.4881	0.5264	0.5930	**0.6365***	7.34%
	NDCG@10	0.4403	0.4739	0.5141	0.5112	0.5602	0.4962	0.5554	0.6023	**0.6441***	6.94%
New York City	Recall@2	0.3789	0.4912	0.5655	0.5734	0.5859	0.5754	0.6027	0.6501	**0.6893***	6.03%
	Recall@5	0.4363	0.5517	0.5994	0.6175	0.6253	0.6020	0.6358	0.6725	**0.7011***	4.25%
	Recall@10	0.4519	0.5908	0.6251	0.6424	0.6478	0.6312	0.6533	0.6873	**0.7091***	3.17%
	NDCG@2	0.3819	0.4701	0.5629	0.5490	0.5702	0.5524	0.5887	0.6411	**0.6846***	6.79%
	NDCG@5	0.3867	0.4972	0.5716	0.5559	0.5881	0.5596	0.6092	0.6513	**0.6902***	5.97%
	NDCG@10	0.3903	0.5099	0.5805	0.5613	0.5956	0.5681	0.6124	0.6561	**0.6928***	5.59%

Table 1: The test results of GeoMamba and all baselines. The highest performance is emphasized with bold font and the second highest is marked with underlines. Results marked with * indicates that the corresponding results outperform the most competitive baseline models at a significance level with a p-value$<$0.05 level of t-test.

on long sequences have revealed the fact that Markovian SSMs struggle in recalling from distant historical information (Arora et al. 2023), we propose to address this limitation and incorporate multi-grained information by including multiple observation endpoints within the signal sequences. The representations extracted from these endpoints can accurately reflect the user's historical geographical and preference tendency towards other POI features by the end of observation times.

In practice, to comprehensively represent the historical sequences, we evenly select N observation points across the sequence: $t_{obs} = \{t_i | t_i = \frac{i}{N} n\}$, given the visiting sequence $\mathcal{S}_u$ that involves n visits. The corresponding multi-scale representations are subsequently obtained by:

$$[y^{(L)}_{s,t_1}; y^{(L)}_{s,t_2}; ...; y^{(L)}_{s,t_N}] \in \mathbb{R}^{N \times d}. \tag{13}$$

To obtain the comprehensive representation of user's historical preference, the general user embedding is formed by concatenating the two sets of representations:

$$R^u_{comb} = [\mathrm{Concat}(y^{(L)}_{g,t_1}, y^{(L)}_{f,t_1}); ...; \mathrm{Concat}(y^{(L)}_{g,t_N}, y^{(L)}_{f,t_N})]. \tag{14}$$

Fine-grained Matryoshka Learning Trick After obtaining the multi-scaled representation of input sequences R^u_{comb} , it is crucial to find a way to fully utilize these representations and exploit the implicit patterns from multiple facets. Common approaches involve either directly aggregating the representations (Huang et al. 2024), or using gate-controlled mechanisms (Wang et al. 2022c). However, it is noteworthy that these representations share overlapping information and thus vary in their significance for predicting the next POI. Specifically, representations with later observation time would inherently contain information from previous sequences. This overlap and imbalance can lead to suboptimal results when processed with straightforward aggregation methods introduced above.

To address the aforementioned redundant information issue, we draw inspirations from previous advances in sentence representation learning (Kusupati et al. 2022) and adopt a Matryoshka Representation Learning (MRL) trick to capture the multi-granular characteristics with different representation scales. Under MRL, representations are transformed into latent spaces of varying scales according to their carried information. Specifically, the overall objective of GeoMamba at T-th target POI is formulated as:

$$\mathcal{L}_u = \sum_{k=1}^{N} \mathcal{L}_{CE}(W_k \cdot F_k(R^u_{comb,k}; \theta_{F_k}), p^u_{T+1}), \tag{15}$$

where $\mathcal{L}_{CE}(\cdot,\cdot)$ stands for the cross-entropy loss, $W_k \in \mathbb{R}^{|\mathcal{P}| \times 2^{-k}d}$ are series of trainable linear transformations for multi-scaled representations. The transformation $F_k(\cdot; \theta_{F_k}) : 2d \to 2^{-k}d$, parameterized by θ_{F_k}, maps sequence representations into multi-granularity spaces. The Matryoshka representations of sequences enable the model to capture initial sub-sequences of visits with smaller embedding dimensions, while modeling long-term historical data with larger representations. These finely-grained, multi-granularity representations are then independently utilized for next POI prediction and model optimization.

The recommendation results of GeoMamba are then obtained by jointly considering the multi-scale factor of input visiting signals to compute the recommendation score:

$$\mathrm{Score}^u_{p_i} = [\sum_{k=1}^{N} (W_k \cdot F_k(R^u_{comb,k}, \theta_{F_k}))]_i. \tag{16}$$

Where the scores towards candidates POIs are then used for downstream top-K recommendation process.

Variant	Singapore		Tokyo		NYC	
	R@2	N@2	R@2	N@2	R@2	N@2
V-Attn.	0.2719	0.2580	0.5684	0.5583	0.5667	0.5571
V-Mamba	0.2833	0.2688	0.5453	0.5842	0.6067	0.5954
w/o E-SSM	0.3095	0.2964	0.5636	0.5882	0.6268	0.6566
w/o G-SSM	0.3541	0.3468	0.6115	0.5998	0.6752	0.6683
w/o MRL	0.3543	0.3424	0.6229	0.6128	0.6724	0.6657
Original	0.3740	0.3671	0.6329	0.6224	0.6893	0.6846

Table 2: The ablation results of GeoMamba. R@2 and N@2 are short for Recall@2 and NDCG@2 repsectively.

Experiment

To evaluate the effectiveness of GeoMamba, as well as the functionality of model components, we conduct series of experiments and studies in this section to validate the capability of the proposed GeoMamba.

Experimental Settings

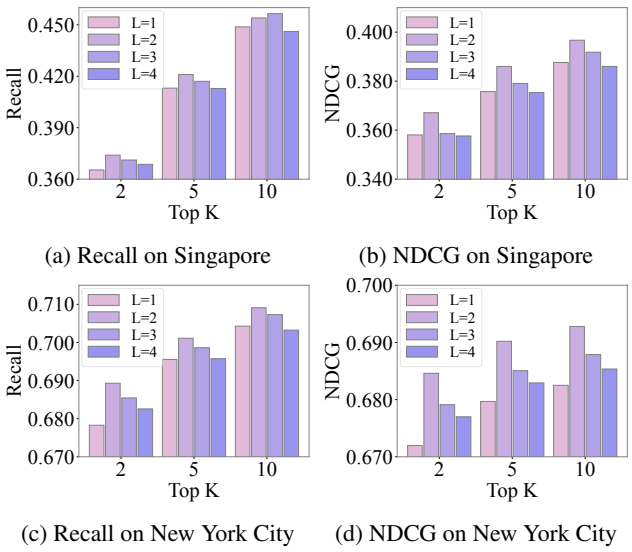
Datasets and Evaluation Protocols To make a comprehensive evaluation of the model performance, the experiments of the proposed GeoMamba and baseline models are conducted on three widely used datasets for POI check-in recommendation. The visit data are collected on a real-world check-in platform Foursquare (Yang et al. 2014) from Singapore, Tokyo, and New York City respectively.

We adopt the same data split strategy from previous works (Wang et al. 2022a; Qin et al. 2023b) and split the sequences in chronological order by $80\%, 10\%, 10\%$ ratio into train, valid, and test sets. To evaluate model performance, we use the commonly adopted Recall@K and NDCG@K as evaluation metrics, with the cut-off K is chosen from $\{2, 5, 10\}$.

Baseline Models We choose a wide range of representative baseline methods from three perspectives:

- **POI correlation-based models**: Include the classical user-item graph-based model LightGCN (He et al. 2020) and a hypergraph framework HCCF (Xia et al. 2022).
- **Location Graph-based Models**: Include location representation-based model GeoIE (Wang et al. 2018), a sequence graph augmentation model SGRec (Li et al. 2021), and a disentangled dual graph representation-based model DRAN (Wang et al. 2022a).
- **Spatio-temporal Sequence Models**: Include a spatio-temporal LSTM model LSTPM (Sun et al. 2020), an attention-based STAN (Luo, Liu, and Liu 2021), and a diffusion generative model Diff-POI (Qin et al. 2023b).

Implementation Details We implement GeoMamba and the baseline methods in PyTorch based on the open-sourced implementations or acquire from the authors. The embedding sizes are fixed to 64 and models are optimized by Adam optimizer with L2 normalization weight of 0.001. For GeoMamba, we finetune the scale number N from $\{2, 3, 4\}$ and the learning rate is fixed as 0.01. The filter coefficient K and J for GaPPO are selected from $\{1, 2, 3\}$, and we fix $\gamma_\varphi = \gamma_\psi = 1$. The number of SSM layers is set as $L = 2$. We conduct experiments with randomly chosen seeds 5 times to represent model performance.



(a) Recall on Singapore (b) NDCG on Singapore

(c) Recall on New York City (d) NDCG on New York City

Figure 3: Model performance w.r.t. model depth L.

General Comparison

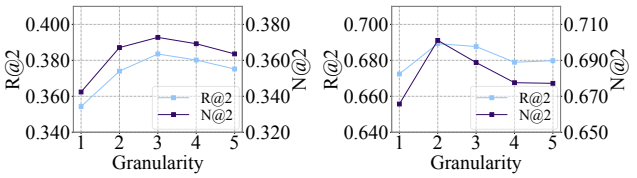
We conduct the general experiments to evaluate the recommendation performance of the aforementioned models on the three datasets, and the corresponding results are listed in Table 1. From the results we can observe that:

- The proposed GeoMamba outperforms all the state-of-the-art methods and achieves significant improvements on all datasets. Specifically, GeoMamba achieves more than 16.0%, 8.0%, and 6.0% performance gain in Recall@2, as well as more than 17.3%, 7.6%, and 6.8% performance gain in NDCG@2 on three datasets respectively, compared with the baseline models.
- Compared graph-based models that depicts geographical relationship with POI graphs, sequential models exhibit slight advantages in next POI recommendation. By extending the state space model architectures to incorporate graph-described locational relationship, the GaPPO-based GeoMamba is able to fully exploit the potential

Ablation Studies

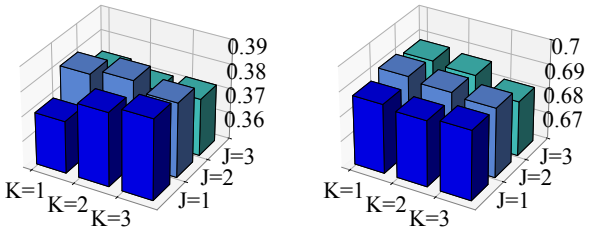
To validate the effectiveness of the proposed components of GeoMamba, we conduct several ablation studies. Specifically, we evaluate the performance of several variants of GeoMamba: model with the SSM encoders replaced by vanilla transformer and Mamba layers (V-Attn. and V-Mamba), model without the geographical SSM encoder (G-SSM) and the enhanced Mamba encoder (E-SSM), and model without multi-scaled Matryoshka representation learning. From the results in Table 2 we can observe that:

- The vanilla transformer and Mamba layer have similar performance on POI recommendation tasks, both inferior to the corresponding specialized recommendation model.



(a) Singapore (b) New York City

Figure 4: Model performance w.r.t. granularity N.



(a) Singapore (b) New York City

Figure 5: Model Recall@2 w.r.t. filter parameters.

- The geographical SSM, the Matryoshka representation learning technique, and the enhanced Mamba encoder all contribute to GeoMamba's recommendation performance. Specifically, the feature-enhanced Mamba encoder plays a more significant role in model performance.

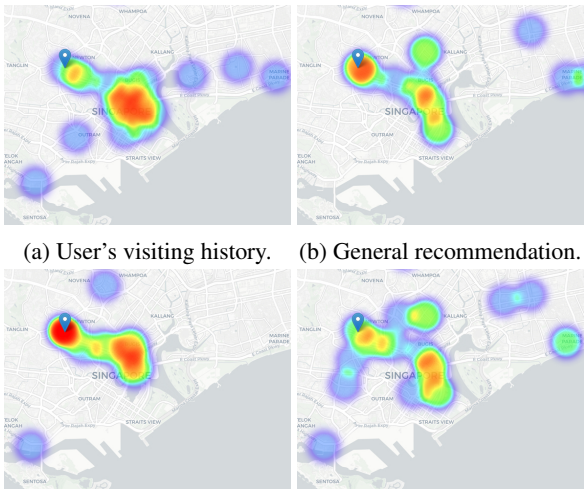
Parameter Analyses

To investigate model's sensitivity and robustness towards different settings, we conduct a series of parameter analyses on the model layers L, the granularity number N, as well as the GaPPO filter parameter K and J.

From Figure 3 which illustrates the model performance w.r.t. number of layers L, we can observe that as the SSM layers go deeper, the model could benefit from more expressive sequence encoders which yield informative visit embeddings, the performance gain decreases as the model depth L growing and eventually results in performance decline.

Additionally, we tune the representation granularity N and the experimental results are reported in Figure 4. From the figure we can observe that the appropriate granularity varies across different datasets. While the granularity with from 2 to 4 scales would be suitable for most settings, the optimal N depends on the specific datasets.

To validate model's sensitivity towards the settings of GaPPO operator, we simultaneously adjust the filter size of K and J, which control the graph filter from both temporal and spectral dimensions. From the results in Figure 5, we can conclude that GaPPOs with larger window sizes may not necessarily have better performance.



(a) User's visiting history. (b) General recommendation.

(c) Geographical SSM. (d) Enhanced Mamba.

Figure 6: Visualization of model recommendation results and user's visiting trajectory.

In-depth Studies

To gain a better understanding on the nature of the proposed graph SSM encoder, as well as investigate the interpretability of the model's recommendation results, we carried out series of in-depth studies on Singapore dataset that reflect the nature of GeoMamba's recommendation process.

Specifically, We adopt the heat map visualization of the check-in trajectory of a randomly-selected user from Singapore dataset, and the personalized top-50 recommended POIs from GeoMamba model and the two modules are chosen for illustrative comparison. The blue map markers represent the ground-truth POI of the user's next visit. From the visualization results in Figure 6 we can observe that The user's previous check-ins were shifted between two hot spots, *i.e.* from the cultural attractions around Marina Centre to the shopping centers on the Orchard Road. Both the general model outputs and the two SSM encoders are capable of capturing such intention transition patterns. Compared with the enhanced Mamba encoder, the geographical SSM encoder makes more concentrated recommendations around the user's historical check-ins and its next visit.

Conclusion

In this paper, we propose GeoMamba, a state space model-based POI recommendation model that is specified for representing the spatio-temporal state transitions in visit sequences. To extend the expressiveness of the sequence state space, we propose a novel GaPPO-based graph SSM that captures the geographical correlation between POIs, as well as a enhanced Mamba encoder to capture the non-locational POI characteristics. The finely-grained representations are further extracted via a Matryoshka representation learning module. Experimental results illustrate the effectiveness of geographical SSMs for next POI recommendations.

Acknowledgements

This paper is partially supported by the National Key Research and Development Program of China with Grant No. 2023YFC3341203 as well as the National Natural Science Foundation of China with Grant Number 62276002.

The authors are grateful to the anonymous reviewers for critically reading the manuscript and for giving important suggestions to improve their paper.

References

Arora, S.; Eyuboglu, S.; Timalsina, A.; Johnson, I.; Poli, M.; Zou, J.; Rudra, A.; and Ré, C. 2023. Zoology: Measuring and improving recall in efficient language models. *arXiv preprint arXiv:2312.04927*.

Cheng, C.; Yang, H.; Lyu, M. R.; and King, I. 2013. Where you like to go next: successive point-of-interest recommendation. In *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence*, IJCAI '13, 2605–2611. AAAI Press. ISBN 9781577356332.

Cho, E.; Myers, S. A.; and Leskovec, J. 2011. Friendship and mobility: user movement in location-based social networks. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, 1082–1090.

Dao, T.; and Gu, A. 2024. Transformers are SSMs: Generalized Models and Efficient Algorithms Through Structured State Space Duality. In *Forty-first International Conference on Machine Learning*.

Feng, S.; Li, X.; Zeng, Y.; Cong, G.; Chee, Y. M.; and Yuan, Q. 2015. Personalized ranking metric embedding for next new POI recommendation. In *Proceedings of the 24th International Conference on Artificial Intelligence*, IJCAI'15, 2069–2075. AAAI Press. ISBN 9781577357384.

Gu, A.; and Dao, T. 2023. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752*.

Gu, A.; Dao, T.; Ermon, S.; Rudra, A.; and Ré, C. 2020. Hippo: Recurrent memory with optimal polynomial projections. *Advances in neural information processing systems*, 33: 1474–1487.

Gu, A.; Goel, K.; and Re, C. 2022. Efficiently Modeling Long Sequences with Structured State Spaces. In *International Conference on Learning Representations*.

Gu, A.; Johnson, I.; Goel, K.; Saab, K.; Dao, T.; Rudra, A.; and Ré, C. 2021. Combining recurrent, convolutional, and continuous-time models with linear state space layers. *Advances in neural information processing systems*, 34: 572–585.

Guo, Q.; Sun, Z.; Zhang, J.; and Theng, Y.-L. 2020. An attentional recurrent neural network for personalized next location recommendation. In *Proceedings of the AAAI Conference on artificial intelligence*, volume 34, 83–90.

Han, H.; Zhang, M.; Hou, M.; Zhang, F.; Wang, Z.; Chen, E.; Wang, H.; Ma, J.; and Liu, Q. 2020. STGCN: a spatial-temporal aware graph learning method for POI recommendation. In *2020 IEEE International Conference on Data Mining (ICDM)*, 1052–1057. IEEE.

Hasani, R.; Lechner, M.; Wang, T.-H.; Chahine, M.; Amini, A.; and Rus, D. 2023. Liquid Structural State-Space Models. In *The Eleventh International Conference on Learning Representations*.

He, X.; Deng, K.; Wang, X.; Li, Y.; Zhang, Y.; and Wang, M. 2020. Lightgcn: Simplifying and powering graph convolution network for recommendation. In *Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval*, 639–648.

Huang, T.; Pan, X.; Cai, X.; Zhang, Y.; and Yuan, X. 2024. Learning Time Slot Preferences via Mobility Tree for Next POI Recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 8535–8543.

Isufi, E.; Loukas, A.; Simonetto, A.; and Leus, G. 2016. Separable autoregressive moving average graph-temporal filters. In *2016 24th European Signal Processing Conference (EUSIPCO)*, 200–204. IEEE.

Kipf, T. N.; and Welling, M. 2017. Semi-Supervised Classification with Graph Convolutional Networks. In *International Conference on Learning Representations*.

Kusupati, A.; Bhatt, G.; Rege, A.; Wallingford, M.; Sinha, A.; Ramanujan, V.; Howard-Snyder, W.; Chen, K.; Kakade, S.; Jain, P.; et al. 2022. Matryoshka representation learning. *Advances in Neural Information Processing Systems*, 35: 30233–30249.

Li, Y.; Chen, T.; Luo, Y.; Yin, H.; and Huang, Z. 2021. Discovering Collaborative Signals for Next POI Recommendation with Iterative Seq2Graph Augmentation. In Zhou, Z.-H., ed., *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*, 1491–1497. International Joint Conferences on Artificial Intelligence Organization. Main Track.

Lian, D.; Zhao, C.; Xie, X.; Sun, G.; Chen, E.; and Rui, Y. 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*, KDD '14, 831–840. New York, NY, USA: Association for Computing Machinery. ISBN 9781450329569.

Lim, N.; Hooi, B.; Ng, S.-K.; Goh, Y. L.; Weng, R.; and Tan, R. 2022. Hierarchical multi-task graph recurrent network for next poi recommendation. In *Proceedings of the 45th international ACM SIGIR conference on Research and development in Information Retrieval*, 1133–1143.

Luo, Y.; Liu, Q.; and Liu, Z. 2021. STAN: Spatio-Temporal Attention Network for Next Location Recommendation. In *Proceedings of the Web Conference 2021*, WWW '21, 2177–2185. New York, NY, USA: Association for Computing Machinery. ISBN 9781450383127.

Manotumruksa, J.; Macdonald, C.; and Ounis, I. 2018. A Contextual Attention Recurrent Architecture for Context-Aware Venue Recommendation. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, SIGIR '18, 555–564. New York, NY, USA: Association for Computing Machinery. ISBN 9781450356572.

Poli, M.; Massaroli, S.; Nguyen, E.; Fu, D. Y.; Dao, T.; Baccus, S.; Bengio, Y.; Ermon, S.; and Ré, C. 2023. Hyena hierarchy: Towards larger convolutional language models. In *International Conference on Machine Learning*, 28043–28078. PMLR.

Qin, Y.; Wang, Y.; Sun, F.; Ju, W.; Hou, X.; Wang, Z.; Cheng, J.; Lei, J.; and Zhang, M. 2023a. DisenPOI: Disentangling sequential and geographical influence for point-of-interest recommendation. In *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining*, 508–516.

Qin, Y.; Wu, H.; Ju, W.; Luo, X.; and Zhang, M. 2023b. A Diffusion Model for POI Recommendation. *ACM Trans. Inf. Syst.*, 42(2).

Sun, K.; Qian, T.; Chen, T.; Liang, Y.; Nguyen, Q. V. H.; and Yin, H. 2020. Where to go next: Modeling long-and short-term user preferences for point-of-interest recommendation. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, 214–221.

Wang, H.; Shen, H.; Ouyang, W.; and Cheng, X. 2018. Exploiting POI-specific geographical influence for point-of-interest recommendation. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, IJCAI'18, 3877–3883. AAAI Press. ISBN 9780999241127.

Wang, W.; Yin, H.; Chen, L.; Sun, Y.; Sadiq, S.; and Zhou, X. 2015. Geo-SAGE: A geographical sparse additive generative model for spatial item recommendation. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*, 1255–1264.

Wang, Z.; Zhu, Y.; Liu, H.; and Wang, C. 2022a. Learning graph-based disentangled representations for next POI recommendation. In *Proceedings of the 45th international ACM SIGIR conference on research and development in information retrieval*, 1154–1163.

Wang, Z.; Zhu, Y.; Zhang, Q.; Liu, H.; Wang, C.; and Liu, T. 2022b. Graph-enhanced spatial-temporal network for next POI recommendation. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 16(6): 1–21.

Wang, Z.; Zhu, Y.; Zhang, Q.; Liu, H.; Wang, C.; and Liu, T. 2022c. Graph-Enhanced Spatial-Temporal Network for Next POI Recommendation. *ACM Trans. Knowl. Discov. Data*, 16(6).

Xia, L.; Huang, C.; Xu, Y.; Zhao, J.; Yin, D.; and Huang, J. 2022. Hypergraph contrastive collaborative filtering. In *Proceedings of the 45th International ACM SIGIR conference on research and development in information retrieval*, 70–79.

Xie, M.; Yin, H.; Wang, H.; Xu, F.; Chen, W.; and Wang, S. 2016. Learning Graph-based POI Embedding for Location-based Recommendation. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, CIKM '16, 15–24. New York, NY, USA: Association for Computing Machinery. ISBN 9781450340731.

Yang, D.; Zhang, D.; Zheng, V. W.; and Yu, Z. 2014. Modeling user activity preference by leveraging user spatial temporal characteristics in LBSNs. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 45(1): 129–142.

Ye, M.; Yin, P.; Lee, W.-C.; and Lee, D.-L. 2011. Exploiting geographical influence for collaborative point-of-interest recommendation. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, 325–334.

Zhao, S.; Zhao, T.; Yang, H.; Lyu, M.; and King, I. 2016. STELLAR: Spatial-temporal latent ranking for successive point-of-interest recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 30.