

Open Access

Human mobility prediction with causal and spatial-constrained multi-task network



Zongyuan Huang¹, Shengyuan Xu¹, Menghan Wang², Hansi Wu², Yanyan Xu^{1*} and Yaohui Jin^{1*}

*Correspondence: yanyanxu@sjtu.edu.cn; jinyh@sjtu.edu.cn ¹ MoE Key Lab of Artificial Intelligence, Al Institute, Shanghai Jiao Tong University, Shanghai, 200240, China Full list of author information is available at the end of the article

Abstract

Modeling human mobility helps to understand how people are accessing resources and physically contacting with each other in cities, and thus contributes to various applications such as urban planning, epidemic control, and location-based advertisement. Next location prediction is one decisive task in individual human mobility modeling and is usually viewed as sequence modeling, solved with Markov or RNN-based methods. However, the existing models paid little attention to the logic of individual travel decisions and the reproducibility of the collective behavior of population. To this end, we propose a Causal and Spatial-constrained Long and Short-term Learner (CSLSL) for next location prediction. CSLSL utilizes a causal structure based on multi-task learning to explicitly model the "when \rightarrow what \rightarrow where", a.k.a. "time \rightarrow activity \rightarrow location" decision logic. We next propose a spatial-constrained loss function as an auxiliary task, to ensure the consistency between the predicted and actual spatial distribution of travelers' destinations. Moreover, CSLSL adopts modules named Long and Short-term Capturer (LSC) to learn the transition regularities across different time spans. Extensive experiments on three real-world datasets show promising performance improvements of CSLSL over baselines and confirm the effectiveness of introducing the causality and consistency constraints. The implementation is available at https://github.com/urbanmobility/CSLSL.

Keywords: Next location prediction; Human mobility; Causality; Spatial consistency; Multi-task network

1 Introduction

Human mobility modeling aims to explore the regularities and patterns of human behavior [1, 2] and plays a significant role in numerous applications, such as urban planning [3], travel demand management [4, 5], health risk assessment [6], epidemic spreading modeling and control [7–9], and so on. In the big data era, the accessibility to GPS, mobile phone records, and location-based social networks (LBSNs) provides an unprecedented chance to understand and model human mobility [2, 10].

In the research community of human mobility, physicists focus on statistical analysis from a macroscopic perspective and have summarized empirical rules [2]. For example, they found that, truncating the power law distribution can well fit the displacement distribution [1]; despite the significant differences in the travel patterns, a majority of users'

© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.



mobility behaviors are predictable [11]. Computer scientists, on the other hand, prefer to model the transition regularities from location sequences, using Markov models [12], recurrent neural networks (RNNs) [13], etc. In summary, statistical physics study the collective behavior at population level, while deep learning methods emphasize modeling individual travel trajectories. Thus, we can expect that integrating physical domain knowledge into a deep learning model encourages the model to pay attention to group behaviors and promotes the performance of deep learning models at population level.

Here we place our emphasis on next location prediction, a vital task in human mobility modeling at individual level [14]. A body of work leverages machine learning methods to tackle this problem due to the sequential nature of mobility behavior. A common thread of these studies is to efficiently capture behavior patterns from sparse data [10, 15–17]. Traditional methods mainly adopt Markov chains to model transition probability matrices across locations, along with techniques like factorization [12, 18, 19] or metric embedding [20]. In recent years, deep learning methods are gaining increasing attention in next location prediction as the recurrent neural network (RNN) presents its capability to capture sequential dependency. To model multi-scale spatio-temporal periodicity, researchers designed attention or gate mechanisms and introduced time and distance interval information [13, 21–24]. Also a few studies incorporate semantic information such as location categories to cope with the data sparsity [16, 25, 26]. However, methods that capture dependencies only from location sequences are difficult to fully fit complex human travel behaviors, especially with sparse data.

To tackle this challenge, we seek to integrate physical knowledge into deep learning methods to enhance the capability of human mobility prediction. Specifically, we propose two physical constraints. The first one is summarized as "*when* \rightarrow *what* \rightarrow *where*" causal relationship. "*When*", "*what*", and "*where*" are the three core elements of human travel behavior and the dependencies between them can explain the motivation of location transfer. For example, as shown in Fig. 1(a), people have specific demands at different times, causing the shifts between locations. Considering causal dependencies enables more comprehensive modeling of human mobility. The second constraint is the macro-statistical characteristics reflecting group behavior. Figure 1(b) illustrates the deviation of the modeled displacement distribution via LSTM from the true distribution in New York City and Tokyo, suggesting LSTM is more likely to focus on shorter trips with higher frequency. Ensuring the consistency between the model output and the macro-statistical characteristics is expected to improve the model's capability to fit travel behavior. We summarize these two constraints as causality and consistency constraints and incorporate them into deep learning models.

To this end, we propose a Causal and Spatial-constrained Long and Short-term Learner (CSLSL), a model integrating the decision logic and the consistency constraints of human mobility modeling. To model the "*when*—*what*—*where*" decision logic, we introduce a causal structure in CSLSL. Based on a multi-task learning, the causal structure utilizes three similar network branches to model the regularities of time, activity, and location, respectively. In line with the "*when*—*what*—*where*" logic, we explicitly build connections between the three branches in the causal structure. As for consistency, we exploratively propose a spatial-constrained loss to reduce the distance between the predicted and actual locations, and indirectly ensure the consistency of the spatial density distribution. In addition, we adopt a Long and Short-term Capturer (LSC) to learn the transition patterns of



different time spans. There are two units in LSC that focus on long-term and short-term regularities respectively.

The main contributions of this work are summarized as follows:

- We propose CSLSL to integrate the travel decision logic and the macro-statistical consistency for human mobility modeling. To our best knowledge, CSLSL is the first model to learn the causality and consistency constraints for next location prediction.
- We introduce a causal structure that can capture not only the separate regularities of time, activity, and location, but also the "*when*→*what*→*where*" causal dependencies. In this way, CSLSL models more essential travel logic in addition to sequence relationships.
- To ensure the consistency in spatial distribution, we propose a spatial-constrained loss to reduce the gap between the predicted and actual destinations.
- We evaluate CSLSL on three real-world datasets to confirm the performance improvements. We also conduct ablation studies and visualization analyses of results such as displacement distribution to demonstrate the effectiveness of our design.

2 Related work

2.1 Next location prediction

Here we classify the approaches to the next location prediction problem into two categories: traditional and deep learning methods. Traditional methods mainly apply Markov chain (MC) and focus on constructing a better location transition probability matrix [12, 18, 20, 27]. For instance, factorized personalized Markov chain (FPMC) combines the matrix factorization technique with Markov chains to learn users' personalized transition matrices [12]. The limitation of the MC-based methods lies in the difficulty in capturing long-term and high-order regularity [16, 17].

Deep learning methods have advantages of learning dense representation and complex dependency. Recently, RNN-based methods show promising performance in mining sequential information. A popular scheme of deep learning methods is incorporating time and distance intervals to assist the model in learning the spatio-temporal regularities of human mobility. Specifically, these methods integrate spatio-temporal information into hidden state transition [28], gate mechanisms [21, 22, 29], or self-attention mechanisms [15, 17], and exploit spatio-temporal contexts in a implicit manner. To leverage the spatio-temporal contexts, researchers explicitly used spatial and temporal factors as attention weights to select the historical hidden states [24, 30]. Another scheme emphasizes the long-term patterns of human behavior, such as DeepMove [13] and LSTPM [23]. They introduce two different components to model long-term and short-term preferences respectively. There is also another scheme that utilizes semantic information such as location categories to improve the performance of location prediction [16, 26, 31]. However, methods that focus on modeling location transfer patterns in sequences cannot effectively capture complex human decision logic. In our work, we propose a causal structure to explicitly capture the "*when*→*what*→*where*" decision logic.

2.2 Time- or activity-jointed location prediction

The methods that jointly predict time or activity learn knowledge from related tasks to improve the prediction performance of location. RMTPP [32] combines RNN and temporal point process (TPP) to jointly model the time and location information. He et al. [25] proposed a two-fold approach that predicts category with Bayesian Personalized Ranking (BPR) technique and then predicts the category-based location. Krishna et al. [33] utilized two distinct LSTM networks to predict activities and durations. DeepJMT [34] fuses spatio-temporal information and social context to predict time and location with a hierarchical RNN and TPP technique. Sun et al. [35] proposed a hybrid LSTM and a sequential LSTM with a self-attention mechanism to jointly model location and travel time. The limitation of these approaches is that they attempt to implicitly and passively learn the correlation between time, category, and location information, but this relationship is explicit and can be directly exploited. In contrast, CSLSL explicitly models the causal dependencies between time, category, and location information through two structural designs.

2.3 Statistical physics-informed human mobility modeling

Explicitly integrating knowledge of statistical physics contributes to guiding model optimization and improving the performance of machine learning methods. On the task of trajectory generation, researchers introduced knowledge of statistical physics to constrain the macroscopic performance of their models, such as the individual trajectory generation model TimeGeo [36] and DITRAS [37], and flow generation model DeepGravity [38]. Unlike the trajectory generation task, only a limited amount of work on individual mobility prediction incorporates knowledge of statistical physics. Zhao et al. [30] integrated domain knowledge, specifically a power-law decay for distances and an exponential decay for time intervals, into an attention mechanism to adjust the impact of historical information on current prediction. However, deep learning-based methods focus more on fitting individual behavior while neglecting group behavior constraints described by macrostatistical characteristics, for example, the regional attractiveness of city blocks. The spatial distribution of model prediction results should be consistent with the actual statistical distribution. Predicted locations closer to the actual location are more expected. Toward this, we propose a spatial-constrained loss function to narrow the distance between the predicted and actual locations, thereby ensuring the consistency of the spatial distribution.

3 Problem formulation

A person's travel behavior can be represented as a sequence of locations, associated with the timestamps and her user ID. In LBSN datasets, each location is also associated with its functional category to support the analysis of user's activity. Let $\mathcal{U} = \{u_1, \ldots, u_{|\mathcal{U}|}\}$, $\mathcal{L} = \{l_1, \ldots, l_{|\mathcal{L}|}\}$ and $\mathcal{C} = \{c_1, \ldots, c_{|\mathcal{C}|}\}$ denote a set of users, locations and functional categories, respectively. Each location l_i is associated with its category and geographical coordinate $(c_i, \operatorname{lat}_i, \operatorname{lon}_i)$.

Definition 1 (Record) Record *r* is a 3-tuple (u_i, l_j, t_k) , representing that the user u_i visited location l_j at time t_k , where $u_i \in U$, $l_j \in \mathcal{L}$.

Definition 2 (Individual Trajectory) A person's trajectory is defined as a record sequence $\mathcal{R} = \{r_1, r_2, \dots, r_{|\mathcal{R}|}\}$, which consists of the person's all records arranged in chronological order. Note that the time interval between two consecutive records is heterogeneous due to the irregular travel behavior.

Definition 3 (Session) Session *S* is a subsequence of records in a time slot. One user's trajectory \mathcal{R} can be split into a series of sessions with various strategies. For example, DeepMove adopts a specific time interval between two consecutive records to split the trajectory [13]. Other strategies segment users' trajectories using a fixed number of records [17, 24] or a meaningful time window such as days, weeks, etc [23]. We define the session where the prediction target is located as the short-term session S_p and the previous historical sessions as long-term sessions { S_q }, $q \in \{1, ..., p-1\}$.

The location prediction problem is formulated as: given a record sequence of a user $\mathcal{R}_{t-1} = \{r_1, \ldots, r_{t-1}\}$, the goal is to predict where the user *u* is most likely to go in her next trip. We use \hat{l}_t to denote the predicted next location. Note that the timestamp of the next trip *t* is also unknown.

4 Methodology

In this section, we first analyze the causality and consistency constraints in human mobility modeling, and then elaborately introduce the design of the proposed model, Casual and Spatial-constrained Long and Short-term Learner (CSLSL). The architecture of CSLSL is illustrated in Fig. 2. It mainly consists of two parts, an embedding part for learning the representations of arrival time, category, and location, from users' recent and historical records; and the second part for learning the regularities of mobility behavior in a multitask learning based causal module and making predictions.

4.1 Causality and consistency constraints

A common practice for next location prediction is to discover similar subsequences or location transition relationships from historical records. This is accomplished by integrating the context information such as distance or time intervals into attention or RNN-centered framework [17, 22–24]. We can formulate such mainstream scheme as $P(\hat{l}_t | \mathcal{R}_{t-1})$, where \mathcal{R}_{t-1} is the historical record sequence. Another scheme adopts multi-task learning techniques to jointly predict next location with time or activity [34, 35], formulated as $P(\hat{l}_t, \hat{c}_t, \hat{t} | \mathcal{R}_{t-1}) = P(\hat{l}_t | \mathcal{R}_{t-1})P(\hat{c}_t | \mathcal{R}_{t-1})P(\hat{t} | \mathcal{R}_{t-1})$, where we assume that the location category can approximate the type of activity. Although these two schemes combine con-



textual information to capture hidden regularities of location transition, they ignore the causal dependencies in the context information.

As aforementioned, we regard "*when*", "*what*" and "*where*" as three crucial elements to describe human mobility [39, 40]. "*When*" refers to the time the trip takes place, e.g. "midday". "*What*" tells about the activities people participate in and also answers the reasons for the trip, such as "having lunch". "*Where*" is the destination of the trip, like "steakhouse". Periodic activities exist in human mobility and occur at specific times and places, such as going to work in the morning and going to a restaurant for lunch, which reveals the correlation between the three elements. When we mention a specific timestamp, we have various activity choices. But we are accustomed to doing certain activities at certain times, such as going to the gym in the evening. Similarly, one activity (category) corresponds to multiple locations (POIs), while one location ID only corresponds to one activity, also reflected in the dataset. Moreover, our target is location prediction, thus location should be the final subtask to leverage the predicted time and activity information. Therefore, we summarize a "*when*→*what*→*where*", a.k.a. "*time*→*activity*→*location*" causal relationship, which is in line with the coarse-to-fine logic of the human decision. The proposed scheme can be formulated as:

$$P(\hat{l}_t, \hat{c}_t, \hat{t} | \mathcal{R}_{t-1}) = P(\hat{l}_t | \hat{c}_t, \hat{t}, \mathcal{R}_{t-1}) P(\hat{c}_t | \hat{t}, \mathcal{R}_{t-1}) P(\hat{t} | \mathcal{R}_{t-1}).$$
(1)

The scheme explicitly models the dependencies between time, activity, and location, and alleviates the difficulty of location prediction. For example, people are accustomed to going to restaurants at midday instead of bar, that is, $P(\hat{c}_t = \text{restaurant}|\hat{t} = \text{midday}, \mathcal{R}_{t-1}) > P(\hat{c}_t = \text{bar}|\hat{t} = \text{midday}, \mathcal{R}_{t-1})$. Each individual has her personalized $P(\hat{l}_t|\hat{c}_t = \text{restaurant}, \hat{t} = \text{midday}, \mathcal{R}_{t-1})$, and the casual constrained location distribution is easier to learn than $P(\hat{l}_t|\mathcal{R}_{t-1})$. In CSLSL, we introduce a causal structure to implement the scheme. In experiments, we also demonstrate that "time \rightarrow acitivity \rightarrow location" outperforms "activity \rightarrow time \rightarrow location" and "time, activity, location", which does not involve logical connections.



On the other side, integrating physical knowledge provides more information and prior constraints to guide the optimization of deep learning models [41, 42]. In human mobility modeling, one can expect that properly introducing the physical laws and domain knowledge would narrow down the gap between the output of deep learning-based approaches and the observed macro-statistical characteristics of human behavior. Due to the difficulty in applying statistical constraints in the training of deep learning models, here we consider the geographic spatial consistency in an indirect way. Specifically, we devise a loss function to constrain the distance between the predicted and actual locations. That is, the closer the predicted location is to the ground truth, the smaller loss we have. By this way, we can indirectly ensure the consistency of the displacement distribution and the consistency of the spatial distribution of travelers' destinations.

4.2 Long and short-term capturer

Human travel behavior has long and short-cycle repetitive patterns, such as going to work every day and going to the supermarket once a week. Inspired by DeepMove [13] and LSTPM [23], we devise a Long and Short-term Capturer (LSC) to learn the behavioral patterns in different observation cycles. In the whole framework shown in Fig. 2, we apply three LSCs to model the time, activity, and location sequences, respectively.

Let $\mathbf{e}^l \in \mathbb{R}^{d^l}$, $\mathbf{e}^c \in \mathbb{R}^{d^c}$, $\mathbf{e}^t \in \mathbb{R}^{d^t}$ and $\mathbf{e}^u \in \mathbb{R}^{d^u}$ denote the embedded representation of location, category, time and user, respectively. Given a historical record sequence \mathcal{R} , CSLSL embeds each record as $(\mathbf{e}^l, \mathbf{e}^c, \mathbf{e}^t, \mathbf{e}^u)$ in hidden spaces. Note that we first convert the continuous timestamp as the hour in a day t^h and the day in a week t^d to present the daily and weekly periodicity. By doing so, we have $\mathbf{e}^t = \mathbf{e}^{t^n} \oplus \mathbf{e}^{t^d}$. Then these representations in a record are concatenated together, $\mathbf{e}^r = \mathbf{e}^l \oplus \mathbf{e}^c \oplus \mathbf{e}^t \oplus \mathbf{e}^u$. We next split each user's record sequence into multiple sessions with a certain time window, like days or weeks. The records in short-term session and long-term sessions are represented as $\widetilde{\mathbf{e}}_p^r = \{\widetilde{\mathbf{e}}_1^r, \dots, \widetilde{\mathbf{e}}_{p-1}^r\}$ and $\{\widetilde{\mathbf{e}}_q^r\} = \{\widetilde{\mathbf{e}}_1^r, \dots, \widetilde{\mathbf{e}}_{p-1}^r\}$, respectively.

Our proposed LSC consists of two capturers to learn the transition regularities in the short-term and long-term sessions, respectively, as shown in Fig. 3. We formulate LSC as:

$$\boldsymbol{h}_{t} = \mathrm{LSC}(\widetilde{\boldsymbol{e}}_{p}^{r}, \{\widetilde{\boldsymbol{e}}_{q}^{r}\}, \boldsymbol{h}_{0}), \qquad (2)$$

where \mathbf{h}_0 is the initial hidden state. In the LSC structure, the short-term capturer takes the hidden state $h_{|S_{p-1}|}$ of the long-term capturer as the initial hidden state to combine the historical information. Because GRU is simple but efficient in modeling temporal data, we apply a layer of GRU in both of the capturers:

$$\boldsymbol{h}_{i} = \mathrm{GRU}(\boldsymbol{e}_{i-1}^{r}, \boldsymbol{h}_{i-1}), \tag{3}$$

where $i \in \{1, 2, ..., |S_{p-1}|\}$ for long-term capturer and $i \in \{1, 2, ..., t\}$ for short-term capturer.

4.3 Causal structure

To model the "time \rightarrow activity \rightarrow location" logic relationship discussed in Sect. 4.1, we introduce a causal structure based on multi-task learning techniques. As illustrated in Fig. 2, we utilize three branches with the same architecture to model the change patterns of time, activity, and location, respectively. Specifically, in each branch, we convey the same record representations to the LSC module and then transfer the output hidden states to the predictor. To explicitly model the summarized causal relation in human travel behavior, we next design two paths for information transfer between various tasks. The first path lies between two LSC modules, passing on the task-specific hidden states. The second path lies between two predictors. In this path, the predicted result of the upstream task is processed by the converter and then conveyed to the downstream task. Here we use the fully connected layer as the predictor (P) and converter(C). That is $\mathbf{y} = \text{Linear}(\mathbf{x}) = \mathbf{W}\mathbf{x} + \mathbf{b}$.

Mathematically, the branch of "time" is formulated as:

$$\boldsymbol{h}_{t}^{t} = \text{LSC}(\widetilde{\boldsymbol{e}}_{\nu}^{r}, \{\widetilde{\boldsymbol{e}}_{a}^{r}\}, 0), \tag{4}$$

$$\hat{t} = \text{Linear}^{(P^{t})}(\boldsymbol{h}_{t}^{t}),$$
(5)

where $\boldsymbol{W}^{(P^t)} \in \mathbb{R}^{1 \times |\boldsymbol{h}^t|}$, \boldsymbol{h}_t^t is the hidden state of the next time, and \hat{t} is the predicted time. As the downstream task of "*time*" in causal structure, the branch of "*activity*" can be formulated as:

$$\boldsymbol{h}_{t}^{c} = \mathrm{LSC}(\widetilde{\boldsymbol{e}}_{p}^{r}, \{\widetilde{\boldsymbol{e}}_{q}^{r}\}, \boldsymbol{h}_{t}^{t}), \tag{6}$$

$$\boldsymbol{c}_{t} = \operatorname{Linear}^{(P^{c})}(\boldsymbol{h}_{t}^{c} \oplus \operatorname{Linear}^{(C^{t})}(\hat{t})), \tag{7}$$

$$\hat{c}_t = \operatorname{argmax}(\boldsymbol{c}_t), \tag{8}$$

where $\boldsymbol{W}^{(C^t)} \in \mathbb{R}^{|\boldsymbol{e}^t| \times 1}$, $\boldsymbol{W}^{(P^c)} \in \mathbb{R}^{|\mathcal{C}| \times (|\boldsymbol{h}^c| + |\boldsymbol{e}^t|)}$, \boldsymbol{h}_t^c is the hidden state of the next activity, and \hat{c}_t is the predicted activity. Eventually, we can formulate the branch of "*location*" as:

$$\boldsymbol{h}_{t}^{l} = \text{LSC}(\widetilde{\boldsymbol{e}}_{p}^{r}, \{\widetilde{\boldsymbol{e}}_{q}^{r}\}, \boldsymbol{h}_{t}^{c}), \tag{9}$$

$$\boldsymbol{I}_{t} = \operatorname{Linear}^{(P^{l})} (\boldsymbol{h}_{t}^{l} \oplus \operatorname{Linear}^{(C^{c})}(\boldsymbol{c}_{t})), \tag{10}$$

$$\hat{l}_t = \operatorname{argmax}(\boldsymbol{I}_t), \tag{11}$$

where $\boldsymbol{W}^{(C^c)} \in \mathbb{R}^{|\boldsymbol{e}^c| \times |\mathcal{C}|}$, $\boldsymbol{W}^{(p^l)} \in \mathbb{R}^{|\mathcal{L}| \times (|\boldsymbol{h}^l| + |\boldsymbol{e}^c|)}$, \boldsymbol{h}_t^l is the hidden state of the next location, \boldsymbol{I}_t is the distribution of predicted location, and \hat{l}_t is the predicted location.

4.4 Spatial-constrained loss function

As discussed in Sect. 4.1, for seeking an alignment of spatial distribution of destinations, we propose a spatial-constrained loss function to shorten the distance from the predicted location to the ground truth at individual level. The distance constraint can be regarded as a self-supervised auxiliary task, integrating the geographical information and restricting the candidate set for better next location prediction. The existing methods introduce the



distance constraints in a regression subtask, directly predicting the geographical locations of destinations [43]. However, in the classification scheme, we must query the coordinates of location IDs to calculate their distance. This operation is not derivable. We get inspiration from REINFORCE [44], which introduces the reward in the loss function to train a policy network, and also consider the distance error as a coefficient to weight the crossentropy between ground truth and the predicted location ID with their physical distance. The spatial-constrained loss function is defined as:

$$L_{s} = -\sum_{i=1}^{N} \text{distance}(\hat{l}_{t,i}, l_{t,i}) \cdot \log(\sigma(\boldsymbol{l}_{t,i})),$$
(12)

where *N* is the total number of records and σ is the softmax function.

We next employ MAE loss for time prediction and cross entropy loss for category and location prediction. Thus we have $L_t = MAE(\hat{t}, t) = \sum_{i=1}^{N} |\hat{t}_i - t_i|$ and $L_* = \text{CrossEntropy}(*) = -\sum_{i=1}^{N} \log(\sigma(*_{t,i})), * \in \{c, l\}$. Thus, the total loss function can be written as

$$L_{\text{total}} = L_l + \lambda_t L_t + \lambda_c L_c + \lambda_s L_s, \tag{13}$$

where λ_t , λ_c , λ_s are weights for their loss functions.

4.5 Structure comparison

There are various strategies for task combination in multi-task learning, such as sharebottom structure [45, 46], hierarchical structure [34, 47], and multi-expert structure [46, 48]. Inspired by these structures, we propose five variants, as shown in Fig. 4, to demonstrate the advantages of our causal structure. Note that these variants use the same basic components as CSLSL, such as GRUs and fully connected layers.

Long and Short-term Learner (LSL) is a basic approach with only one branch to predict location. To jointly predict "*time*", "*activity*", and "*location*", Share-Bottom LSL (SBLSL) introduces two additional predictors that share the same bottom LSC module with the original one. Multi-Experts LSL (MELSL) is an advanced version of SBLSL, with a similar structure to Mixture of Sequential Expert (MoSE) [46]. MELSL employs several GRUs as experts to focus on different aspects of sequence dependencies and gate networks to combine relevant aspects for each task.

Unlike the share-bottom structure, Separate LSL (SLSL) employs a separate branch for each task and the only shared information between each task is the same record representations. Considering the dependencies between tasks, Hierarchical LSL (HLSL) concatenates the record embedding and the output hidden state of the upstream task as the input

	NYC			TKY			Dallas				
	$ \mathcal{U} $	$ \mathcal{L} $	$ \mathcal{R} $	$ \mathcal{U} $	$ \mathcal{L} $	$ \mathcal{R} $	$ \mathcal{U} $	$ \mathcal{L} $	$ \mathcal{R} $		
Raw	1083	38,333	227,428	2293	61,858	573,703	5894	5767	167,016		
Processed	1083	4638	139,183	2293	7222	427,746	2412	5642	146,117		
FPMC-D	1083	4638	94,168	2293	7222	334,561	2412	5642	100,743		
FPMC-W	1083	4638	92,260	2293	7222	303,854	2412	5642	122,419		
DeepMove	1061	4627	111,968	2284	7206	333,215	1193	5346	93,911		
Flashback	439	4093	96,771	1451	6998	366,604	340	5434	82,287		
LSTPM	579	4128	62,511	1866	7007	273,880	587	5169	55,784		
GeoSAN	1073	4611	138,229	2289	7209	427,157	2300	5357	142,980		
STAN	1000	4503	82,783	2200	7047	201,831	2400	5615	95,656		
GETNext	1066	4621	131,920	2280	7200	414,993	-	-	-		
Ours	1065	4635	133,944	2280	7204	422,432	1357	5428	120,531		

 Table 1
 Statistical information of the three datasets, NYC, TKY, and Dallas. The number of users, locations, and records are collected from raw and processed data

of the downstream task. Thus the equation (3) changes to:

$$\boldsymbol{h}_{i}^{k} = \mathrm{GRU}(\boldsymbol{e}_{i-1}^{r} \oplus \boldsymbol{h}_{i}^{k-1}, \boldsymbol{h}_{i-1}^{k}), \tag{14}$$

where \mathbf{h}_{i}^{k} is the hidden state of the *k*-th task at *i*-th time step, and the equation (2) changes to:

$$\boldsymbol{h}_{t}^{k} = \text{LSC}\left(\left[\widetilde{\boldsymbol{e}}_{p}^{r}, \widetilde{\boldsymbol{h}}_{p}^{k-1}\right], \left\{\left[\widetilde{\boldsymbol{e}}_{q}^{r}, \widetilde{\boldsymbol{h}}_{q}^{k-1}\right]\right\}, 0\right), \tag{15}$$

where $[\widetilde{\boldsymbol{e}}_{p}^{r}, \widetilde{\boldsymbol{h}}_{p}^{k-1}] = \{\boldsymbol{e}_{1}^{r} \oplus \boldsymbol{h}_{1}^{k-1}, \dots, \boldsymbol{e}_{t-1}^{r} \oplus \boldsymbol{h}_{t-1}^{k-1}\}.$

5 Experiments

5.1 Data description

We leverage three publicly available check-in datasets in the experiments: two datasets from Foursquare [49] in New York (NYC) and Tokyo (TKY) and one dataset from Gowalla [50] in Dallas. Data in NYC and TKY were collected from 3 April 2012 to 16 February 2013, and data in Dallas was collected from 4 February 2009 to 22 October 2010. The number of users, locations, and records in three datasets are summarized in Table 1, where |*| denotes the number of *. The number of location categories |C| in NYC and TKY are 400 and 385, while Dallas does not contain category information.

To prepare data for baselines and proposed models, we first filter out both users and locations with fewer than 10 records, in line with previous work [13, 23]. We then merge the consecutive records with the same user and location on the same day. The statistical information of the raw and processed data is depicted in Table 1. After pre-processing, the number of categories for NYC and TKY are reduced to 308 and 286. For CSLSL and its variations, we split trajectories into sessions according to week due to the data sparsity. In addition, we require that each session contains at least two records and a user contains at least five sessions to guarantee a training/testing split of 8/2, following [13]. All baselines have their own further data preparation strategies and the model-specific dataset information is also shown in Table 1. It's noteworthy that, LSTPM [23] requires at least three records in each session and Flashback [24] limits the minimum records of each user to 100. These practices filter out more sparse data and reduce the challenge of predic-

tion. Moreover, GETNext requires category information as input, thus it cannot work on dataset Dallas.

5.2 Baselines and settings

Baselines. We compare CSLSL with the state-of-the-art baselines:

- FPMC [12] is a Markov-based model that uses factorization to learn individual transition matrices.
- DeepMove [13] adopts an attention mechanism to learn long-term preference and a GRU module to capture short-term preference.
- Flashback [24] uses spatio-temporal distances as attention weights to search the historical hidden states for current prediction.
- LSTPM [23] considers temporal similarity and distance factor to model long-term preferences and geographical relevance to model short-term preferences.
- GeoSAN [15] designs a geography encoder to implicitly capture spatial proximity and introduces a loss function based on importance sampling to better use the informative negative samples.
- STAN [17] introduces a two-layer attention architecture with spatio-temporal relation matrices to explicitly capture the spatio-temporal correlations.
- GETNext [51] utilizes a GCN to integrate collective movement patterns and a transformer encoder to capture transition regularities. Besides, it introduces location categories as inputs and prediction targets.

Settings. To convincingly compare these baselines with our CSLSL, we collected the open-source codes released by the authors and attempted to find the optimal hyperparameters in the experiments. It's worth noting that most of the baselines only predict next location, without category and time of visitation. Thus, we match the predicted location ID to its category for comparison and exclude the performance comparison of time prediction. Besides, we split users' trajectories by day and week for FPMC model, referred to as FPMC-D and FPMC-W, respectively. For CSLSL, the dimensions of representation vectors e^{t} , e^{c} , $e^{t^{h}}$, $e^{t^{d}}$ and e^{u} are set to 200, 100, 10, 20, and 20 for all datasets. The dimension of the hidden state in all GRUs is set to 600. We use the Adam optimizer with the learning rate of 0.0001, and λ_t , λ_c , and λ_s are set to 10.

Metrics. In the next location prediction task, what we care about is whether the actual location is in the top N of our predictions, $N = \{1, 5, 10\}$. Recall@N is the most commonly used metric and is equal to Accuracy@N because we don't have false positive (FP) and true negative (TN). The definition of Recall@N is

$$\operatorname{Recall}@N = \frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} \frac{|\mathcal{L}_{u}^{T} \cap \mathcal{L}_{u}^{P}|}{|\mathcal{L}_{u}^{T}|},\tag{16}$$

where \mathcal{L}_{u}^{T} and \mathcal{L}_{u}^{P} are the target and top *N* prediction location sets, respectively.

5.3 Performance comparison with baselines

The experimental results are averaged over 10 independent runs and shown in Table 2. For each city, the results are presented in three pieces, representing the results of baselines (lines 1–8), variants (lines 9–14), and ablations (lines 15–18), respectively. The best performance in each column is highlighted in bold text and the second best one is underlined. For NYC and TKY, we present the predicted results for categories and locations,

	NYC						ТКҮ						Dallas		
	Category			Location			Category			Location			Location		
	R@1	R@5	R@10	R@1	R@5	R@10	R@1	R@5	R@10	R@1	R@5	R@10	R@1	R@5	R@10
FPMC-D	0.217	0.510	0.631	0.173	0.424	0.526	0.462	0.623	0.688	0.176	0.408	0.503	0.086	0.204	0.261
FPMC-W	0.196	0.394	0.482	0.162	0.323	0.382	0.428	0.513	0.557	0.117	0.245	0.313	0.055	0.125	0.161
DeepMove	0.244	0.499	0.586	0.198	0.407	0.470	0.398	0.567	0.626	0.161	0.331	0.400	0.089	0.185	0.229
Flashback	0.280	0.618	0.737	0.223	0.521	0.639	<u>0.479</u>	0.730	0.801	0.207	<u>0.486</u>	0.583	0.090	0.199	0.256
LSTPM*	0.335	0.659	0.765	0.267	0.560	0.662	0.461	0.727	0.797	0.231	0.457	0.543	0.123	0.247	0.316
GeoSAN	0.193	0.433	0.602	0.166	0.430	0.584	0.317	0.551	0.694	0.158	0.392	0.528	0.078	0.186	0.265
STAN	0.218	0.480	0.591	0.192	0.411	0.494	0.376	0.575	0.668	0.167	0.388	0.478	0.074	0.153	0.196
GETNext	0.303	0.646	0.749	0.246	0.536	0.622	0.452	0.758	0.844	0.216	0.456	0.550	-	-	-
LSL	0.288	0.587	0.684	0.242	0.506	0.589	0.446	0.697	0.762	0.225	0.462	0.548	0.101	0.203	0.254
SBLSL	0.290	0.595	0.682	0.242	0.488	0.571	0.446	0.769	0.851	0.229	0.460	0.545	0.107	0.194	0.234
MELSL	0.275	0.586	0.662	0.227	0.432	0.526	0.439	0.648	0.745	0.211	0.419	0.475	0.078	0.121	0.164
SLSL	0.281	0.611	0.722	0.253	0.534	0.632	0.409	0.731	0.825	0.233	0.458	0.559	0.114	0.220	0.272
HLSL	0.296	0.628	0.734	0.256	0.536	0.625	0.441	0.760	0.847	0.232	0.472	0.562	0.115	0.230	0.284
CLSL-ctl	0.315	0.647	0.745	0.257	0.543	0.636	0.472	0.780	0.861	0.227	0.472	0.560	-	-	-
CLSL	0.322	0.658	0.747	0.261	0.553	0.643	0.459	0.782	0.864	0.238	0.476	0.567	0.120	0.229	0.282
CSLSL-c	0.315	0.638	0.757	0.247	0.546	0.643	0.450	0.775	0.858	0.230	0.463	0.551	-	-	-
CSLSL-t	0.319	0.648	0.749	0.264	0.556	0.652	0.479	0.740	0.829	0.233	0.478	0.568	0.118	0.231	0.284
CSLSL	<u>0.327</u>	0.661	<u>0.759</u>	0.268	0.568	<u>0.656</u>	0.488	0.801	0.875	0.240	0.488	<u>0.580</u>	0.126	<u>0.243</u>	<u>0.297</u>

Table 2 Performance comparison between baselines, CSLSL, its variants and ablations on three real-world datasets

^{*}Flashback and LSTPM filter out much more sparse users in their data preparation, reducing the challenge of prediction.

while for Dallas, we only show the location prediction results due to the lack of category information.

From the experiment results, we can observe that the proposed CSLSL shows promising performances compared with baselines. In terms of Recall@1 in location prediction, CSLSL achieves 27%, 37%, and 43% averaged performance improvements over these deep learning baselines in three datasets. For Recall@1 in category prediction, the improvements are 34% and 23% in NYC and TKY, respectively. Considering Recall@{5,10} in location prediction, CSLSL achieves similar performances with LSTPM and Flashback, which filter more than 46%, 18%, and 57% of sparse users than we do on three datasets shown in Table 1. These similar performances in the more challenging dataset settings can also reflect the superiority of our model. Disregarding these two models, CSLSL still obtains over 20% averaged improvements than the rest of deep learning baselines. Moreover, CSLSL has improved by 8.9% and 11.1% in Recall@1 in NYC and TKY compared to GETNext, which has similar dataset statistics to ours. The poor performances of all models on the Dallas dataset may be due to the data sparseness. Even so, CSLSL can still utilize the time-location relationship and the spatial constraints to achieve performance gains.

Among the baselines, we observe that the overall performance of location prediction on TKY is lower than that on NYC. This is probably because the TKY dataset has a larger number of users and locations than NYC, increasing the difficulty of mobility prediction. However, CSLSL obtains more performance improvement on TKY than NYC for location prediction compared with baselines. For instance, the performance of CSLSL improves by 8.9% on NYC compared with GETNext, while the improvement is 11.1% on TKY. On the other side, the category predictions for all models have higher accuracy on TKY than NYC mainly owes to the proper modeling of the dependencies.

5.4 Performance comparison with variants

To fairly demonstrate the effectiveness of the proposed causal structure, here we develop an ablated version of CSLSL via dropping the spatial-constraint loss, namely CLSL, and compare it with the 5 variants discussed in the Sect. 4.5. Moreover, we also consider the "*what*→*when*→*where*" relationship, thus we change the order of these three branches in CLSL from "*time*→*category*→*location*" to "*category*→*time*→*location*" and this variant is named CLSL-ctl. We present the results of the variants and CLSL in the second and third pieces of Table 2. The category prediction results of LSL are obtained in the same way as the baselines.

Compared with LSL, SBLSL has a similar performance of location prediction and slightly improved performance of category prediction, suggesting that the shared bottom of SBLSL has indeed learned the category transfer regularities. However, these learned regularities make no contribution to the location prediction. Besides, the performance of MELSL is weaker than LSL and SBLSL, which may be because MELSL does not clarify the relationship between tasks and its experts cannot find suitable optimization directions. The better performance of SLSL than SBLSL indicates that the separate modules to learn transition relationships are better than the shared one. HLSL achieves the best performance in the second pieces of Table 2, suggesting that there are dependencies between *time, category*, and *location*, and that capturing the dependencies facilitates location prediction.

The performances of these variants are weaker than CLSL, suggesting that although these variants utilize temporal and categorical information, they cannot effectively and autonomously capture the dependencies between *time*, *category*, and *location*. In contrast, the causal structure explicitly captures the dependencies between tasks in two ways, thereby fully exploiting their dependencies to improve performance. Moreover, the better performance of CLSL than CLSL-ctl is in line with expectations, because *location* has stronger dependencies with *category* than *time* and category information can bring more performance gains for location prediction. Therefore, our proposed causal structure explicitly models "*time* \rightarrow *category* \rightarrow *location*" rather than "*category* \rightarrow *time* \rightarrow *location*".

5.5 Ablation study

We also conduct ablation studies to examine the contributions of different components in CSLSL. The ablated models include:

- LSL: the version that only keeps the location branch.
- CLSL: the version that removes the spatial-constraint loss.
- CSLSL-t: the version that removes the time branch.
- CSLSL-c: the version that removes the category branch.

As shown in Table 2, we can find that CSLSL-t achieves better performance than CSLSLc, indicating the "category \rightarrow location" relationship has stronger dependency constraints than "time \rightarrow location". This result is also consistent with what we discussed in the Sect. 5.4. The best performance of the complete CSLSL demonstrates the significance of the entire "time \rightarrow category \rightarrow location" decision logic. Comparing the performance of CLSL and CSLSL, we can confirm that the spatial-constraint loss function has a positive impact on performance improvement. Moreover, LSL achieves decent performance compared with baselines, probably because it leverages category information and the LSC module is capable of capturing the long-term and short-term preferences.



5.6 Results visualization analysis

We conduct result visualization analysis to further understand the effectiveness of the causal structure and the spatial-constrained loss. For the causal structure, we compare the category and the location prediction results of GETNext and CSLSL, as shown in Fig. 5. The successfully predicted locations are divided into two parts in the figure based on whether the category prediction is accurate. We can observe that for CSLSL, the records with successfully predicted both categories and locations on NYC and TKY account for 21% and 20% of all records. Compared with GETNext, CSLSL successfully predicted 10% and 18% more locations with more accurately predicted categories on NYC and TKY, respectively. This intuitively indicates that the causal structure can enhance location prediction with more accurate category prediction results. Interestingly, the location can be predicted correctly with a unsuccessfully predicted category. This is because the category information is introduced as additional auxiliary information without imposing mandatory constraints on the location prediction.

To further explore the relationship between categories and location prediction, we examine the accuracy of location predictions for different categories, as depicted in Fig. 6. The category classification is derived from the Foursquare platform. The results exhibit varying levels of predictability for different categories. For instance, the *Community and Government* category shows higher accuracy, while *Retail* demonstrates lower accuracy. This disparity may be attributed to the complex relationship between categories and locations. A greater number of location options within the same category in proximity to the user's location would result in higher prediction difficulty. Additionally, the periodicity of visits to different categories also affects the accuracy of predictions.

The quantity and frequency of individuals' visited locations can reflect the predictability of their travel behavior. Therefore, we utilize entropy to describe the patterns of individual location visits. Entropy(u) = $-\sum_{i}^{n} p_{i} \log p_{i}$, where p_{i} denotes the frequency of *i*-th location and *n* is the total location number the user *u* visited. Figure 7 (a) depicts the correlation between category entropy and location prediction accuracy, while Fig. 7 (b) illustrates the relationship between location entropy and location prediction accuracy. The results





indicate a negative correlation between entropy and accuracy. Users with higher entropy tend to visit more diverse locations, making their travel predictions more challenging.

Regarding the spatial-constrained loss, we examine whether the distances between predicted and actual locations are successfully constrained, and compare CSLSL with four baselines. As shown in Fig. 8, the results show that the predicted locations of CSLSL are





closer to the actual locations, which indicates that the proposed loss can successfully constrain the distance errors. Furthermore, we inspect the constraining effect of the proposed loss on spatial consistency. Figure 9(a) shows the comparison of the predicted displacement with the ground truth. It can be seen that the predicted displacement of CSLSL is closer to the true distribution. This is because the constraint between the predicted and actual locations can indirectly ensure the consistency of the predicted displacements and the ground truth. Figure 9(b) shows the prediction error of regional attractiveness. We divided the geographic regions into square grids with side lengths of 500 m, and counted the difference between the predicted and actual visits in each grid. As presented in Fig. 9(b), CSLSL has a smaller prediction error of regional attractiveness, suggesting that the proposed loss successfully constrains the spatial consistency.



5.7 Sensitivity analysis

We perform sensitivity analysis on dataset NYC and TKY to examine how the performance of CSLSL is affected by $\lambda_*, * \in \{t, c, s\}$. We first vary λ_t and λ_c to analyze the effect of *time* and *category* prediction subtasks with a fixed $\lambda_s = 1$. Then we fix λ_t and λ_c and vary λ_s to observe the impact of spatial-constrained auxiliary tasks. Recall@1 is chosen as the evaluation metric and the results of location prediction are averaged of three runs, shown in Fig. 10.

From Fig. 10 (a), we can observe that the performance of location prediction is more sensitive to λ_c than λ_t , reflecting that the accurate category prediction exerts more influence on the location prediction accuracy, which is also consistent with our proposed decision logic. In addition, the best performance on NYC is obtained with $\lambda_t = 5$ and $\lambda_c = 10$ when $\lambda_s = 1$, while that on TKY is obtained with $\lambda_t = 100$ and $\lambda_c = 50$. Figure 10 (b) shows that CSLSL reaches a more stable accuracy on NYC when $\lambda_s = 5.0$, while the average accuracy is higher when $\lambda_s = 10.0$. The results on TKY show that when λ_t and λ_c are set to smaller values, better performance can be achieved when λ_s is varied. In summary, CSLSL is robust to these parameters and does not suffer from large performance fluctuations with parameter changes.

6 Conclusion

In this work, we propose a Causal and Spatial-constrained Long and Short-Term Learner (CSLSL) to incorporate the individual travel decision logic and the group consistency for next location prediction. In CSLSL, we introduce a causal structure based on multi-task

learning to explicitly capture the "when \rightarrow what \rightarrow where" decision logic and enhance location prediction by fully exploiting the temporal and categorized information. We further propose a simple but effective spatial-constrained loss function that acts as a selfsupervised auxiliary task to incorporate geographical information and indirectly ensure spatial consistency. We conducted extensive experiments to confirm the effectiveness of the design. Specifically, we compared our model with seven baseline models on three datasets, demonstrating the superior performance of the proposed model. Besides, we conducted variant experiments and ablation experiments to validate the effectiveness of the proposed causal structure and spatial constraint loss. Furthermore, we performed additional visualization analyses on the prediction outcomes of the model. These included exploring the relationship between categories and location predictions, analyzing the influence of individual behavioral diversity on predictability, and examining distance relationships and disparities in spatial distribution. Finally, we conducted sensitivity analysis experiments on hyperparameters to examine the robustness of our model. Although we evaluated our model on check-in data, the performance improvement was limited due to the sparse nature of the data. We expect to experiment on dense datasets with comprehensive travel behavior. Such datasets would exhibit more regular patterns in human behavior, enabling the model to more effectively utilize time and activity information to enhance location prediction.

Acknowledgements

The authors thank Wenging Chen for inspiring a part of the model design.

Funding

This work was jointly supported by the National Natural Science Foundation of China (62102258), Shanghai Pujiang Program (21PJ1407300), Shanghai Municipal Science and Technology Major Project (2021SHZDZX0102), and the Fundamental Research Funds for the Central Universities.

Data availability

The Foursquare datasets NYC and TKY can be accessed through the following link: https://sites.google.com/site/yangdingqi/home/foursquare-dataset?authuser=0; The Dallas dataset from Gowalla is available at: https://snap.stanford.edu/data/loc-gowalla.html. The code implementation of this work is accessible online at: https://github.com/urbanmobility/CSLSL.

Declarations

Competing interests

The authors declare that they have no competing interests.

Author contributions

ZH and YX conceived the research and designed the analyses. ZH processed data, conducted experiments, analyzed results, and wrote the paper. SX assisted with the baseline experiments. SX, MW, HW, YX, and YJ provided advice for paper writing. YX supervised the research. All authors read and approved the final manuscript.

Author details

¹MoE Key Lab of Artificial Intelligence, AI Institute, Shanghai Jiao Tong University, Shanghai, 200240, China. ²eBay Inc., Shanghai, China.

Received: 1 June 2023 Accepted: 4 March 2024 Published online: 19 March 2024

References

- 1. Gonzalez MC, Hidalgo CA, Barabasi A-L (2008) Understanding individual human mobility patterns. Nature 453(7196):779–782
- Barbosa H, Barthelemy M, Ghoshal G, James CR, Lenormand M, Louail T, Menezes R, Ramasco JJ, Simini F, Tomasini M (2018) Human mobility: models and applications. Phys Rep 734:1–74
- 3. Xu F, Li Y, Jin D, Lu J, Song C (2021) Emergence of urban growth patterns from human mobility behavior. Nat Comput Sci 1(12):791–800
- 4. Çolak S, Lima A, González MC (2016) Understanding congested travel in urban areas. Nat Commun 7(1):1–8

- 5. Xu Y, Çolak S, Kara EC, Moura SJ, González MC (2018) Planning for electric vehicle needs by coupling charging profiles with urban mobility. Nat Energy 3(6):484–493
- 6. Xu Y, Jiang S, Li R, Zhang J, Zhao J, Abbar S, González MC (2019) Unraveling environmental justice in ambient pm2.5 exposure in Beijing: a big data approach. Comput Environ Urban Syst 75:12–21
- Arenas A, Cota W, Gómez-Gardeñes J, Gómez S, Granell C, Matamalas JT, Soriano-Paños D, Steinegger B (2020) Modeling the spatiotemporal epidemic spreading of Covid-19 and the impact of mobility and social distancing interventions. Phys Rev X 10(4):041055
- Jia JS, Lu X, Yuan Y, Xu G, Jia J, Christakis NA (2020) Population flow drives spatio-temporal distribution of Covid-19 in China. Nature 582(7812):389–394
- 9. Luca M, Lepri B, Frias-Martinez E, Lutu A (2022) Modeling international mobility using roaming cell phone traces during Covid-19 pandemic. EPJ Data Sci 11(1):22
- 10. Luca M, Barlacchi G, Lepri B, Pappalardo L (2020) Deep learning for human mobility: a survey on data and models. ArXiv preprint. arXiv:2012.02825
- 11. Song C, Qu Z, Blumm N, Barabási A-L (2010) Limits of predictability in human mobility. Science 327(5968):1018–1021
- 12. Rendle S, Freudenthaler C, Schmidt-Thieme L (2010) Factorizing personalized Markov chains for next-basket recommendation. In: Proceedings of the 19th international conference on world wide web, pp 811–820
- 13. Feng J, Li Y, Zhang C, Sun F, Meng F, Guo A, Jin D (2018) Deepmove: predicting human mobility with attentional recurrent networks. In: Proceedings of the 2018 world wide web conference, pp 1459–1468
- Couto Teixeira D, Almeida JM, Viana AC (2021) On estimating the predictability of human mobility: the role of routine. EPJ Data Sci 10(1):49
- 15. Lian D, Wu Y, Ge Y, Xie X, Chen E (2020) Geography-aware sequential location recommendation. In: Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining, pp 2009–2019
- 16. Guo Q, Sun Z, Zhang J, Theng Y-L (2020) An attentional recurrent neural network for personalized next location recommendation. In: Proceedings of the AAAI conference on artificial intelligence, vol 34, pp 83–90
- 17. Luo Y, Liu Q, Liu Z (2021) Stan: spatio-temporal attention network for next location recommendation. In: Proceedings of the web conference 2021, pp 2177–2185
- Cheng C, Yang H, Lyu MR, King I (2013) Where you like to go next: successive point-of-interest recommendation. In: Twenty-third international joint conference on artificial intelligence
- He J, Li X, Liao L, Song D, Cheung W (2016) Inferring a personalized next point-of-interest recommendation model with latent behavior patterns. In: Proceedings of the AAAI conference on artificial intelligence, vol 30
- 20. Feng S, Li X, Zeng Y, Cong G, Chee YM, Yuan Q (2015) Personalized ranking metric embedding for next new poi recommendation. In: Twenty-fourth international joint conference on artificial intelligence
- Manotumruksa J, Macdonald C, 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, pp 555–564
- 22. Zhao P, Zhu H, Liu Y, Xu J, Li Z, Zhuang F, Sheng VS, Zhou X (2019) Where to go next: a spatio-temporal gated network for next poi recommendation. In: Proceedings of the AAAI conference on artificial intelligence, vol 33, pp 5877–5884
- 23. Sun K, Qian T, Chen T, Liang Y, Nguyen QVH, 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, vol 34, pp 214–221
- 24. Yang D, Fankhauser B, Rosso P, Cudre-Mauroux P (2020) Location prediction over sparse user mobility traces using rnns: flashback in hidden states! In: Proceedings of the twenty-ninth international joint conference on artificial intelligence, pp 2184–2190
- He J, Li X, Liao L (2017) Category-aware next point-of-interest recommendation via listwise Bayesian personalized ranking. In: IJCAI, vol 17, pp 1837–1843
- 26. Yu F, Cui L, Guo W, Lu X, Li Q, Lu H (2020) A category-aware deep model for successive poi recommendation on sparse check-in data. In: Proceedings of the web conference 2020, pp 1264–1274
- 27. Zhao S, Zhao T, Yang H, Lyu MR, King I (2016) Stellar: spatial-temporal latent ranking for successive point-of-interest recommendation. In: Thirtieth AAAI conference on artificial intelligence
- 28. Liu Q, Wu S, Wang L, Tan T (2016) Predicting the next location: a recurrent model with spatial and temporal contexts. In: Thirtieth AAAI conference on artificial intelligence
- 29. Kong D, Wu F (2018) Hst-lstm: a hierarchical spatial-temporal long-short term memory network for location prediction. In: IJCAI, vol 18, pp 2341–2347
- 30. Zhao K, Zhang Y, Yin H, Wang J, Zheng K, Zhou X, Xing C (2020) Discovering subsequence patterns for next poi recommendation. In: IJCAI, pp 3216–3222
- Wang H, Yu Q, Liu Y, Jin D, Li Y (2021) Spatio-temporal urban knowledge graph enabled mobility prediction. Proc ACM Interact Mob Wearable Ubiquitous Technol 5(4):1–24
- 32. Du N, Dai H, Trivedi R, Upadhyay U, Gomez-Rodriguez M, Song L (2016) Recurrent marked temporal point processes: embedding event history to vector. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, pp 1555–1564
- 33. Krishna K, Jain D, Mehta SV, Choudhary S (2018) An Istm based system for prediction of human activities with durations. In: Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies 1(4), pp 1–31
- 34. Chen Y, Long C, Cong G, Li C (2020) Context-aware deep model for joint mobility and time prediction. In: Proceedings of the 13th international conference on web search and data mining, pp 106–114
- 35. Sun J, Kim J (2021) Joint prediction of next location and travel time from urban vehicle trajectories using long short-term memory neural networks. Transp Res, Part C, Emerg Technol 128:103114
- Jiang S, Yang Y, Gupta S, Veneziano D, Athavale S, González MC (2016) The timegeo modeling framework for urban mobility without travel surveys. Proc Natl Acad Sci 113(37):5370–5378
- Pappalardo L, Simini F (2018) Data-driven generation of spatio-temporal routines in human mobility. Data Min Knowl Discov 32(3):787–829
- Simini F, Barlacchi G, Luca M, Pappalardo L (2021) A deep gravity model for mobility flows generation. Nat Commun 12(1):1–13

- Zhang W, Shen Q, Teso S, Lepri B, Passerini A, Bison I, Giunchiglia F (2021) Putting human behavior predictability in context. EPJ Data Sci 10(1):42
- Pacheco D, Oliveira M, Chen Z, Barbosa H, Foucault-Welles B, Ghoshal G, Menezes R (2022) Predictability states in human mobility. ArXiv preprint. arXiv:2201.01376
- Karniadakis GE, Kevrekidis IG, Lu L, Perdikaris P, Wang S, Yang L (2021) Physics-informed machine learning. Nat Rev Phys 3(6):422–440
- Willard J, Jia X, Xu S, Steinbach M, Kumar V (2020) Integrating physics-based modeling with machine learning: a survey, vol 1 pp 1–34. ArXiv preprint. arXiv:2003.04919
- Xue H, Salim F, Ren Y, Oliver N (2021) Mobtcast: leveraging auxiliary trajectory forecasting for human mobility prediction. Adv Neural Inf Process Syst 34:30380–30391
- 44. Williams RJ (1992) Simple statistical gradient-following algorithms for connectionist reinforcement learning. Reinf. Learn: 5–32
- 45. Ruder S (2017) An overview of multi-task learning in deep neural networks. ArXiv preprint. arXiv:1706.05098
- 46. Qin Z, Cheng Y, Zhao Z, Chen Z, Metzler D, Qin J (2020) Multitask mixture of sequential experts for user activity streams. In: Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining, pp 3083–3091
- 47. Sanh V, Wolf T, Ruder S (2019) A hierarchical multi-task approach for learning embeddings from semantic tasks. In: Proceedings of the AAAI conference on artificial intelligence, vol 33, pp 6949–6956
- Ma J, Zhao Z, Yi X, Chen J, Hong L, Chi EH (2018) Modeling task relationships in multi-task learning with multi-gate mixture-of-experts. In: Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining, pp 1930–1939
- Yang D, Zhang D, Zheng VW, Yu Z (2014) Modeling user activity preference by leveraging user spatial temporal characteristics in lbsns. IEEE Trans Syst Man Cybern Syst 45(1):129–142
- Cho E, Myers SA, 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, pp 1082–1090
- 51. Yang S, Liu J, Zhao K (2022) Getnext: trajectory flow map enhanced transformer for next poi recommendation. In: SIGIR

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen^o journal and benefit from:

- Convenient online submission
- ► Rigorous peer review
- ► Open access: articles freely available online
- ► High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at > springeropen.com