ST-PIL: Spatial-Temporal Periodic Interest Learning for Next Point-of-Interest Recommendation

Qiang Cui, Chenrui Zhang, Yafeng Zhang, Jinpeng Wang*, Mingchen Cai

Meituan, Beijing, China

{cuiqiang04,zhangchenrui02,zhangyafeng,wangjinpeng04,caimingchen}@meituan.com

ABSTRACT

Point-of-Interest (POI) recommendation is an important task in location-based social networks. It facilitates the relation modeling between users and locations. Recently, researchers recommend POIs by long- and short-term interests and achieve success. However, they fail to well capture the periodic interest. People tend to visit similar places at similar times or in similar areas. Existing models try to acquire such kind of periodicity by user's mobility status or time slot, which limits the performance of periodic interest. To this end, we propose to learn spatial-temporal periodic interest. Specifically, in the long-term module, we learn the temporal periodic interest of daily granularity, then utilize intra-level attention to form long-term interest. In the short-term module, we construct various short-term sequences to acquire the spatial-temporal periodic interest of hourly, areal, and hourly-areal granularities, respectively. Finally, we apply inter-level attention to automatically integrate multiple interests. Experiments on two real-world datasets demonstrate the state-of-the-art performance of our method.

CCS CONCEPTS

• **Information systems → Recommender systems**.

KEYWORDS

spatial-temporal, periodic, point-of-interest

1 INTRODUCTION

Location-based Social Networks (LBSNs), such as Foursquare, and Yelp, enable users to share check-in experiences and opinions on Point-of-Interests (POIs). As one of the core services of LBSNs, POI recommendation [6] focuses on utilizing user's check-ins, location-based information, contextual information, and social relationships

---

*Jinpeng Wang is the corresponding author.

to recommend POIs. This task is beneficial not only to the improvement of user experiences and social networking services but also to the advertising between potential locations and users.

A large number of methods are proposed to solve the next POI recommendation and many researchers recently combine long- and short-term interests to boost performance [4, 7–10, 13]. The long-term interest is usually acquired with all his/her historical check-ins, while the short-term interest is modeled with recent check-ins [13]. These interests are usually encoded by feeding the user's check-ins to an encoder network, e.g., a long short-term memory (LSTM) [5] to capture transitional regularities. The way of organizing inputs as successive check-ins works well as users' recent check-ins are more likely to impact his/her next check-in [3, 7, 8, 10].

However, in the case when the user's check-ins are abundant, it is difficult for the model to encode all POIs. Moreover, the location preference of users is generally periodic [14]. As a result, only a small subset of historical check-ins are highly related to the user's next visit. Hence, the successive check-ins would weaken the signals of the related parts. Only a few methods attempt to get out of this limitation by modeling periodicity. For example, DeepMove [4] applies the current mobility status to capture periodicity from all check-ins, LSTPM [8] uses time slot to compute temporal similarity to obtain time-weighted trajectory representation, but they still suffer from the limited capability of learning periodic interest.

The context of check-in, e.g., spatial or temporal information when a user is visiting the POI recommendation system, can provide some hints to the way of modeling periodic interest. Firstly, user activity is constrained by time and exhibits daily and hourly patterns. For instance, some users probably visit different places for vacation only during leisure time, e.g., weekends. Some other users would go to restaurants in the evening. Secondly, user behavior shows an areal pattern when it comes to different areas. Such as some users visit historical sites in the scenic area but play chess in the entertainment area. Thirdly, users will visit the same POI at a specific time, which refers to an hourly-areal pattern. For example, many people go to work at the company at 8 o'clock. Previous methods like DeepMove [4] and LSTPM [8] do not make full use of the spatial-temporal information when learning periodicity.

Motivated by the above analysis, we propose a **S**patial-**T**emporal **P**eriodic **I**nterest **L**earning network (**ST-PIL**) for next POI recommendation. Specifically, ST-PIL adopts the structure of long- and short-term interest learning, but fully utilizes spatial-temporal context to retrieve related parts from historical check-ins for periodic interest. For the long-term module, we split a user's all check-ins into different sequences, according to temporal context called day of week. Then we obtain the daily pattern by average pooling and acquire long-term interest by intra-level attention. This captures

the user's temporal mobility of daily granularity. For the short-term module, we obtain various sequences by contexts called time slot and geohash[1]. Correspondingly, LSTMs are deployed to obtain short-term interests. This module acquires periodicities of hourly, areal, and hourly-areal granularities. For overall representation, inter-level attention and weighted concatenation are used to adaptively combine all interests. A multilayer perceptron is used to obtain the final probability for a candidate POI. The main contributions are summarized as follows:

- We propose to fully consider the spatial-temporal context for periodic interest learning. Specifically, periodicities of daily, hourly, areal, and hourly-areal granularities are constructed.
- We propose two levels of attention. Intra-level attention learns long-term interest from the user's daily pattern. Inter-level attention integrates long- and short-term interests to obtain a better overall representation.

2 RELATED WORK

In this section, we briefly review some recent studies for the POI recommendation and summarize the differences.

Combining long- and short-term interests attracts much attention recently and it achieves state-of-the-art. DRCF [7] captures long-term interest by collaborative filtering and adds short-term interest by a recurrent neural network (RNN). STGN [13] modifies LSTM by devising spatial-temporal gates to enhance interest. Deep-Move [4] applies an RNN-based method to capture transitional regularities and designs a historical attention module to exploit mobility periodicity by user status. LSPL [9] and PLSPL [10] use attention layer to obtain long-term interest and use LSTM to model recent sequential behaviors for short-term interest. LSTPM [8] uses all trajectories with various techniques to obtain long-term interest, such as a time-weighted operation of periodicity, an LSTM and a geo-dilated LSTM are used to capture short-term interest.

The differences between previous work and our study are obvious. Previous work usually captures periodicity from limited granularity, such as time slot. We fully utilize spatial-temporal context and devise effective attention operations to automatically combine various periodic interests.

3 METHODOLOGY

In this section, we introduce the details of our ST-PIL. Fig. 1 shows the architecture, including an embedding layer, long- and short-term modules, inter-level attention and a prediction layer. Various context-aware short-term sequences are illustrated in Fig. 2.

3.1 Embedding Layer

First of all, we introduce rich features. Let $U = \{u_1, u_2, ..., u_{|U|}\}$, $P = \{p_1, p_2, ..., p_{|P|}\}$, and $C = \{c_1, c_2, ..., c_{|C|}\}$ denote the set of users, POIs, and categories, respectively, where each POI has a category $c \in C$. Next is spatial-temporal information. As each POI has a geocoding of latitude and longitude, we transform it into the geohash-integer version. We use geohash-5 $G = \{g_1, g_2, ..., g_{|G|}\}$ to express spatial information as areas, and each POI only belongs to one geohash-5. Spatial features are represented by day of week and
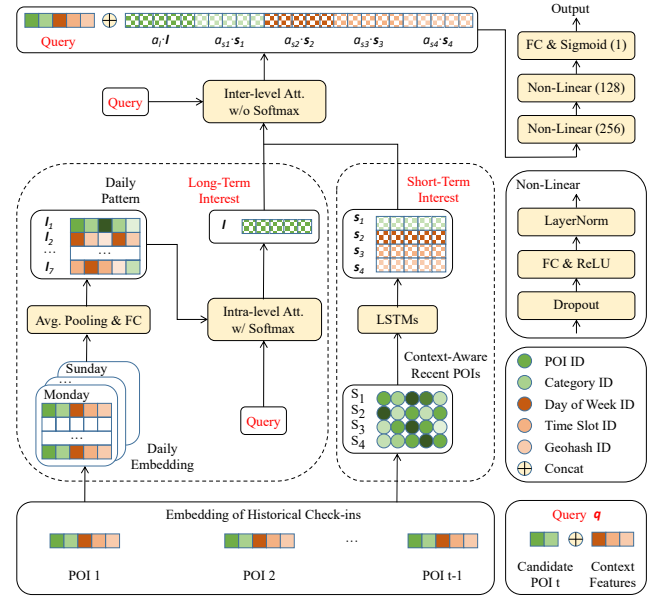
[1]https://en.wikipedia.org/wiki/Geohash

Figure 1: The overall architecture of ST-PIL.

time slot, where we have 7 days in a week $W = \{w_1, ..., w_7\}$ and 24 slots of hours in a day $M = \{m_1, ..., m_{24}\}$. Then, a user's check-in at t-th time step is $h^t = (p^t, c^t, w^t, m^t, g^t)$. In our work, each feature has a unique ID and is represented by a learned embedding.

Next is the problem formulation. Given a user's all historical check-ins $H = \{h^1, h^2, ..., h^{t-1}\}$ as well as his/her current time and area (mapped as (w^t, m^t, g^t)), our goal is to select top-k POIs that the user would check-in. For example, we now know that a user is near a certain mall, we need to predict which POI he/she will go to.

3.2 Long-Term Module

In the long-term module, we use temporal context called day of week to find user's daily pattern and use intra-level attention to establish long-term interest.

Daily pattern aims to acquire user's daily periodicity, from Monday to Sunday. Firstly, we construct daily mask sequences with the same length as the user check-in histories and divide all check-ins into different days to form daily embedding. For example, we use a sequence $d = [1, 0, 1, ..., 0]$ to indicate the user performed check-ins on Monday, where symbols '1' and '0' refer to whether or not a check-in occurs on that day. Through the element-wise product between the mask d and user's all historical check-ins, the check-ins on Monday are taken out separately. Next, through average pooling and a shared fully connected layer on each daily embedding, we end up with seven daily patterns, denoted as $L = [l_1; l_2; ...; l_7]$. Specifically, we calculate the cumulative sums of a daily embedding and a mask and then divide them to achieve average pooling.

Long-term interest can summarize important information from the daily pattern with intra-level attention. We first create a query q for attention. A query at t-th time step is built by the concatenation of a candidate POI and the already known context (w^t, m^t, g^t):

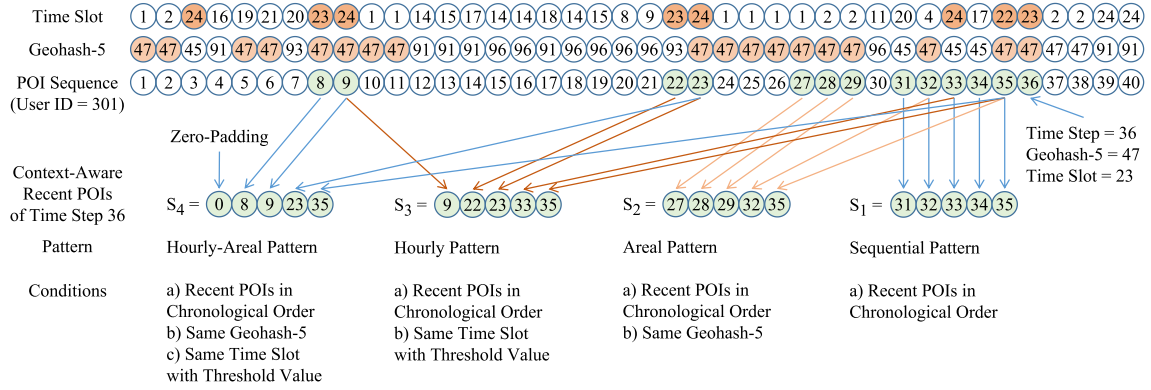$$q^t = [p^t, c^t, w^t, m^t, g^t], \tag{1}$$

Figure 2: Illustration of obtaining various short-term sequences for a certain time step by spatial-temporal context.

where every POI can be a candidate but only one is positive, and each feature ID is embedded by a dense vector. Then, we apply the widely-used Bahdanau attention[2] [1] to aggregate daily pattern L^t:

$$
\begin{aligned}
e_j^t &= \boldsymbol{v}_e^{\mathrm{T}} \tanh(V_1 \boldsymbol{q}^t + V_2 \boldsymbol{l}_j^t), \\
\boldsymbol{e}^t &= \mathrm{softmax}(\{e_1^t, e_2^t, ..., e_7^t\}), \\
\boldsymbol{l}^t &= \sum_{j=1}^{7} e_j^t \boldsymbol{l}_j^t,
\end{aligned}
\tag{2}
$$

where $\boldsymbol{v}_e, V_1$ and V_2 are weight matrices, each weight e_j^t is normalized by $softmax$ and $\boldsymbol{l}^t$ is the t-th long-term interest.

3.3 Short-Term Module

The short-term module plays an important role to capture the fundamental sequential pattern and context-aware patterns, e.g., hourly, areal, and hourly-areal periodicities.

Fig. 2 illustrates the rule conditions of obtaining short-term POI sequences $S^t = \{S_1^t, S_2^t, S_3^t, S_4^t\}$ from the user's all historical check-ins. In detail, at current time step $t = 36$, we have spatial-temporal context called time slot $m^t = 23$ and area $g^t = 47$. Correspondingly, S_1 uses recent POIs to capture the sequential transition regularity, S_2 and S_3 finds POIs that the user has checked-in before in the current area $g^t = 47$ and in similar time slots $\{22, 23, 24\}$ respectively, and S_4 is the insection of S_2 and S_3.

For each short-term sequence at t-th time step S_k^t, we apply a LSTM [5] to obtain its interest:

$$
\boldsymbol{s}_k^t = \mathrm{LSTM}(S_k^t),
\tag{3}
$$

where $\boldsymbol{s}_k^t$ is a separate short-term interest and $k = \{1, 2, 3, 4\}$. As each sequence implies a specific mobility pattern, we use four LSTMs instead of one to avoid collision.

3.4 Inter-level Attention

In this section, we obtain the final interest representation through inter-level attention.

At current time step, we have five interests $\{\boldsymbol{l}^t, \boldsymbol{s}_1^t, \boldsymbol{s}_2^t, \boldsymbol{s}_3^t, \boldsymbol{s}_4^t\}$, where $\boldsymbol{l}^t$ is the long-term interest and others are short-term interests. We still use Bahdanau attention to compute weights, e.g.,

$$
a_l^t = \boldsymbol{v}_a^{\mathrm{T}} \tanh(V_3 \boldsymbol{q}^t + V_4 \boldsymbol{l}^t),
\tag{4}
$$

where $\boldsymbol{v}_a, V_3$ and V_4 are new weight matrices, and a_l^t is the attention weight for $\boldsymbol{l}^t$. Correspondingly, other weights are computed for the rest interests. Next, we concatenate all these interests:

$$
\boldsymbol{x}^t = [a_l^t \boldsymbol{l}^t, a_{s1}^t \boldsymbol{s}_1^t, a_{s2}^t \boldsymbol{s}_2^t, a_{s3}^t \boldsymbol{s}_3^t, a_{s4}^t \boldsymbol{s}_4^t],
\tag{5}
$$

where $\boldsymbol{x}^t$ is the overall interest.

Different from intra-level attention in Section 3.2, inter-level attention has its characteristics. First, as the interests all come from historical check-ins, we add a dropout layer for query $\boldsymbol{q}^t$ and interests to avoid overfitting before computing attention weights. Second, as these interests capture different mobility patterns, we aggregate them by weighted concatenation to preserve unique features. Besides, we abandon the $softmax$ on the attention weights to improve the expression ability of important patterns.

3.5 Prediction Layer

We compute the probability of a candidate POI based on the spatial-temporal context in the prediction layer, consisting of an input, a multilayer perceptron (MLP), and a loss function. The input for MLP is a concatenation of the query $\boldsymbol{q}^t$ and the overall interest $\boldsymbol{x}^t$:

$$
input^t = [\boldsymbol{q}^t, \boldsymbol{x}^t].
\tag{6}
$$

The following process of MLP is illustrated in Fig. 1. Finally, we use the popular cross-entropy as our loss:

$$
\mathcal{L} = -\sum_{i=1}^{N} y_i \cdot \log(p(y_i)) + (1 - y_i) \cdot \log(1 - p(y_i)),
\tag{7}
$$

where N is the number of all candidates, including positives and negatives, $y_i \in \{0, 1\}$ and $p(y_i)$ are label and probability for the candidate respectively.

4 EXPERIMENTS

4.1 Experimental Settings

Datasets. We choose two public real-world datasets collected from New York city (NYC) and Tokyo (TKY)[3] [12]. Each check-in contains a user ID, POI ID, category ID, latitude, longitude, and UTC. We delete POIs that are checked by less than 5 times, and we only keep the most recent 500 check-ins for each user. Following the previous work [7], we adopt the leave-one-out evaluation methodology. We

---

[2]https://www.tensorflow.org/tutorials/text/nmt_with_attention

[3]http://www-public.it-sudparis.eu/~zhang_da/pub/dataset_tsmc2014.zip

Table 1: Performance comparison on two datasets. The best value and the runner-up are boldfaced and underlined, respectively.

		NYC					TKY				
		Acc@1	Acc@5	Acc@10	MRR@5	MRR@10	Acc@1	Acc@5	Acc@10	MRR@5	MRR@10
Baselines	LSTM	0.1816	0.3991	0.4663	0.2610	0.2701	0.1506	0.3553	0.4511	0.2256	0.2384
	STGN	0.2071	0.4051	0.4645	0.2829	0.2909	0.2134	0.4352	0.5161	0.2956	0.3066
	LSTPM	0.1846	0.3693	0.4484	0.2549	0.2660	0.1928	0.3889	0.4622	0.2647	0.2746
	PLSPL	0.2188	0.4367	0.5162	0.3000	0.3109	0.2054	0.4252	0.5225	0.2853	0.2983
Ours	ST-PIL$_L$	0.3135	0.5543	0.6243	0.4055	0.4151	0.2842	0.5340	0.6287	0.3773	0.3900
	ST-PIL$_S$	0.3790	0.5798	0.6363	0.4543	0.4619	0.3464	0.5872	0.6708	0.4386	0.4498
	ST-PIL	**0.3807**	**0.5850**	**0.6454**	**0.4584**	**0.4666**	**0.3523**	**0.5963**	**0.6772**	**0.4455**	**0.4564**
	improvement (%)	+74.00	+33.96	+25.03	+52.80	+50.08	+71.52	+37.02	+29.61	+50.70	+48.86

treat each user's last POI as ground truth, compute probabilities for all POIs during the test and rank them.

Baselines. We compare our ST-PIL with several baselines, which are good at capturing long- and short-term interests. **LSTM** [5] is a widely-used method for sequential modeling. **STGN** [13] controls both interests by devising spatial-temporal gates. **LSTPM** [8] designs a non-local network for long-term interest and a geo-dilated LSTM for short-term interest. **PLSPL** [10] uses user-based attention and LSTMs for two interests, respectively. And it adopts another user-based attention to obtain overall predicted probability.

Parameters and Evaluations. The dimensions of POI, category, day of week, time slot, and geohash-5 are 64, 32, 8, 16, and 32, respectively. We make these dimensions approximately proportional to the logarithm of the number of their unique ID values [2]. The number of negative candidates is 20. The dimension of each interest is 64. We grid search the dropout rate from $\{0.0, 0.1, 0.2\}$. Correspondingly, the best rates of inter-level attention are 0.0 and 0.1 on NYC and TKY, respectively, and the best rate of MLP is 0.1. We apply Acc@k [11] and MRR@k to evaluate methods. Our model has two variants, called ST-PIL$_L$ and ST-PIL$_S$ which only have the long-term module and short-term module, respectively.

4.2 Results and Discussions

Performance. The results are listed in Table 1 and our ST-PIL significantly outperforms the others. Given the comparison with baselines, our ST-PIL achieves a great improvement on the two datasets. Specifically, for Acc@1, ST-PIL outperforms the runner-up by about 70%. For Acc@5 and Acc@10, our model improves by about 30%. The overall improvements of MRR@5 and MRR@10 are about 50%. In light of the model variant, ST-PIL$_S$ outperforms ST-PIL$_L$ and ST-PIL is the best, which clearly shows the advantage of our inter-level attention in integrating multiple interests.

Settings of Long-Term Module. We set up three settings, called att-qk, att-k, and seq-avg, to examine how to deal with all historical check-ins. The att-qk refers to our long-term module. The att-k does not use the query in intra-level attention. The seq-avg averages the entire POI sequence. The comparison is illustrated in Fig. 3, where symbols 'L' and 'L+S' refer to ST-PIL$_L$ and ST-PIL, respectively. Obviously, the setting att-qk is best. In summary, capturing daily periodicity with a query can enhance the expressions of long-term interest and overall interest.

Ablation Study of Short-Term Module. We compare four short-term interests to explore different mobility patterns. Results are illustrated in Fig. 4 by blue and red colors. For the first discussion,
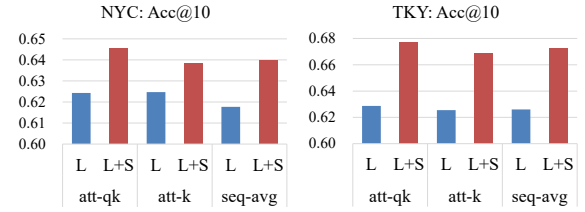
Figure 3: Examine different settings for long-term module.
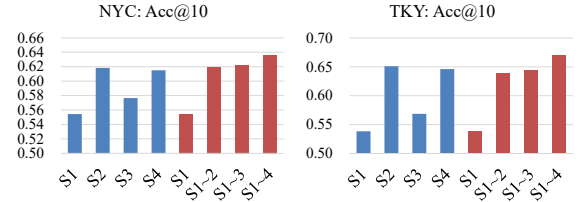
Figure 4: Ablation study of short-term module.

the four interests perform differently. The performance of S_2 and S_4 are better than the other two. This means the spatial and spatial-temporal periodicities are more influential than sequential effect and temporal periodicity. Perhaps when a user chooses a POI in a certain area, the POIs he/she has visited before in this area have a greater impact. For the second discussion, the overall performance of ST-PIL$_S$ is better than any short-term interest. Besides, when we gradually add four short-term interests, the overall performance is getting better and better. This comparison also verifies the effectiveness of our inter-level attention.

5 CONCLUSION AND FUTURE WORK

In this work, we propose a spatial-temporal periodic interest learning network (ST-PIL) with long- and short-term modules for the next POI recommendation. In the long-term module, we explore the daily pattern and capture long-term interest by intra-level attention. In the short-term module, we extract various short-term sequences to model sequential, spatial, temporal, and spatial-temporal patterns. Inter-level attention can effectively integrate different interests. The experiment demonstrates the effectiveness of our ST-PIL. In the future, we would design a more automated modeling methodology to replace the pre-operations of constructing different sequences.

REFERENCES

[1] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. Neural machine translation by jointly learning to align and translate. In *ICLR*.

[2] Paul Covington, Jay Adams, and Emre Sargin. 2016. Deep neural networks for youtube recommendations. In *RecSys*. 191–198.

[3] Qiang Cui, Yuyuan Tang, Shu Wu, and Liang Wang. 2019. Distance2Pre: Personalized spatial preference for next point-of-interest prediction. In *PAKDD*. Springer, 289–301.

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

[5] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural Computation* 9, 8 (1997), 1735–1780.

[6] Md Islam, Mir Mahathir Mohammad, Sarkar Snigdha Sarathi Das, Mohammed Eunus Ali, et al. 2020. A Survey on Deep Learning Based Point-Of-Interest (POI) Recommendations. *arXiv preprint arXiv:2011.10187* (2020).

[7] Jarana Manotumruksa, Craig Macdonald, and Iadh Ounis. 2017. A deep recurrent collaborative filtering framework for venue recommendation. In *CIKM*. 1429–1438.

[8] Ke Sun, Tieyun Qian, Tong Chen, Yile Liang, Quoc Viet Hung Nguyen, and Hongzhi Yin. 2020. Where to go next: Modeling long-and short-term user preferences for point-of-interest recommendation. In *AAAI*, Vol. 34. 214–221.

[9] Yuxia Wu, Ke Li, Guoshuai Zhao, and Xueming Qian. 2019. Long-and short-term preference learning for next POI recommendation. In *CIKM*. 2301–2304.

[10] Yuxia Wu, Ke Li, Guoshuai Zhao, and QIAN Xueming. 2020. Personalized long-and short-term preference learning for next POI recommendation. *TKDE* (2020).

[11] Min Xie, Hongzhi Yin, Hao Wang, Fanjiang Xu, Weitong Chen, and Sen Wang. 2016. Learning graph-based poi embedding for location-based recommendation. In *CIKM*. 15–24.

[12] Dingqi Yang, Daqing Zhang, Vincent W Zheng, and Zhiyong Yu. 2014. Modeling user activity preference by leveraging user spatial temporal characteristics in LBSNs. *IEEE Transactions on SMC: Systems* 45, 1 (2014), 129–142.

[13] Pengpeng Zhao, Haifeng Zhu, Yanchi Liu, Jiajie Xu, Zhixu Li, Fuzhen Zhuang, Victor S Sheng, and Xiaofang Zhou. 2019. Where to Go Next: A Spatio-Temporal Gated Network for Next POI Recommendation. In *AAAI*, Vol. 33. 5877–5884.

[14] Shenglin Zhao, Irwin King, and Michael R Lyu. 2016. A survey of point-of-interest recommendation in location-based social networks. *arXiv preprint arXiv:1607.00647* (2016).