CHA: Categorical Hierarchy-based Attention for Next POI Recommendation

HONGYU ZANG, DONGCHENG HAN, XIN LI, and ZHIFENG WAN, Beijing Institute of Technology, Bejing, China
MINGZHONG WANG, University of the Sunshine Coast, Queensland, Australia

Next **Point-of-interest (POI)** recommendation is a key task in improving location-related customer experiences and business operations, but yet remains challenging due to the substantial diversity of human activities and the sparsity of the check-in records available. To address these challenges, we proposed to explore the category hierarchy knowledge graph of POIs via an attention mechanism to learn the robust representations of POIs even when there is insufficient data. We also proposed a spatial-temporal decay LSTM and a Discrete Fourier Series-based periodic attention to better facilitate the capturing of the personalized behavior pattern. Extensive experiments on two commonly adopted real-world **location-based social networks (LBSNs)** datasets proved that the inclusion of the aforementioned modules helps to boost the performance of next and next new POI recommendation tasks significantly. Specifically, our model in general outperforms other state-of-the-art methods by a large margin.

CCS Concepts: • **Information systems** → **Collaborative filtering**; • **Applied computing** → *Sociology;*

Additional Key Words and Phrases: Next POI recommendation, next new POI recommendation, categorical hierarchy-based attention

ACM Reference format:
Hongyu Zang, Dongcheng Han, Xin Li, Zhifeng Wan, and Mingzhong Wang. 2021. CHA: Categorical Hierarchy-based Attention for Next POI Recommendation. *ACM Trans. Inf. Syst.* 40, 1, Article 7 (August 2021), 22 pages.
https://doi.org/10.1145/3464300

1 INTRODUCTION

The users of popular mobile-based social media, such as Facebook, Foursquare, and Twitter, have generated a massive amount of behavioral data with spatial-temporal details, such as real-time locations, check-in time, activities/events, comments/feelings, experiences, and ratings, which

Hongyu Zang and Dongcheng Han contributed equally to this research.

This research has been partially supported by NSFC under Grant (U19B2020 and 61772074) and National Key R&D Program of China under Grant (2017YFB0803300).

Authors' addresses: H. Zang, D. Han, and X. Li (corresponding author), School of Computer Science, Beijing Institute of Technology, 5 South Zhongguancun Street, Haidian District, Beijing, 100081, China; emails: {zanghyu, dchan, xinli}@bit.edu.cn; Z. Wan, School of Mechatronical Engineering, Beijing Institute of Technology, 5 South Zhongguancun Street, Haidian District, Beijing, 100081, China; email: 3120190232@bit.edu.cn; M. Wang, School of Science, University of the Sunshine Coast, 90 Sippy Downs Drive, Sippy Downs, Queensland, Australia; email: mwang@usc.edu.au.

1046-8188/2021/08-ART7 $15.00
https://doi.org/10.1145/3464300

promotes the emergence of LBSNs. LBSN is the social structure connected by one or more specific types of interrelationships between users.

In daily life, people physically visit places according to their interests, such as shopping malls, restaurants, and tourist attractions. These places are referred to as **Points of Interest (POIs)**, which also contribute to the contents of LBSN for users to check-in online and share their personalized information with others in the network. Appropriate recommendations of POIs would benefit both customers and businesses as they help customers to make better selections while helping the expansion of businesses. Therefore, **Point-of-Interest recommendation (POIR)** has become an important topic in both academia and industry.

The demand and competition in the businesses, especially the advertising market, require more specific recommendations which are capable of predicting POI(s) that users will visit in their very next step(s), and this type of problem is termed as *next POIR*.

Next POIR can benefit from the big volume of data generated by users in LBSNs. For example, Foursquare has more than 60 million registered users and more than 55 million monthly active users. The Foursquare Swarm app has about 9 million daily check-ins on average. The amount of data keeps growing at an astonishing rate. However, it is also challenging to make accurate and personalized POI recommendations for users regarding the big volume of available information.

Many next POIR solutions are based on the **Collaborative Filtering (CF)** by adding other features, such as temporal information [39], geographical information [18], and social relationship [6], to boost the performance of the next POIR tasks. For example, Cui et al. [9] utilized a CF method to consider user information, such as photo uploading records and heterogeneous high-order relationship, to capture more accurate personalized preferences. Li et al. [19] proposed the two-fold model to first predict the users' preference at the category level leveraging on a tensor factorization, and then rank the POI candidates within the predicted category with the temporal and distance constraints.

The (deep) neural network structures have also been adapted in the next POIR tasks because (1) recommendation systems are based on the users' past behavior and neural networks have good performance at processing sequences and (2) deep learning can effectively learn the hidden factors from various features' representations. For example, Al-Molegi et al. [1] employed a **recurrent neural network (RNN)** to fit the data and make the prediction. Liu et al. [21] proposed an ST-RNN, which explicitly incorporates the continuous time intervals and distances between the visits in each layer of the recurrent structure, to model the sequential data. Zhao et al. [45] proposed an ST-LSTM with two pairs of time gates and distance gates to capture the short-term interest and long-term interest, respectively. It aims to mitigate RNN's limitation due to the gradient vanishing and gradient exploding problems. Moreover, HST-LSTM [14] stacks ST-LSTM in a hierarchical manner and employs the encoder and decoder to compute the embedding of the historical visits.

However, the aforementioned approaches need a huge amount of relatively dense data to train their deep models to learn the representations of POIs. Unfortunately, most (next) POIR applications suffer from the severe data sparsity as most mobility data is collected on a voluntary basis where users usually do not frequently check in at the POIs that they visit. Moreover, the visited locations of a specific user are rather sparse in comparison with all possible POIs in LBSNs.

We analyzed the datasets used in our experiments and depicted each location's check-in counts and its frequency in the datasets in Figure 1. The figure demonstrates that most locations are visited infrequently, and the locations with a higher visiting frequency account for a smaller proportion of all locations in both Foursquare New York (NYC) and Foursquare Tokyo (TKY) datasets. Apparently, the sparsity of user-location check-in in POIR is dramatically higher than that of user-item rating in traditional recommendation systems, making POIR more challenging.

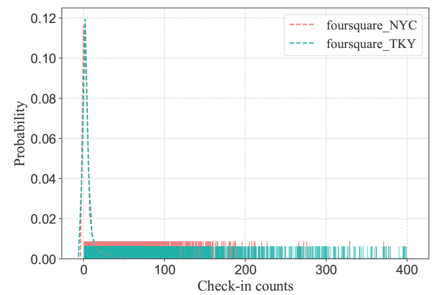
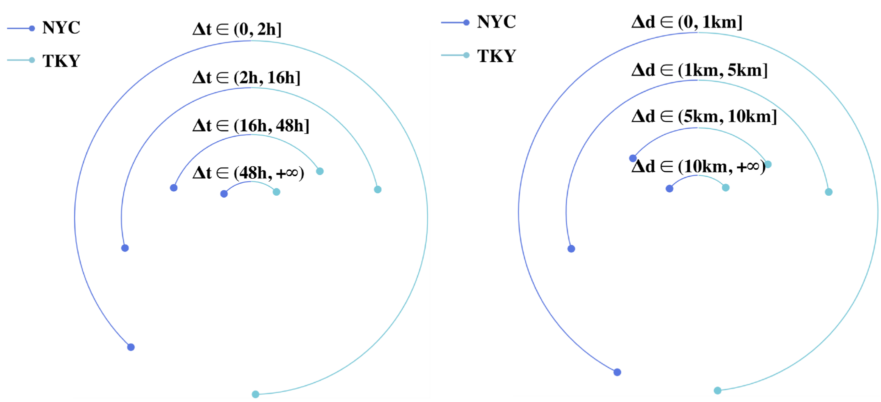


Fig. 1. Check-ins probability vs. counts.



(a) Check-ins probability vs. time interval (b) Check-ins probability vs. distance

Fig. 2. Check-ins probability vs. time interval / distance. The time interval and the distance between two successive check-ins are split into four sections, the blue lines and the green lines describe the NYC and TKY datasets, respectively.

Figure 2 illustrates different time interval and geographical distance between two successive locations and its corresponding check-ins probability. Figure 2(a) depicts check-ins probability regarding the time interval. The data shows that 38.08% of two adjacent check-ins of all users are within 2 hours in the NYC dataset and 50.04% in the TKY dataset. In contrast, the probability is only 13.53% if users visit the next location after 48 hours. In general, the larger the time interval, the lower the probability for the next POI to be visited. Similarly, Figure 2(b) depicts check-ins probability regarding the distance, in which we can conclude that the larger the distance, the lower the probability for the next POI to be visited.

It is evident that the next location visited by a user is closely related to the time interval and the distance to the current location, implying that users prefer nearby locations rather than distant ones. This kind of spatial-temporal features provides a promising direction in improving the performance of next POIR.

Knowledge-graph-based models have shown great potential in reasoning [42, 43] and various recommendation applications, such as news, E-commerce, and movie and music recommendations, to alleviate the sparsity problem as well as to enhance the accuracy and interpretability of the recommendation system by providing additional auxiliary information, such as the background knowledge of items and their relationships [30]. Knowledge graphs can effectively represent entities and their related attributes or relationship information, so as to obtain a more accurate node representation and provide effective prior knowledge for learning tasks. Moreover, knowledge graphs support different connection types of relationships, which is conducive to the divergence of recommendation results and avoids limiting the results to a single type.

For example, KGCN [33] uses graph convolution to mine related attributes on an attribute-constructed knowledge graph to capture the correlation between items. KGAT [34] explicitly utilizes the high-order connectivity in the graph of item, side information a.k.a. knowledge graph in an end-to-end fashion to construct the prediction model. Xian et al. [35] modeled the recommendation problem as a graph-based **Markov Decision Process (MDP)**, and then utilized reinforcement learning for path inference over knowledge-graph-based user-item interactions. In the medical field, Choi et al. [8] proposed a **Graph-based Attention Model (GRAM).** GRAM utilizes the existing medical "knowledge", the medical ontologies with hierarchical information, such as the International Classification of Diseases Version 9 (ICD-9) ontology and the **Anatomical Therapeutic Chemical (ATC)** classification system, to learn robust representations for medical concept (diagnosis and medication), especially for those that rarely appeared in the **Electronic Health Records (EHR)**. These work demonstrated the feasibility of applying graph neural networks and information propagation mechanism for structural knowledge in recommendation systems.

Therefore, we proposed to improve the performance of POIR with knowledge graph-based modeling. We analyzed the Foursquare New York and Foursquare Tokyo data sets and found 3 important features which are valuable to build an effective knowledge-based model, including:

- Categorical information of locations represents more abstract and generalizable relationships among locations. Therefore, a knowledge graph based on the category hierarchy can help with effective location representations.
- Users demonstrate clear behavior patterns regarding the spatial and/or temporal interval information. A user is more likely to move between two locations with smaller time interval or geographic distance. Therefore, the spatial/temporal intervals should be quantified to refine the recommendations.
- Users demonstrate clear routine behavior patterns. A user is likely to regularly visit the same location or location category daily or weekly. Therefore, periodic patterns should be quantified to improve the recommendations.

To address the data sparsity problem in next POIR applications and maximize the benefits of the features found in check-in data sets, we proposed **CHA (Categorical Hierarchy-based Attention)** for next POIR which is an end-to-end knowledge graph-based predictive framework with the aid of the category hierarchy of POIs. Our main contributions include:

(1) We proposed to explore the category hierarchy of POIs to develop an attention-based knowledge graph to help learn robust location representations even when there is insufficient data.
(2) We developed a spatial-temporal decay LSTM, which explicitly models the influence of the time-interval and distance on next POIR, and a discrete Fourier Series-based periodic attention, which models users' innate periodic activities, to better facilitate the capturing of the personalized behavior pattern.
(3) We developed an end-to-end learning framework to integrate the aforementioned components.

The extensive experiments on real-world LBSN datasets show that our model has the strong capability of making accurate recommendations with the relatively sparse data. Our proposed approach significantly outperforms other state-of-the-art methods for both next and next new POI recommendation tasks.

The remainder of this article is organized as follows: Section 2 reviews the related work on POIR. Section 3 explains the proposed model, including the location embedding via attention-based knowledge graph, spatial-temporal decay LSTM framework, periodic attention, and the end-to-end training framework. Section 4 shows the experiment settings and results, then provides the visual interpretation of the effectiveness of our model. Finally, Section 5 concludes the article.

2 RELATED WORK

In comparison with general POIR that mainly exploits users' preferences on POIs, next POIR is even more challenging because it also studies the sequential relationship in the users' check-in history to predict their next move. Since the check-in behavior is closely related to spatial-temporal characteristics, various studies have worked on capturing the most pertinent piece of a check-in sequence to improve the recommendation performance. In this section, we discuss related work on both POIR and next POIR, which can be roughly categorised into two types: context-aware POIR and neural-network-based POIR.

Context-Aware POIR. The Markov-based methods proved to be effective in modeling the sequential correlation between the POIs and the users' preference. FPMC-LR [7] applies personalized Markov chains and matrix factorization to learn the transition matrix among POIs and the general taste of users, respectively. He et al. [12] extended FPMC-LR by explicitly utilizing the discretized check-in timestamps and the category of the locations to cluster/group the latent transition pattern, and then to infer the pattern-level transitions and the pattern distributions to recover the unobserved transition preference. Liu et al. [23] learned the latent representation for a location by using the Skip-gram model to capture the influence of its contextual information. Li et al. [20] introduced a time-decay factor into the tensor-based model to weigh the importance of the spanning time of two successive check-ins. He et al. [13] explicitly modeled the temporal intervals between POIs with a transition interval tensor to predict the users' next move and the time simultaneously. Hang et al. [11] proposed **embedding for dense heterogeneous graphs (EDHG)** to identify the challenges for time-aware POI prediction in more dense educational check-in data, which first constructed a heterogeneous graph, then captured features of connectivity and structural similarity for pairs of nodes. Based on the Skip-gram model, Chang et al. [4] proposed a **content-aware POI embedding (CAPE)** model which utilizes text content that provides information about the characteristics of a POI. Li et al. [19] proposed PPDM (Personalized Pattern Distribution Model), which utilizes personalized latent behavior patterns to learn personalized pattern distribution for each user.

Besides the temporal information, geographical information is also considered as an important factor [6, 18] due to the fact that the range of activities is limited in most people's daily life. Yin et al. [40] proposed a unified probabilistic generative method to model the joint effect of semantic and spatial-temporal patterns of users' check-ins. Zhao et al. [47] proposed the **spatial-temporal latent ranking (STELLAR)** method, which uses a ranking-based pairwise tensor factorization framework to capture the impact of time on successive POI recommendations. In order to capture geographical influence, Feng et al. [10] built a graph by splitting the POIs into a hierarchy of binary regions, and exploited POI2Vec to incorporate this geographical knowledge graph of POIs in learning latent representations. Wang et al. [31] used **Location-Sentiment-Aware Recommendation System (LSARS)** to accurately capture the users' check-in behaviors by adapting to the user's interest drift and crowd sentiments. Wang et al. [32] proposed GeoIE to capture the geographical influence between the two POIs by considering the influence, the susceptibility, and the physical distance of each place. Rahmani et al. [29] utilized the user's main region for activities and the correlation of each location within that region, which is fused into the logistic matrix factorization to capture the user's preference. Liu et al. [22] modeled the geographic features, such as regions and POI characteristics, which were fused with the generative adversarial networks. Qian et al. [27] utilized each spatial-temporal pair as a translation vector and modeled it as the connection between users and locations.

Moreover, the categorical information is also applied. For example, Chen et al. [5] and He et al. [12] proposed to use the categorical label of the location to enrich the feature of the representations. Li et al. [20] proposed a two-fold model, which learn a category transition first, then rank the locations inside the chosen category mainly according to the distance. To cope with the challenges of modeling non-linear user-POI interactions from implicit feedback and incorporating context information, Aliannejadi and Crestani [2] aimed to find the mapping between user tags and location taste keywords via a probabilistic model. As for hierarchical knowledge information, Zhang et al. [41] proposed HCT to recommend precise POIs with uncertain check-in records. Such uncertain check-in records (e.g., shopping mall) are defined as collective POIs in HCT, which exploits category transitions at different granularity to mine the users' preferred transition patterns in collective POIs, and then predict the users' preferred categories inside collective POIs. Instead, our model infuses categorical information into deep models by adaptively combining the ancestors of POIs via attention mechanism.

However, the multi-layer categorical hierarchy is not well exploited in most existing work, and we argue that it could offer more insight information to POIR tasks while alleviating the sparsity challenge. Therefore, in this article, we proposed an end-to-end framework, which starts with learning robust representations of POIs by fusing categorical hierarchy information with the knowledge graph, and then integrates the key features we mentioned in Section 3 to enhance the performance of next POIR.

Neural-Network-Based POIR. Deep models have also been widely applied in next POIR tasks [17, 21, 44], and they differ from each other because of (1) what kind of specific deep structures is utilized, such as RNN, LSTM, or specially designed; and (2) how the temporal information is incorporated, such as by leveraging on temporal representations or weighing the influence induced by the temporal information.

Specifically, Liu et al. [21] proposed ST-RNN to model local spatial and temporal contexts in each layer with time-specific transition matrices for different time intervals and distance-specific transition matrices for different geographical distances. Ma et al. [24] proposed a **self-attentive encoder (SAE)** and a **neighbor-aware decoder (NAD)** by adopting a multi-dimensional attention mechanism and the inner product of POI embeddings together with the **radial basis function (RBF)** kernel, respectively. To overcome the lack of geolocation information in social media posts, Chang et al. [3] proposed the **deep neural POI imputation model (DeepPIM)** to utilize available text, photo, user, and posting time information to automatically add missing POI information. Based on the attention mechanism, text information is encoded in textual features using RNN. DeepPIM also applies CNN pre-trained to classify objects in an image, and extracts visual features from the given photo information.

The sparsity of the user-POI matrix and cold-start problems have also been considered for POIR tasks. To address the challenges, Li et al. [15] defined three types of friends and incorporated three types of check-ins into matrix factorization model using two different loss functions. Li et al. [16] proposed a location-oriented method to adaptively model the missing data, upon which it formulates the squared error based loss. Yang et al. [37] leveraged the expressive neural networks to devise a general and principled **semi-supervised learning (SSL)** framework. Xie et al. [36] developed the **graph-based embedding (GE)** model to embed POI-POI, POI-Region, POI-Time, and POI-Word into a shared low-dimensional space. Qu et al. [28] adopted graph neural networks for neighborhood-neighborhood interactions, further incorporating knowledge-enhanced graphs for recommendation. The major difference between the existing work and our knowledge-enhanced model is that our model focuses more on the categorical hierarchy of POIs to address cold-start problem as well as to learn more accurate POI representations.

Generally, our work differs from these POIR studies in three main aspects: First, we learn the location embedding via attention-based knowledge graph, which includes the category hierarchy information of each location. Our model assigns different weights to different hierarchical categories for each location, resulting in a location-specific representation. Second, we apply spatial-temporal decay LSTM to capture a more accurate representation of each visit sequence which consists of locations with different time and geographic intervals. Last, we study and model the periodic behavior patterns of users and apply the periodic attention mechanism to further enhance the effectiveness of the model.

3 CHA FRAMEWORK

In this section, we first define the notations used in the solution. Thereafter, we introduce each key aspect in the end-to-end framework, including the attention mechanism for the category hierarchy, spatial-temporal decay LSTM, periodic attention, and the end-to-end training.

3.1 Notations

Let $\mathcal{L} = \{l_1, l_2, \ldots, l_{|\mathcal{L}|}\}$ be the set of all locations (POIs) from the visiting records. Let $l_u^i \in \mathcal{L}$ be the POI visited by user u at time t_i, and L_u be the sequence of POIs visited by user u before time t_{i+1}. Then, $L_u = < l_u^{t_1}, l_u^{t_2}, \ldots, l_u^{t_i} >$. The goal is to train the entire framework with the observed visiting history, so as to recommend top-N candidates as accurately as possible as the next POI $l_u^{t_{i+1}}$ for user u.

The hierarchy of categories is represented as a **directed acyclic graph (DAG)** $\mathcal{G}$ in which an edge denotes a parent-child relationship. A parent in $\mathcal{G}$ represents the generalization information over its children, providing a multi-scale view of locations. The leaf nodes in $\mathcal{G}$ are from the location set $\mathcal{L}$. Let $C = \{c_1, c_2, \ldots, c_{|C|}\}$ be the set of all non-leaf nodes.

3.2 Attention-based Knowledge Graph

This section explains the use of category hierarchy to learn robust location representations. The category hierarchy is obtained from the Foursquare Venue API.[1] In Foursquare, a location is represented in a maximally 7-layer hierarchy with one root node and one leaf node which is the location itself. In this design, there are up to five layers to represent the hierarchical information of the location categories.For instance, in Figure 3, the POI l_i is labeled as a 6-layer hierarchical category from the bottom to the top, where "Outdoors&Recreation" is the first-layer (more general) category.

Accordingly, CHA represents locations and their hierarchical category information in the graph $\mathcal{G}$. The nodes in $\mathcal{G}$ include locations (leaf nodes) and their categories (ancestor nodes). The hierarchy has no more than seven levels (Figure 3), and the nodes of each level have their own representations. That is, each location has up to seven different notations corresponding to its category levels. A location can be seen as a combination of itself and its upper-level ancestors in the hierarchy via an attention mechanism.

Unlike previous single-class representations of locations, the model takes full advantage of the hierarchical category of each location and learns to assign different attention to different classes, then obtains the final representations of locations. Thus, robust representations for the locations rarely appearing in the visit records could be generated by leveraging this categorical hierarchy attention, which helps to mitigate the sparsity challenge in next POIR. Meanwhile, the generated representations are interpretable and consistent with the categorical hierarchy.

[1] https://developer.foursquare.com/docs/build-with-foursquare/categories/.

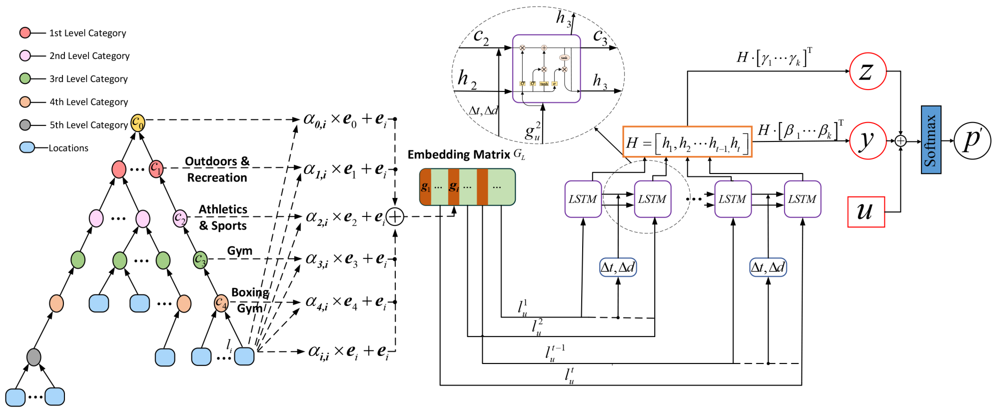


Fig. 3. The illustration of CHA. The leaf nodes l_i represent the POIs from the location set $\mathcal{L}$, while non-leaf nodes represent the categorical information. The final representation g_i of the leaf node l_i is computed by combining the basic embeddings e_i of l_i and its parent nodes e_0,e_1,e_2,e_3,e_4 via an attention mechanism. Then, the embedding matrix G_L for all POIs is used to embed the check-in sequence, which is then fed to our STD-LSTM model to acquire latent representation H. Finally, H will be used to calculate a score function with the help of periodic attention to estimate the probability of next location.

To address the data sparsity problem, when the location l_i is relatively rarely (less frequently) visited, the learning process could focus more (assign higher weight) on its ancestors (categories at different levels) as the ancestors can be learned more sufficiently and offer instructive information about their children. Consequently, in the embedding space, l_i should be close to the locations which share similar ancestors. The process can be facilitated via the attention mechanism and the end-to-end training as illustrated in Figure 3.

Each node in the graph is initialized randomly with an embedding vector $\mathbf{e}_i \in \mathbb{R}^m$, where m denotes the dimensionality. Therefore, $\mathbf{g}_i \in \mathbb{R}^m$, the embedding of the location l_i, is computed as a combination of itself and its ancestors:

$$\mathbf{g}_i = \sum_{j \in \mathcal{A}(i)} (\alpha_{ij}\mathbf{e}_j + \mathbf{e}_i), \sum_{j \in \mathcal{A}(i)} \alpha_{ij} = 1, \quad \alpha_{ij} \geq 0 \text{ for } j \in \mathcal{A}(i), \tag{1}$$

where $\mathcal{A}(i)$ denotes the set of indices regarding the location l_i and its category levels/ancestors. $\alpha_{ij} \in \mathbb{R}^+$ denotes the attention weight of e_j, and it is calculated via a softmax function:

$$\alpha_{ij} = \frac{\exp(s(\mathbf{e}_i, \mathbf{e}_j))}{\sum_{k \in \mathcal{A}(i)} \exp(s(\mathbf{e}_i, \mathbf{e}_k))} \tag{2}$$

where $s(\mathbf{e}_i, \mathbf{e}_j)$ denotes the scoring function indicating how well e_i and e_j match. $s(\mathbf{e}_i, \mathbf{e}_j)$ can be computed via a single-layer perceptron as:

$$s(\mathbf{e}_i, \mathbf{e}_j) = \mathbf{u}_a^\top \tanh\left(\mathbf{W}_a \begin{bmatrix} \mathbf{e}_i \\ \mathbf{e}_j \end{bmatrix} + \mathbf{b}_a\right) \tag{3}$$

where $\mathbf{W}_a \in \mathbb{R}^{r \times 2m}$ is the weight matrix for the concatenation of e_i and e_j, $\mathbf{b}_a \in \mathbb{R}^r$ and $\mathbf{u}_a \in \mathbb{R}^r$ are the bias vector and the weight vector, respectively. Then, the location-embedding matrix $\boldsymbol{G}$ is obtained, in which $\mathbf{g}_i$ is its ith column. $\boldsymbol{G}$ can be used to generate sequence representations including temporal and geographical characteristics for next POIR tasks.

3.3 Spatial-temporal Decay LSTM

LSTM (Long Short Term Memory), which extends the RNN (recurrent neural networks) structure by capturing long-term dependency information, could be used to capture the check-in history of users. LSTM decides what information should be discarded from the cell state $\mathbf{c}_{t-1}$ with a forget gate $\mathbf{f}_t$, then updates c_{t-1} to c_t with an input gate $\mathbf{i}_t$ and a memory cell $\tilde{\mathbf{c}}_t$, using the following equations iteratively:

$$\begin{aligned} \mathbf{f}_t &= \sigma(\mathbf{W}_f \cdot [\mathbf{h}_{t-1}, \mathbf{g}_t] + \mathbf{b}_f) \\ \mathbf{i}_t &= \sigma\left(\mathbf{W}_i \cdot [\mathbf{h}_{t-1}, \mathbf{g}_t] + \mathbf{b}_i\right) \\ \tilde{\mathbf{c}}_t &= \tanh\left(\mathbf{W}_c \cdot [\mathbf{h}_{t-1}, \mathbf{g}_t] + \mathbf{b}_c\right), \end{aligned} \tag{4}$$

where $\mathbf{h}_{t-1}$ is a hidden state output by $\mathbf{c}_{t-1} \in \mathbb{R}^{m \times m}$; $\mathbf{W}_f$, $\mathbf{W}_i$, and $\mathbf{W}_c$ are the corresponding weight matrices; $\mathbf{b}_f$, $\mathbf{b}_i$, and $\mathbf{b}_c$ are the bias vectors; and $\mathbf{g}_t$ denotes the location user u visited at time t. The new cell state is given as:

$$\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tilde{\mathbf{c}}_t. \tag{5}$$

In this process, the hidden state h_t is generated with an output gate o_t via sigmoid function:

$$\begin{aligned} \mathbf{o}_t &= \sigma\left(\mathbf{W}_o \left[\mathbf{h}_{t-1}, \mathbf{g}_t\right] + \mathbf{b}_o\right) \\ \mathbf{h}_t &= \mathbf{o}_t \odot \tanh\left(\mathbf{c}_t\right) \end{aligned} \tag{6}$$

where h_t is used to compute the score of each candidate location. Following the discussion in the introduction that the spatial/temporal intervals should be quantified to refine the POIR, we capture the geographical distance and time interval between two adjacent check-ins in the neural network. That is, for each pair of successive check-ins, an attention weight is assigned regarding the interval value to capture its impact on POIR.

The time interval between two successive visits can span from several seconds to several weeks in a user's history, but the second visit is generally more likely to happen with a short time interval. Therefore, the correlation between two visits is stronger if they have a shorter time interval than a longer interval as long as they do not belong to a periodic pattern. The same discussion applies to the distance between two successive visits that the correlation between them is stronger if they have a shorter distance [20].

However, LSTM is ill suited to deal with check-in data which consists of irregular continuous-time events as it implicitly assumes that the time intervals in a sequence are the same, which is also a challenge for the geographical distance. Moreover, it tends to weigh more on the "recent" event in the sequence, despite of long-term events even if they may have a relatively small time interval/distance to the current event which is usually considered as an important factor for predicting the next location in POIR tasks.

Therefore, LSTM should not be applied directly to learn the check-in sequence for next POIR as it cannot find precise correlations between two visits which are in the continuous spatial-temporal domain. Inspired by Time-LSTM, in this article, we propose an STD-LSTM model to address this challenge, which integrates a spatial-temporal decay function with the numerical value of the time interval Δt and the location distance Δd to Equation (5) to explicitly model the influence of the spatial/temporal intervals between two consecutive visits, as well as to enhance the robustness of LSTM, given as:

$$\begin{aligned} g(\Delta t, \Delta d) &= \sigma(-w(\Delta t + M_d \Delta d) + b) \\ \mathbf{c}_t &= g(\Delta t, \Delta d) \cdot \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tilde{\mathbf{c}}_t. \end{aligned} \tag{7}$$

$g(\Delta t, \Delta d)$ defines the spatial-temporal decay function, in which $w > 0$ is the decay speed, b indicates how to retain the historical information, $\Delta t = \{0, t_2 - t_1, t_3 - t_2, \ldots, t_i - t_{i-1}\}$ and we obtain Δd in a similar way. As the spatial intervals and the temporal intervals have different scales, we utilize a trainable transition matrix M_d to project the spatial intervals into the same embedding space of temporal intervals. As a result, the new cell state computation (Equation (7)) is capable of capturing the diminishing impact with the increase of the time interval and/or the location distance. In this way, STD-LSTM outputs a latent representation $H = \langle \mathbf{h}_{\mathbf{t}_1}, \mathbf{h}_{\mathbf{t}_2}, \ldots, \mathbf{h}_{\mathbf{t}_i} \rangle$ for the visiting sequence L_u of user u. The use of decay function helps to resolve LSTM's deficiency of handling the sequence with irregular patterns, such as unequal time intervals and continuous distance between visits. With the decay function, the influence of all locations in a check-in sequence can be quantified as a weight value. In this case, we utilize the attention mechanism, to compute the spatial-temporal-decay-based attention weight distribution according to the latent representation H. Specifically, to reinforce the finding of important information and overcome the shortcoming of LSTM, we assign weights to the historical check-ins which are vital to the next POI prediction as

$$
\begin{aligned}
b_{t_j} &= (\mathbf{h}_{\mathbf{t}_\mathbf{j}})^\top \cdot \hat{\mathbf{W}} \cdot \mathbf{g}_\mathbf{i} \\
\beta_{t_j} &= \frac{\exp(b_{t_j})}{\sum_{q \in \{t_1, \ldots t_i\}} \exp(b_q)}
\end{aligned}
\tag{8}
$$

where $\hat{\mathbf{W}} \in \mathbb{R}^{m \times m}$ is a trainable weight matrix, indicating how $\mathbf{h}_{\mathbf{t}_\mathbf{j}}$ correlates with $\mathbf{g}_\mathbf{i}$. Based on the attention computed with Equation (8), we can apply it to the hidden state of the location sequence and obtain the representation as

$$
\mathbf{y} = \sum_{k \in \{t_1, \ldots, t_i\}} \beta_k \cdot \mathbf{h}_k, \tag{9}
$$

where $\mathbf{y}$ captures the impact of each historical check-in location on the candidate (next POI) location for any spatial/temporal intervals between them.

3.4 Periodic Attention

In addition to the spatial and temporal correlation between the POIs captured by STD-LSTM, periodic behavior is another type of universal pattern in human's daily life [46], which is hidden in the check-in sequence awaiting discovery. For instance, most workers go to their workplaces in the morning, visit restaurants at noon, and return to their residences at night.

In general, cyclic functions, such as sine function, can be used to model periodic features. However, such a simple function cannot be applied directly for check-in history as human behaviors show complex patterns with multi-extremums with irregular periods. Therefore, it is difficult to train accurate and practical parameters, such as frequency and phase.

To address these challenges, we propose to extend STD-LSTM with **Discrete Fourier Series (DFS)**, which applies a truncated series of sine and cosine functions to approximate the periodic pattern of l_i in a user check-ins sequence and estimate the probability that the next location $l_u^{t_{i+1}}$ for user u is the location u previously visited at t_j, given as:

$$
\begin{aligned}
p\left(l_u^{t_{i+1}} = l_u^{t_j} | \tau_j\right) = \mu^0 + \sum_{n=1}^{N} \Bigg(\mu_n^s \cdot \sin\left(2\pi \frac{n}{N} \tau_j\right) \\
+ \mu_n^c \cdot \cos\left(2\pi \frac{n}{N} \tau_j\right)\Bigg),
\end{aligned}
\tag{10}
$$

where $\tau_j = t_{i+1} - t_j$ denotes the time interval between the $(i+1)$-th and jth check-ins of a user, μ^0 is **Direct Current (DC)** offset of the Fourier Transform, μ^s is the amplitude of sine, μ^c is the amplitude of cosine, and N determines the max base frequency of DFS, which also equals to the max length of periodicity we want to capture. μ^0, μ^s, and μ^c are the parameters of the entire framework.

In order to quantify the periodic influence of each POI in a user's check-in sequence, Equation (10) is used as the scoring function to compute the periodic attention as follows:

$$\eta_{t_j} = p\left(l_u^{t_{i+1}} = l_u^{t_j} | \tau_j\right) \tag{11}$$

$$\gamma_{t_j} = \frac{\exp(\eta_{t_j})}{\sum_{q \in \{t_1, \ldots, t_i\}} \exp(\eta_q)} \tag{12}$$

Integrating the periodic attention with the hidden output of the check-in sequence H, which is from the previous STD-LSTM, the impact of periodic factor on the location recommendation can be modeled as

$$\mathbf{z} = \sum_{k \in \{t_1, \ldots, t_i\}} \gamma_k \cdot \mathbf{h_k}, \tag{13}$$

where z indicates the representation of each check-in sequence under the influence of periodic behavior patterns, and we further utilize it to be an important contextual factor to improve POIR tasks.

3.5 End-to-end Training

Given the corresponding check-in sequence L_u of user u and the candidate POI $\mathbf{g_i}$, the score of recommending $\mathbf{g_i}$ in the proposed model is

$$s\left(u, L_u, l_i\right) = \left(\mathbf{u} + \mathbf{y} + \mathbf{z}\right)^\top \cdot \mathbf{g_i}. \tag{14}$$

where $\mathbf{u}$ is the embedding vector of u, and $\mathbf{y}$ and $\mathbf{z}$ are defined by Equation (9) and Equation (13), respectively. The score function considers the impact of three factors on a candidate (next POI) location $\mathbf{g_i}$, including $\mathbf{u}^\top\mathbf{g_i}$ for the user's preference for $\mathbf{g_i}$, $\mathbf{y}^\top\mathbf{g_i}$ for the correlation between historical check-in places and $\mathbf{g_i}$ regarding the spatial-temporal decay factor, and $\mathbf{z}^\top\mathbf{g_i}$ for the correlation between the check-in sequence and $\mathbf{g_i}$ regarding the periodic attention. Consequently, the probability of l_i being the next POI for u can be computed via a softmax function as:

$$p'\left(l_u^{t_{i+1}} = l_i\right) = \frac{\exp\left(s\left(\mathbf{u}, \mathbf{h_t}, \mathbf{g_i}\right)\right)}{\sum_{j=1}^{|\mathcal{L}|} \exp(s(\mathbf{u}, \mathbf{h_t}, \mathbf{g_j}))}. \tag{15}$$

The prediction loss of the entire framework is then calculated using the cross entropy as follows:

$$J = \sum_{case \in D} \sum_{j=1}^{n} -\hat{p}\left(l_u^{t_{i+1}} = l_j\right) \ln p'\left(l_u^{t_{i+1}} = l_j\right), \tag{16}$$

where $\hat{p}(l_u^{t_{i+1}} = l_j)$ is the probability that l_j is the next POI regarding the ground truth. D is the training set, and *case* refers to a batch of training data.

The model is then trained by minimizing the cross entropy loss J. That is, the goal of the model is to increase the probability $p'(l_u^{t_{i+1}} = l_*)$ and the score of location l_* regarding the ground truth $s\left(u, L_u, l_*\right)$.

4 EXPERIMENTS

Extensive experiments were employed to evaluate the proposed framework CHA for its use of graph-based location representations, spatial-temporal decay LSTM, and periodic attention behaviors in case of insufficient data available. This section first describes the experimental setup, and

Table 1. Dataset Summary

	#User	#POI	#Check-in
Fours.-NYC	1083	38333	227428
Fours.-TKY	2293	61857	573703

then compares our results with various baseline models on both next and next new POIR tasks. At the end, this section qualitatively explains the intuitive interpretations of our model.

4.1 Datasets and Key Parameters

Foursquare New York and Foursquare Tokyo, two popular check-in datasets [38] from real-world LBSNs, were used in the experiments. Table 1 lists the statistic summary of each dataset.

In the experiments, we split two datasets into two non-overlapping sets by using the 70% check-ins of each user as the training set and the remaining 30% check-ins as the test set. The dimensionality of $\mathbf{g_i}$ was set to be 60, and the learning rate and the regularization coefficient were both set as 0.001 empirically. The length of check-in records was set to be 10. Other parameters were learned during the training.

4.2 Metric

The output of the recommendation is a list of top-N predicted POIs for user u in descending order of the probability p'. Recall[2] is used to evaluate the performance of next POIR, which is defined as

$$Recall@N = \frac{1}{|U|} \sum_{u \in U} \frac{\left|S_u^N \cap S_{visited}\right|}{|S_{visited}|} \tag{17}$$

where U is the set of all users, S_u^N denotes the set of top-N POIs recommended to user u, $S_{visited}$ is the ground truth POI visited by user u. We report Recall@N with N $\in \{1, 5, 10, 20\}$.

Similarly, we define the recall metric for next new POIR as

$$Recall@N_{new} = \frac{1}{|U|} \sum_{u \in U} \frac{\left|S_u^N \cap S_{visited}^{new}\right|}{\left|S_{visited}^{new}\right|} \tag{18}$$

where $S_{visited}^{new}$ denotes the location set in which the user u visited at the first time next to the prior locations in each check-in sequence.

Mean Reciprocal Rank (MRR) is used to measure the quality of ranked lists. MRR is the average of the reciprocal sum of the ranking of all user-related items in the given results. When the next correct POI is highly ranked in the recommended list, the evaluation score is also high. To measure the performance of next POIR, MRR is defined as

$$MRR = \frac{1}{|U|} \sum_{u \in U} \frac{1}{rank_u} \tag{19}$$

where $rank_u$ denotes the index of $S_{visited}$ in S_u^N for each user u.

Similarly, we define MRR_{new} to measure the performance of next new POIR as

$$MRR_{new} = \frac{1}{|U|} \sum_{u \in U} \frac{1}{rank_u^{new}} \tag{20}$$

where $rank_u^{new}$ denotes the index of $S_{visited}^{new}$ in S_u^N for each user u.

[2]Precision is inappropriate in the experiments as there is only one correct answer in next POI recommendation tasks. The numerator in the precision equation will always be 1 even if the recommendation is successful. That is, the precision keeps on decreasing with the increase of N, and the precision cannot be higher than 1/N.

Table 2. Performance Comparison of Next POIR

Metrics	NYC							TKY						
	F.LR	ST-RNN	PPDM	TMCA	NEXT	POI2Vec	CHA	F.LR	ST-RNN	PPDM	TMCA	NEXT	POI2Vec	CHA
R@1	0.019	0.021	0.062	0.086	0.065	0.069	**0.126**	0.023	0.027	0.066	0.124	0.064	0.071	**0.137**
R@5	0.031	0.063	0.140	0.198	0.146	0.187	**0.307**	0.030	0.033	0.114	0.263	0.136	0.192	**0.330**
R@10	0.054	0.088	0.190	0.250	0.190	0.225	**0.384**	0.095	0.102	0.147	0.319	0.171	0.227	**0.406**
R@20	0.077	0.094	0.247	0.295	0.236	0.268	**0.452**	0.109	0.112	0.192	0.375	0.204	0.269	**0.472**
MRR	0.024	0.041	0.094	0.134	0.101	0.138	**0.207**	0.028	0.047	0.085	0.133	0.090	0.144	**0.223**

*F.LR: FPMC-LR.

In the experiments with MRR, we set N = 20 and the reciprocal rank of the ground truth as 0 if the rank is above the recommendation list size.

4.3 Baseline Models

The following baseline models are compared with CHA in the next and next new POIR tasks:

- *FPMC-LR* [7]: A matrix factorization-based shallow model leveraging the personalized Markov chains and localized regions.
- *ST-RNN* [21]: An RNN-based model considering different timestamps and distances for next POIR.
- *TMCA* [17]: An LSTM-based encoder-decoder framework integrating multi-level contextual factors including temporal information.
- *PPDM* [19]: A two-fold model which learns latent personalized pattern distribution with contextual features including spatial-temporal/categorical information for next and next new POIR.
- *NEXT* [44]: The model which designs a neural network for the tasks with DeepWalk-based [26] pre-trained POI embeddings as the input and incorporates temporal information and friendship of users.
- *POI2Vec* [10]: A knowledge graph-based latent representation model, which hierarchically splits POIs into different regions and uses a binary tree structure to model the check-in sequences with location and temporal information.
- *CHA-x*: A variation of CHA which ignores the attention-based representation learning of the categorical hierarchy, but uses a random initialization for POIs' embedding $\mathbf{g_i}$.
- *CHA-xz*: A variation of CHA which both ignores the knowledge graph and the periodic attention.
- *CHA-y*: A variation of CHA which removes $\mathbf{y}$ in Equation (14).

4.4 Prediction Performance of Next POIR

Table 2 shows the performance of each model on two datasets. The results show that CHA consistently and significantly outperforms all other approaches for both NYC dataset and TKY dataset.

For other models in comparison, the neural network-based TMCA and the knowledge-graph based POI2Vec appear to be the most competitive approaches, while the conventional matrix/tensor factorization-based prediction models perform poorly. It should be noted that the datasets provided in [38] select only active users who check in at least three times per week. Thus, the datasets are considered relatively dense, which favors the neural structured model.

Table 3. Performance Comparison of Next POIR Ablated in Three Ways

Metrics	NYC			TKY		
	CHA-x	CHA-y	CHA-xz	CHA-x	CHA-y	CHA-xz
R@1	0.117	0.116	0.060	0.108	0.116	0.076
R@5	0.273	0.290	0.139	0.254	0.291	0.176
R@10	0.334	0.371	0.180	0.310	0.365	0.222
R@20	0.383	0.441	0.227	0.358	0.429	0.267
MRR	0.185	0.193	0.096	0.172	0.194	0.121

Table 4. Performance Comparison with Different Training Ratios

Metrics	20%			40%			60%		
	TMCA	NEXT	CHA	TMCA	NEXT	CHA	TMCA	NEXT	CHA
R@1	0.045	0.056	**0.080**	0.079	0.067	**0.103**	0.079	0.060	**0.114**
R@5	0.109	0.124	**0.222**	0.182	0.161	**0.261**	0.185	0.144	**0.290**
R@10	0.137	0.157	**0.281**	0.228	0.206	**0.335**	0.231	0.190	**0.371**
R@20	0.163	0.189	**0.330**	0.269	0.249	**0.394**	0.274	0.232	**0.437**
MRR	0.093	0.086	**0.148**	0.158	0.109	**0.174**	0.160	0.096	**0.193**

NYC Dataset.

Table 5. Performance Comparison with Different Training Ratios

Metrics	20%			40%			60%		
	TMCA	NEXT	CHA	TMCA	NEXT	CHA	TMCA	NEXT	CHA
R@1	0.034	0.052	**0.111**	0.075	0.056	**0.138**	0.116	0.058	**0.138**
R@5	0.084	0.118	**0.283**	0.180	0.122	**0.319**	0.251	0.128	**0.326**
R@10	0.104	0.150	**0.355**	0.226	0.155	**0.392**	0.305	0.161	**0.403**
R@20	0.121	0.181	**0.415**	0.272	0.188	**0.457**	0.359	0.195	**0.471**
MRR	0.054	0.086	**0.187**	0.129	0.082	**0.218**	0.169	0.092	**0.222**

TKY Dataset.

To investigate the effectiveness of different components in CHA, we developed the corresponding ablation models. Table 3 demonstrates that categorical hierarchy information shows significant impact on next POI recommendations. Without the attention-based knowledge graph, although our model is still superior to baseline models, there remains a large performance penalty in comparison with the complete CHA model. Compared with CHA-x, the performance of CHA-xz declines substantially, illustrating that the periodic attention mechanism is the essential component of our model.

To evaluate the performance in case of insufficient data, we tested CHA, TMCA, and NEXT with different training-to-test ratios. We tested the training/testing ratio at 20/80, 40/60, and 60/40 for both datasets. Table 4 and Table 5 show the results on NYC and TKY datasets with different

Table 6. Performance Comparison of Next POIR with Different Lengths

LEN	NYC					TKY				
	R@1	R@5	R@10	R@20	MRR	R@1	R@5	R@10	R@20	MRR
6	0.123	0.297	0.370	0.428	0.200	0.133	0.311	0.385	0.450	0.213
8	0.123	0.297	0.374	0.439	0.199	0.135	0.322	0.397	0.463	0.218
10	0.126	0.307	0.384	0.452	0.207	0.137	0.330	0.406	0.472	0.223
12	0.118	0.302	0.386	0.457	0.200	0.137	0.334	0.405	0.473	0.225

LEN: the length of check-in records.

Table 7. Performance Comparison of Next New POIR

Metrics	NYC									TKY								
	F.LR	ST-RNN	PPDM	TMCA	NEXT	CHA-x	CHA-y	CHA-xz	CHA	F.LR	ST-RNN	PPDM	TMCA	NEXT	CHA-x	CHA-y	CHA-xz	CHA
R@1	0.012	0.012	0.022	0.044	0.041	0.033	0.031	0.015	**0.047**	0.016	0.018	0.021	0.045	**0.047**	0.020	0.027	0.021	0.043
R@5	0.025	0.045	0.082	0.141	0.102	0.136	0.153	0.053	**0.169**	0.021	0.025	0.051	0.149	0.102	0.106	0.135	0.081	**0.160**
R@10	0.040	0.061	0.135	0.176	0.144	0.192	0.235	0.080	**0.246**	0.065	0.078	0.087	0.194	0.130	0.158	0.205	0.115	**0.236**
R@20	0.067	0.072	0.196	0.212	0.187	0.242	0.310	0.114	**0.319**	0.078	0.086	0.135	0.250	0.160	0.209	0.272	0.153	**0.309**
MRR	0.018	0.027	0.060	0.082	0.065	0.079	0.087	0.034	**0.103**	0.020	0.038	0.045	0.085	0.072	0.059	0.076	0.049	**0.098**

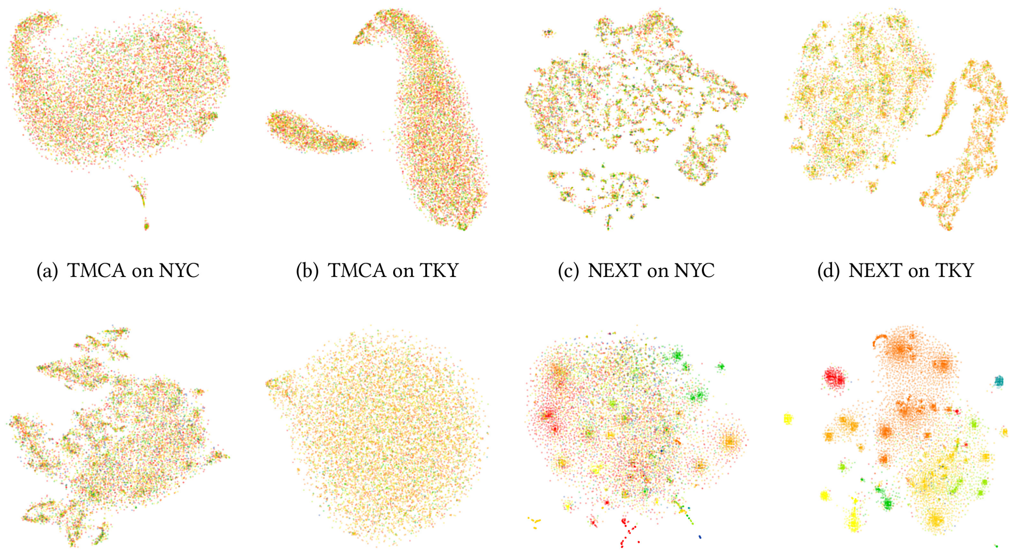
*F.LR: FPMC-LR.

training/testing ratios, respectively. The results show that CHA outperforms TMCA and NEXT on all ratios. It should be noted that NEXT has better results than TMCA at 20/80 ratio, indicating that TMCA is most vulnerable to insufficient training data. Moreover, it is evident that even with the low volume (as low as 20%) of the data for training, CHA can still achieve good predictions. The results prove the robustness and effectiveness of our method.

As the proposed model utilizes the historical check-in sequence to predict the next location each user is most likely to visit, increasing the length of the check-in sequence will generally provide more behavior information to learn, thereby improving the model capability. Table 6 shows the effects of various sequence lengths on the performance. In general, the model performs better along with the increase of sequence length, indicating that the model learns more spatial-temporal- and periodic-based information. Specifically, the model has competitive performance with the length at about 10. Thereafter, the performance of the model remains stable, or slightly worse. It is recommended to select an appropriate length when using the model as the overlength of the check-in sequence may lead to overfitting problem. Moreover, the overlength makes the model more complicated with higher overhead of time and memory usage in training.

4.5 Prediction Performance of Next New POIR

Table 7 shows the performance of next new POIR. The results show that CHA performs well regarding both Recall and MRR metrics on two datasets. Specifically, CHA has better results than TMCA with 5%–50% improvement, and significantly improves the performance of PPDM by around 62%–147%. It shows that (1) the attention-based knowledge graph can utilize the category hierarchy to make more precise representations on new location embeddings for users, (2) spatial-temporal-decay-based LSTM makes better use of the contextual information of each visit sequence so that the locations with different time and geographic intervals have different weights in the sequence, and (3) periodic attention helps to increase the weight of a location which has an approximate periodicity so as to better predict a new location that is to be visited at a specific time.



(a) TMCA on NYC (b) TMCA on TKY (c) NEXT on NYC (d) NEXT on TKY

(e) POI2Vec on NYC (f) POI2Vec on TKY (g) CHA on NYC (h) CHA on TKY

Fig. 4. t-SNE scatterplots of the POIs representations learned from TMCA, NEXT, STRNN, and CHA on two datasets. The colors of the visualized nodes represent the first-level category types in the hierarchy.

Table 7 demonstrates that, without any part of our model, the prediction of next new POIR will perform worse, showing the importance of attention-based knowledge graph, spatial-temporal-decay LSTM, and periodic attention in next new POIR tasks. It is evident that the results are consistent with the observations in the ablation study in Section 4.4.

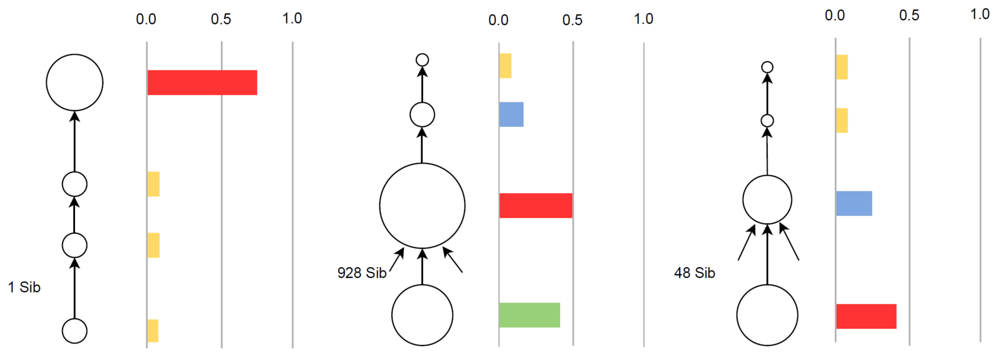
TMCA has better results than PPDM by 8%–100% on NYC dataset and by 85%–192% on TKY dataset, and NEXT outperforms PPDM by 7%–86% on NYC dataset (excluding recall@20) and by 19%–124% on TKY data set, indicating that the importance of utilizing neural networks and contextual information in boosting the performance of predicting new location for each user. Moreover, PPDM outperforms FPMC-LR and ST-RNN, demonstrating that learning personalized pattern distribution by capturing latent behavior patterns of users has the ability to discover more accurate new POIs in next new POIR tasks.

4.6 Interpretable Observations

This section visually demonstrates the effectiveness of using location representations, categorical hierarchy attention, and periodic attention in our model. It provides the interpretations of the observations on applying the model for POIR.

4.6.1 Location-Specific Representation. To evaluate the interpretability of location embedding, we applied t-SNE [25] to visualize each representation $\mathbf{g_i}$ in the entire location set learned by different models. Figure 4 show the scatterplots of POIs representations learned by CHA and its most competitive baselines, TMCA, NEXT, and POI2Vec on both NYC and TKY datasets. The colors of the dots represent the first-level categories in the hierarchy of location representations.

It is clear that the representations learned by CHA are significantly more consistent regarding the categorical knowledge of the hierarchy, while the embeddings learned by the most competitive baselines (TMCA, NEXT, and POI2Vec) are intertwined with each other. Based on the quantitative



(a) Location ID: 1157 (1) (b) Location ID: 24743 (1) (c) Location ID: 16048 (146)

Fig. 5. Category Hierarchy Attention Behaviors.

evaluations we have shown in Table 2, we can conclude that CHA produces more interpretable representations, which in turn help to deal with the data sparsity issue and to make more accurate predictions.

4.6.2 Categorical Hierarchy Attention. Our model learns an attention-based knowledge graph (Figure 3) to represent locations. The hierarchical-category-based attention learned in the model can be explained intuitively based on the data availability and the structure of knowledge graph $\mathcal{G}$. The attention weights of each location can be computed with Equation (1). Figure 5 depicts hierarchical-category-based attention weights for three representative locations when performing prediction tasks. In each subfigure, the bottom node represents the location while the upper nodes are its ancestors in the categorical hierarchy. The size of the node visualizes the amount of attention it receives, which is also shown in the corresponding bar chart. The number on the left side of each subfigure indicates the number of siblings the location has. The number in the parenthesis next to the location ID is the frequency of the location in the training data. The root of the knowledge graph $\mathcal{G}$ is not considered as it has no actual use.

The interpretation of the location attention graphs is:

(1) Location 1157 in Figure 5(a) is rarely observed in the dataset (only once) and has only one sibling. Therefore, in this case, most of the information is derived from its highest ancestor.
(2) Location 24743 in Figure 5(b) is also rarely observed (only once) but has a large number of siblings (928). In this case, the parent of Location 24743 receives strong attention because it aggregates sufficient samples from its children to learn a more accurate representation. In contrast with Location 1157 which has only one sibling, Location 24743 itself also learns a higher weight to help distinguish itself from its siblings.
(3) Location 16048 in Figure 5(c) is frequently observed (146 occurrences) but has a relatively small number of siblings in comparison with Location 24743. Thus, in this case, CHA assigns very strong attention to the location, which is reasonable as the more we observe a location, the stronger our confidence becomes.

4.6.3 Periodic Attention. Periodic attention is used in our model to capture periodic behavior patterns of users. We selected the trajectories of two users in NYC to demonstrate the effect of periodic attention on next POIR.

For the user u_1, a sequence of six check-in records is illustrated in Figure 6. The boxes on the right side show the details of each visit, including address, time, and the value of period attention

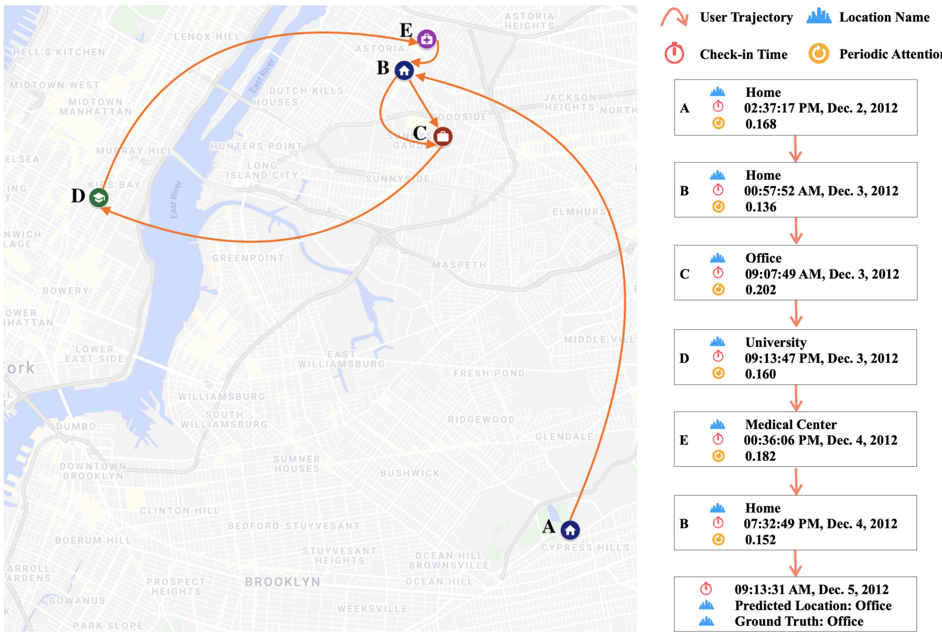


Fig. 6. User u_1's Period Attention Trajectory.

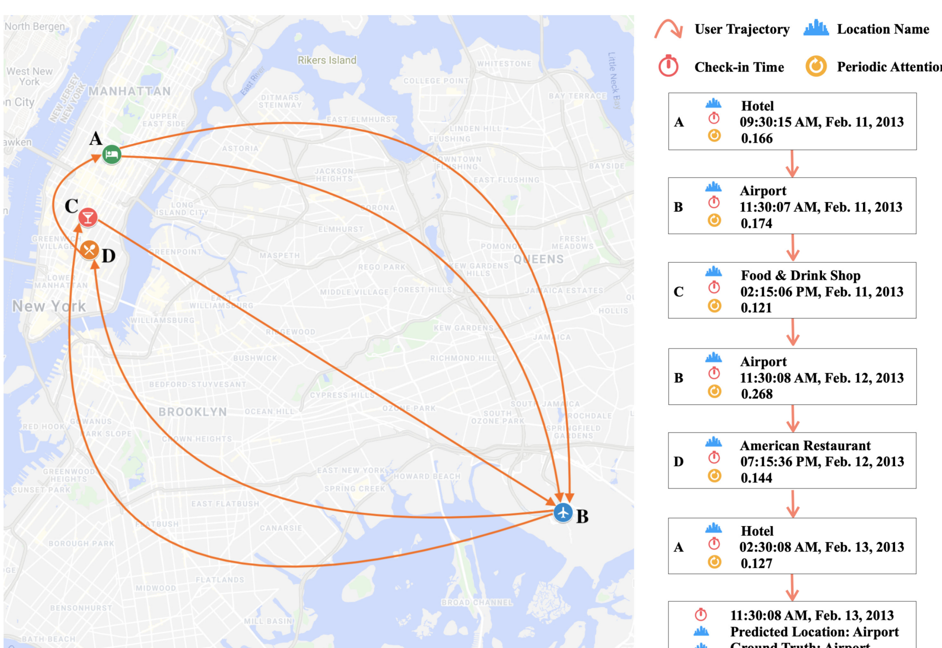


Fig. 7. User u_2's Period Attention Trajectory.

computed by Equation (12). u_1 first departed from the home in Brooklyn and finally arrived at the home in Astoria, Queens. As the user visited the office at around 09:10 AM on different days, a periodic pattern is detected. In comparison, the user had no obvious periodic visits to other places in the sequence. Therefore, the office has the highest attention weight in our model, and then it is selected as the prediction of next visit.

Similarly, Figure 7 illustrates the user u_2's trajectory. The records show that u_2 periodically visited the airport at 11:30 AM. Correspondingly, our model assigns a higher attention score for the airport than other places in the sequence, and makes the correct prediction that the airport is the next visit for u_2. The prediction conforms to the user's actual activity and proves the effectiveness of using periodic attention.

5 CONCLUSION

With the prevalence of location-based services, the demand of accurate location predictions in businesses has driven the fast growth of POI recommendations in both academia and industry. In this paper, we addressed 2 specific challenges or application domains in POIR, including next POIR, which predicts the very next visit of a user given the current location, and next new POIR, which predicts new locations for a user as next POIR.

After investigating the real-world data collected by LBSNs, such as Foursquare New York and Foursquare Tokyo datasets, we found three important features, including the importance of categorical information of locations, the spatial/temporal intervals between check-ins, and the periodic behavior patterns of users, which can help to boost the performance of next (new) POIR given the data sparsity challenge.

Following these features, we proposed an end-to-end framework CHA for next and next new POI recommendation tasks. Leveraging the knowledge of the category hierarchy, CHA introduces a graph-based attention mechanism for a robust and interpretable representation learning of locations, which also effectively addresses the data sparsity challenge in POIR tasks. CHA further applies the spatial-temporal decay LSTM and the periodic attention to better capture the personalized behavior patterns.

With the prevalence of location-based services, the demand of accurate location predictions in businesses has driven the fast growth of POI recommendations in both academia and industry. In this paper, we addressed 2 specific challenges or application domains in POIR, including next POIR, which predicts the very next visit of a user given the current location, and next new POIR, which predicts new locations for a user as next POIR.

After investigating the real-world data collected by LBSNs, such as Foursquare New York and Foursquare Tokyo datasets, we found three important features, including the importance of categorical information of locations, the spatial/temporal intervals between check-ins, and the periodic behavior patterns of users, which can help to boost the performance of next (new) POIR given the data sparsity challenge.

Following these features, we proposed an end-to-end framework CHA for next and next new POI recommendation tasks. Leveraging the knowledge of the category hierarchy, CHA introduces a graph-based attention mechanism for a robust and interpretable representation learning of locations, which also effectively addresses the data sparsity challenge in POIR tasks. CHA further applies the spatial-temporal decay LSTM and the periodic attention to better capture the personalized behavior patterns.

ACKNOWLEDGMENTS

We thank the associate editor and the anonymous reviewers for their insightful and detailed comments. Furthermore, we would like to thank Mr. Li Zhang for the help of debugging our evaluation script. Finally, we would also like to thank Mr. Hao Wu and Miss Yajing Gao for valuable discussions and the help of proof reading the manuscript.

REFERENCES

[1] Abdulrahman Al-Molegi, Mohammed Jabreel, and Baraq Ghaleb. 2016. STF-RNN: Space time features-based recurrent neural network for predicting people next location. In *Proceedings of the 2016 IEEE Symposium Series on Computational Intelligence (SSCI 2016)* (Athens, Greece, December 6–9, 2016). 1–7.

[2] Mohammad Aliannejadi and Fabio Crestani. 2018. Personalized context-aware point of interest recommendation. *ACM Trans. Inf. Syst.* 36, 4 (2018), 45:1–45:28. https://doi.org/10.1145/3231933

[3] Buru Chang, Yonggyu Park, Seongsoon Kim, and Jaewoo Kang. 2018. DeepPIM: A deep neural point-of-interest imputation model. *Inf. Sci.* 465 (2018), 61–71. https://doi.org/10.1016/j.ins.2018.06.065

[4] Buru Chang, Yonggyu Park, Donghyeon Park, Seongsoon Kim, and Jaewoo Kang. 2018. Content-aware hierarchical point-of-interest embedding model for successive POI recommendation. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence, (IJCAI 2018)* (July 13–19, 2018, Stockholm, Sweden). 3301–3307.

[5] Xuefeng Chen, Yifeng Zeng, Gao Cong, Shengchao Qin, Yanping Xiang, and Yuanshun Dai. 2015. On information coverage for location category based point-of-interest recommendation. In *Proceedings of the 29th AAAI Conference on Artificial Intelligence* (January 25–30, 2015, Austin, Texas). 37–43.

[6] Chen Cheng, Haiqin Yang, Irwin King, and Michael R. Lyu. 2012. Fused matrix factorization with geographical and social influence in location-based social networks. In *Proceedings of the 26th AAAI Conference on Artificial Intelligence* (July 22–26, 2012, Toronto, Ontario, Canada).

[7] Chen Cheng, Haiqin Yang, Michael R. Lyu, and Irwin King. 2013. Where you like to go next: Successive point-of-interest recommendation. In *Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI 2013)* (Beijing, China, August 3–9, 2013*)*. 2605–2611.

[8] Edward Choi, Mohammad Taha Bahadori, Le Song, Walter F. Stewart, and Jimeng Sun. 2017. GRAM: Graph-based attention model for healthcare representation learning. In *Proceedings of the 23rd ACM SIGKDD* (Halifax, NS, Canada, August 13–17, 2017). 787–795.

[9] Chaoran Cui, Jialie Shen, Liqiang Nie, Richang Hong, and Jun Ma. 2017. Augmented collaborative filtering for sparseness reduction in personalized POI Recommendation. *ACM Trans. Intell. Syst. Technol.* 8, 5 (2017), 71:1–71:23. https://doi.org/10.1145/3086635

[10] Shanshan Feng, Gao Cong, Bo An, and Yeow Meng Chee. 2017. POI2Vec: Geographical latent representation for predicting future visitors. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence* (February 4–9, 2017, San Francisco, CA), Satinder P. Singh and Shaul Markovitch (Eds.). AAAI Press, 102–108. http://aaai.org/ocs/index.php/AAAI/AAAI17/paper/view/14902

[11] Mengyue Hang, Ian Pytlarz, and Jennifer Neville. 2018. Exploring student check-in behavior for improved point-of-interest prediction. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD 2018*) (London, UK, August 19–23, 2018), Yike Guo and Faisal Farooq (Eds.). ACM, 321–330. https://doi.org/10.1145/3219819.3219902

[12] Jing He, Xin Li, Lejian Liao, Dandan Song, and William K. Cheung. 2016. Inferring a personalized next point-of-interest recommendation model with latent behavior patterns. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence* (February 12–17, 2016, Phoenix, AZ). 137–143.

[13] Jing He, Xin Li, Lejian Liao, and Mingzhong Wang. 2018. Inferring continuous latent preference on transition intervals for next point-of-interest recommendation. In *Proceedings of the ECML PKDD 2018* (Dublin, Ireland, September 10–14, 2018). 741–756. https://doi.org/10.1007/978-3-030-10928-8_44

[14] Dejiang Kong and Fei Wu. 2018. HST-LSTM: A hierarchical spatial-temporal long-short term memory network for location prediction. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI 2018)* (July 13–19, 2018, Stockholm, Sweden). 2341–2347.

[15] Huayu Li, Yong Ge, Richang Hong, and Hengshu Zhu. 2016. Point-of-interest recommendations: learning potential check-ins from friends. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (San Francisco, CA, August 13–17, 2016), Balaji Krishnapuram, Mohak Shah, Alexander J. Smola, Charu C. Aggarwal, Dou Shen, and Rajeev Rastogi (Eds.). ACM, 975–984. https://doi.org/10.1145/2939672.2939767

[16] Huayu Li, Yong Ge, Defu Lian, and Hao Liu. 2017. Learning user's intrinsic and extrinsic interests for point-of-interest recommendation: A unified approach. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI 2017)* (Melbourne, Australia, August 19–25, 2017), Carles Sierra (Ed.). 2117–2123. https://doi.org/10.24963/ijcai.2017/294

[17] Ranzhen Li, Yanyan Shen, and Yanmin Zhu. 2018. Next point-of-interest recommendation with temporal and multi-level context attention. In *Proceedings of the IEEE International Conference on Data Mining (ICDM 2018) (Singapore, November 17–20, 2018)*. 1110–1115. https://doi.org/10.1109/ICDM.2018.00144

[18] Xutao Li, Gao Cong, Xiaoli Li, Tuan-Anh Nguyen Pham, and Shonali Krishnaswamy. 2015. Rank-GeoFM: A ranking based geographical factorization method for point of interest recommendation. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Santiago, Chile, August 9–13, 2015). 433–442.

[19] Xin Li, Dongcheng Han, Jing He, Lejian Liao, and Mingzhong Wang. 2019. Next and next new POI recommendation via latent behavior pattern inference. *ACM Trans. Inf. Syst.* 37, 4 (2019), 46:1–46:28. https://doi.org/10.1145/3354187

[20] Xin Li, Mingming Jiang, Huiting Hong, and Lejian Liao. 2017. A time-aware personalized point-of-interest recommendation via high-order tensor factorization. *ACM Trans. Inf. Syst.* 35, 4 (2017), 31:1–31:23.

[21] Qiang Liu, Shu Wu, Liang Wang, and Tieniu Tan. 2016. Predicting the next location: A recurrent model with spatial and temporal contexts. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence* (February 12–17, 2016),. 194–200.

[22] Wei Liu, Zhi-Jie Wang, Bin Yao, and Jian Yin. 2019. Geo-ALM: POI recommendation by fusing geographical information and adversarial learning mechanism. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI 2019)* (Macao, China, August 10–16, 2019). 1807–1813.

[23] Xin Liu, Yong Liu, and Xiaoli Li. 2016. Exploring the context of locations for personalized location recommendations. In *Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI 2016)* (New York, NY, 9–15 July 2016), Subbarao Kambhampati (Ed.). IJCAI/AAAI Press, 1188–1194. http://www.ijcai.org/Abstract/16/172

[24] Chen Ma, Yingxue Zhang, Qinglong Wang, and Xue Liu. 2018. Point-of-interest recommendation: Exploiting self-attentive autoencoders with neighbor-aware influence. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management (CIKM 2018)* (Torino, Italy, October 22–26, 2018), Alfredo Cuzzocrea, James Allan, Norman W. Paton, Divesh Srivastava, Rakesh Agrawal, Andrei Z. Broder, Mohammed J. Zaki, K. Selçuk Candan, Alexandros Labrinidis, Assaf Schuster, and Haixun Wang (Eds.). ACM, 697–706. https://doi.org/10.1145/3269206.3271733

[25] Laurens van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-SNE. *Journal of Machine Learning Research* 9, (Nov. 2008), 2579–2605.

[26] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. 2014. DeepWalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD* (New York, NY, August 24–27, 2014). 701–710.

[27] Tieyun Qian, Bei Liu, Quoc Viet Hung Nguyen, and Hongzhi Yin. 2019. Spatiotemporal representation learning for translation-based POI recommendation. *ACM Trans. Inf. Syst.* 37, 2 (2019), 18:1–18:24. https://doi.org/10.1145/3295499

[28] Yanru Qu, Ting Bai, Weinan Zhang, Jian-Yun Nie, and Jian Tang. 2019. An end-to-end neighborhood-based interaction model for knowledge-enhanced recommendation. *CoRR* abs/1908.04032 (2019). arXiv:1908.04032 http://arxiv.org/abs/1908.04032

[29] Hossein A. Rahmani, Mohammad Aliannejadi, Sajad Ahmadian, Mitra Baratchi, Mohsen Afsharchi, and Fabio Crestani. 2019. LGLMF: Local Geographical based logistic matrix factorization model for POI recommendation. In *Proceedings of the Information Retrieval Technology - 15th Asia Information Retrieval Societies Conference (AIRS 2019), Hong Kong, November 7–9, 2019, Proceedings (Lecture Notes in Computer Science)*, Fu Lee Wang, Haoran Xie, Wai Lam, Aixin Sun, Lun-Wei Ku, Tianyong Hao, Wei Chen, Tak-Lam Wong, and Xiaohui Tao (Eds.), Vol. 12004. Springer, 66–78. https://doi.org/10.1007/978-3-030-42835-8_7

[30] John K. Tarus, Zhendong Niu, and Ghulam Mustafa. 2018. Knowledge-based recommendation: A review of ontology-based recommender systems for e-learning. *Artif. Intell. Rev.* 50, 1 (2018), 21–48.

[31] Hao Wang, Yanmei Fu, Qinyong Wang, Hongzhi Yin, Changying Du, and Hui Xiong. 2017. A location-sentiment-aware recommender system for both home-town and out-of-town users. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Halifax, NS, Canada, August 13–17, 2017). ACM, 1135–1143. https://doi.org/10.1145/3097983.3098122

[32] Hao Wang, Huawei Shen, Wentao Ouyang, and Xueqi Cheng. 2018. Exploiting POI-specific geographical influence for point-of-interest recommendation. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI 2018)* (July 13–19, 2018, Stockholm, Sweden), Jérôme Lang (Ed.). ijcai.org, 3877–3883. https://doi.org/10.24963/ijcai.2018/539

[33] Hongwei Wang, Miao Zhao, Xing Xie, Wenjie Li, and Minyi Guo. 2019. Knowledge graph convolutional networks for recommender systems. In *Proceedings of the World Wide Web Conference (WWW 2019)* (San Francisco, CA, May 13–17, 2019). 3307–3313.

[34] Xiang Wang, Xiangnan He, Yixin Cao, Meng Liu, and Tat-Seng Chua. 2019. KGAT: Knowledge graph attention network for recommendation. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD 2019)* (Anchorage, AK, August 4–8, 2019). 950–958.

[35] Yikun Xian, Zuohui Fu, S. Muthukrishnan, Gerard de Melo, and Yongfeng Zhang. 2019. Reinforcement knowledge graph reasoning for explainable recommendation. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2019)* (Paris, France, July 21–25, 2019), Benjamin Piwowarski,

Max Chevalier, Éric Gaussier, Yoelle Maarek, Jian-Yun Nie, and Falk Scholer (Eds.). ACM, 285–294. https://doi.org/10.1145/3331184.3331203

[36] Min Xie, Hongzhi Yin, Hao Wang, Fanjiang Xu, Weitong Chen, and Sen Wang. 2016. Learning graph-based POI embedding for location-based recommendation. In *Proceedings of the 25th ACM International Conference on Information and Knowledge Management (CIKM 2016)* (Indianapolis, IN, October 24–28, 2016), Snehasis Mukhopadhyay, ChengXiang Zhai, Elisa Bertino, Fabio Crestani, Javed Mostafa, Jie Tang, Luo Si, Xiaofang Zhou, Yi Chang, Yunyao Li, and Parikshit Sondhi (Eds.). ACM, 15–24.

[37] Carl Yang, Lanxiao Bai, Chao Zhang, Quan Yuan, and Jiawei Han. 2017. Bridging collaborative filtering and semi-supervised learning: A neural approach for POI recommendation. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Halifax, NS, Canada, August 13–17, 2017). ACM, 1245–1254. https://doi.org/10.1145/3097983.3098094

[38] Dingqi Yang, Daqing Zhang, Vincent W. Zheng, and Zhiyong Yu. 2015. Modeling user activity preference by leveraging user spatial temporal characteristics in LBSNs. *IEEE Trans. Systems, Man, and Cybernetics: Systems* 45, 1 (2015), 129–142. https://doi.org/10.1109/TSMC.2014.2327053

[39] Zijun Yao, Yanjie Fu, Bin Liu, Yanchi Liu, and Hui Xiong. 2016. POI recommendation: A temporal matching between POI popularity and user regularity. In *Proceedings of the IEEE 16th International Conference on Data Mining (ICDM 2016)* (December 12–15, 2016, Barcelona, Spain). 549–558.

[40] Hongzhi Yin, Bin Cui, Xiaofang Zhou, Weiqing Wang, Zi Huang, and Shazia W. Sadiq. 2016. Joint modeling of user check-in behaviors for real-time point-of-interest recommendation. *ACM Trans. Inf. Syst.* 35, 2 (2016), 11:1–11:44. https://doi.org/10.1145/2873055

[41] Lu Zhang, Zhu Sun, Jie Zhang, Horst Kloeden, and Felix Klanner. 2020. Modeling hierarchical category transition for next POI recommendation with uncertain check-ins. *Inf. Sci.* 515 (2020), 169–190.

[42] Richong Zhang, Yongyi Mao, and Weihua Zhao. 2020. Knowledge graphs completion via probabilistic reasoning. *Inf. Sci.* 521 (2020), 144–159.

[43] Richong Zhang, Samuel Mensah, Fanshuang Kong, Zhiyuan Hu, Yongyi Mao, and Xudong Liu. 2020. Pairwise link prediction model for out of vocabulary knowledge base entities. *ACM Trans. Inf. Syst.* 38, 4 (2020), 36:1–36:28.

[44] Zhiqian Zhang, Chenliang Li, Zhiyong Wu, Aixin Sun, Dengpan Ye, and Xiangyang Luo. 2020. NEXT: A neural network framework for next POI recommendation. *Frontiers Comput. Sci.* 14, 2 (2020), 314–333. https://doi.org/10.1007/s11704-018-8011-2

[45] Pengpeng Zhao, Haifeng Zhu, Yanchi Liu, Zhixu Li, Jiajie Xu, and Victor S. Sheng. 2018. Where to go next: A spatio-temporal LSTM model for next POI recommendation. *CoRR* abs/1806.06671 (2018).

[46] Shenglin Zhao, Irwin King, and Michael R. Lyu. 2016. A survey of point-of-interest recommendation in location-based social networks. *CoRR* abs/1607.00647 (2016). arXiv:1607.00647 http://arxiv.org/abs/1607.00647

[47] Shenglin Zhao, Tong Zhao, Haiqin Yang, Michael R. Lyu, and Irwin King. 2016. STELLAR: Spatial-temporal latent ranking for successive point-of-interest recommendation. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence* (February 12–17, 2016, Phoenix, AZ), Dale Schuurmans and Michael P. Wellman (Eds.). AAAI Press, 315–322. http://www.aaai.org/ocs/index.php/AAAI/AAAI16/paper/view/12249.

Received September 2020; revised April 2021; accepted April 2021