Received: Added at production | Revised: Added at production | Accepted: Added at production

DOI: xxx/xxxx

ARTICLE TYPE

An Attention-based Category-aware GRU Model for Next POI Recommendation

Yuwen Liu[1] | Aixiang Pei[2] | Fan Wang[1] | Yihong Yang[1] | Xuyun Zhang[3] | Hao Wang[4] | Hongning Dai[5] | Lianyong Qi*[1] | Rui Ma[6]

[1]School of Computer Science, Qufu Normal University, Rizhao, China
[2]Weifang Key Laboratory of Blockchain on Agricultural Vegetables, Weifang University of Science and Technology, Shouguang, China
[3]Department of Computing, Macquarie University, Sydney, Australia
[4]Department of Computer Science, Norwegian University of Science and Technology, Norway
[5]Faculty of information technology, Macau University of Science and Technology, Macau, China
[6]General Education Department, Shandong First Medical University (Shandong Academy of Medical Sciences), Taian, China

Correspondence
*Lianyong Qi,School of Computer Science, Qufu Normal University, Rizhao 276800, China. Email: yuwenliu97@gmail.com

Summary

With the continuous accumulation of users' check-in data, we can gradually capture users' behavior patterns and mine users' preferences. Based on this, the next point-of-interest (POI) recommendation has attracted considerable attention. Its main purpose is to simulate users' behavior habits of check-in behavior. Then, different types of context information are used to construct a personalized recommendation model. However, the users' check-in data are extremely sparse, which leads to low performance in personalized model training using recurrent neural network (RNN). Therefore, we propose a category-aware gated recurrent unit (CA-GRU) model to mitigate the negative impact of sparse check-in data, capture long-range dependence between user check-ins and get better recommendation results of POI category. We combine the spatiotemporal information of check-in data and take the POI category as users' preference to train the model. Also, we develop an attention-based category-aware gated recurrent unit (ATCA-GRU) model for next POI category recommendation. The ATCA-GRU model can selectively utilize the attention mechanism to pay attention to the relevant historical check-in trajectories in the check-in sequence. We evaluate ATCA-GRU using a real-world dataset, named Foursquare. The experimental results indicate that our ATCA-GRU model outperforms the existing similar methods for next POI recommendation.

KEYWORDS:
next POI recommendation, category-aware, attention, gated recurrent unit, embedding

1 | INTRODUCTION

Nowadays, location-based social network services (LBSNs) are developing rapidly, such as Foursquare, Gowalla and Yelp[1]. Users can use their mobile devices to publish the check-in location when they find their points-of-interest (POI). They can share their location and post their feelings at the time. With the increasing check-in data, LBSNs have accumulated considerable users' check-in records. As a result, it provides precious opportunity to mine users' interests and make POI recommendations to users according to their check-in records. POI recommendation has been widely considered at present. Time is closely related to the check-in behavior of users. However, if only predicts the POI that users may visit in a long time, such as a few months later, has little value in real applications. So the next POI recommendation has become hot research now. Although LBSNs collect massive users' data, the check-in data is still extremely sparse, and it isn't easy to mine users' behavior habits from users' sparse check-ins. While the user's preference for POI category is beneficial in alleviating data sparsity and improving recommendation performance[2]. Therefore, for sparse data, we regard the POI category as users' behavior preference, combine preference with time effect, and focus on POI category prediction.

Our goal is to predict POI category which users most likely to visit next. This research is helpful to advertising, urban traffic planning[3] and other practical applications[4 5 6].

The main purpose of our research is to predict the POI category. By utilizing the users' check-in history and some other contexts, users may visit these POI categories in a short time. Unlike items such as application programming interfaces (API), music and movie recommendation without context, next POI recommendation focuses on connection between users and the real world. Time is an important factor affecting people's check-in behavior. For example, some office workers may go to the gym after work; some students may go to cram school on the rest day; people may go to the restaurant for lunch when they need. Moreover, there are complicated time dependencies between users' different check-ins. A user's former check-in trajectories may have a crucial influence on the next check-in. In a word, the temporal contexts have an important influence on analyzing users' historical behavior habits and recommending POIs to users.

In previous studies, Markov Chain (MC) has been widely used to capture the time dependency between users' check-ins[7]. Recently, deep learning research is very popular in various fields[8]. Among them, Recurrent Neural Network (RNN) and its variant Long Short-Term Memory (LSTM)[9]or Gated Recurrent Unit (GRU) play a crucial role in capturing time dependence. Among them, RNN is difficult in capturing the long-term dependence, but its variants LSTM and GRU can capture the long-term dependence of check-in sequence by improving the architecture of the cell. At present, the related research on the next POI recommendation is based on spatiotemporal context information, mining users' sequential patterns, Other researches combine weather information, users' emotional information and POI category information. However, these methods are greatly affected by the actual situation of dataset. The reason for sparsity may be that users haven't motivation to check-in or for privacy protection[10 11]. If the dataset is very sparse, the accuracy of the prediction results will be reduced. Moreover, privacy protection has been widely considered in many fields[12 13]. Unfortunately, although the current LBSN services have collected large number of users' check-in information, it is still not enough. And the data sparsity still greatly affects the performance of the model. To address data sparsity and achieve high quality, we develop an Attention-based Category-aware Gated Recurrent Unit (ATCA-GRU) model to make the next POI category recommendation. The approaches of our work are summarized below.

(1) In order to make full use of the users' historical check-in sequences on the LBSN and mine the complex dependency between users' check-ins, we use GRU model to reduce the negative impact of data sparsity on recommendation results, we propose a category-aware Gated Recurrent Unit (CA-GRU) model with POI categories as users preferences.

(2) We combine the attention mechanism with GRU and propose an attention-based category-aware GRU (ATCA-GRU) model to predict the POI category that users are most likely to visit in the next 24 hours.

(3) We evaluate the ATCA-GRU model using a real-world LBSN dataset collected from Foursquare. The empirical results show that compared with the baselines, our model significantly improves the POI performance of POI category recommendation.

The rest of the paper is organized as follows. Section II presents the works related to the topics of POI category recommendation. Section III introduces some related professional knowledge of our research. Section IV details our attention-based category-aware GRU model, Section V introduces our experiments and Section VI concludes the paper and discusses our future work directions.

2 | RELATED WORK

Conventional POI recommendation. Collaborative Filtering (CF) and Matrix factorization (MF) are popular in conventional POI prediction. And it mainly considers spatial and temporal information. Ye et al.[14] provided a solution based on collaborative filtering (CF) to predict users may interest POI according to geo-localization information and similar users. Yuan et al.[15] combined temporal effects with user-oriented. They used the Bayes rule to capture the spatial-temporal impact and made POI recommendation. Zhang et al.[16] proposed a method based on matrix factorization (MF) to infer users' interests according to social and geographical influences. Social influence is also an extremely important factor. Zhou et al.[17] made group recommendation considering the social impact. Zhao et al.[18] provided an extreme learning machine-based method. It considered the popularity of POI, users' location and preference to recommend POI to a set of users. And they made POI recommendation by mining the users' evaluation information of POI and POI location. And then inferred the users' preference for POIs. In addition, various contexts are also important for recommendation. Huang et al.[19] proposed a unified probabilistic generative model which associated with social, spatial, temporal and sequential of users' check-in records, so as to mine the patterns of the user check-in records.

Next POI recommendation. It mainly considers the POI that users most recent check-ins at a specific time, so that it can meet the needs of POI recommendation in reality. Therefore, the next POI recommendation is more challenging. Jiao et al.[20] simulated user mobile processes with users' preferences and locations. And it used tensor to model the dynamic user behavior to make POI recommendation. Zhao et al.[7] proposed a personalized Markov chain (MC) based to recommend POI. Besides, RNN is also widely used in many fields, such as next POI recommendation[21] and medical pre diagnosis[22]. Chen et al.[23] proposed a memory enhancement neural network for the recommendation. Huang et al.[24] utilized the improved structure of RNN called Long Short-Term Memory (LSTM). Besides, attention mechanism has been widely used in many fields recently. Whether it is image caption[25], speech recognition[26], statistical learning[27], or natural language processing[28], it is easy to use attention models. It helps to overcome some challenges in the recurrent neural network. For example, as the input length increases, the performance decreases, and the calculation efficiency is low due to the unreasonable input sequence. In[24], an ATST-LSTM model is designed to predict POI. In order to better mine different dependence between users' check-ins, Huang et al. combined the attention mechanism in human vision with LSTM, and proposed the ATST-LSTM model. Pang et al.[29] proposed a hierarchical attention mechanism (HAM-POIRec), which effectively improves the utilization of data and mines more hidden information. Wang et al.[30] proposed a SPENT model that can use an RNN to model users' continuous transition behavior. Kong et al.[31] proposed a spatiotemporal LSTM model (ST-LSTM), which is an improvement on the ST-RNN model. The advantage of this model is that it can better capture the long-time dependence between users' check-in. These methods also achieved good prediction results. However, the sparsity of users' check-in data contribute to low accuracy of model prediction.

POI category recommendation. The POI category plays a vital role in POI recommendation, and it can reflect users' preferences. At present, many POI recommendation methods have used POI category factors to assist in recommendation. He et al.[32] developed a category-based method to make next POI recommendation using the category influence. Zhang et al.[33] used the geographic, social and classification correlations on POI the correlation score between users and unvisited POI. And then provided users with suggestions. In addition, Chen et al.[34] proposed a successive POI category prediction method. They used three-dimensional tensor decomposition to predict the user's next POI category. In the follow-up, they combined successive POI category prediction with the group influence of users, and then carried out continuous POI recommendation. Cheng et al.[35] also decomposed the next POI prediction problem into POI category prediction and specific POI prediction. But the difference is that they used the matrix decomposition method to predict the category, and then combined the user preferences, time influence and geographic influence to recommend the next POI. Although some methods are used to predict the POI category, they do not regard the users' POI category as the users' preferences for fine-grained POI category prediction.

It is difficult to find users' behavior patterns and predict a POI because of the sparse check-in data. Moreover, the low-accuracy POI prediction has low value in real applications. Therefore, efficient use of information is particularly important[36].POI category prediction can not only alleviate the sparsity but also have better practicality value. For example, recommending a certain type of restaurant to users is more diverse than recommending a certain restaurant, and recommending entertainment venues is more selective than recommending a certain entertainment venue.

At present, the related research on POI recommendation is mainly based on spatiotemporal context to mine the users' sequential patterns, and some researches combine weather information, users' emotional information and POI category information[37]. However, these methods are greatly affected by the actual situation of the dataset. Unfortunately, although the current LBSN services have collected large number of user check-in records, it is still not enough. And the data sparsity still greatly affects the performance of the model. In particular, we devise an ATCA-GRU model, which can not only effectively alleviate sparsity, but also accurately forecast the POI categories that users may be interested in. Our study addresses data sparsity and makes accurate and efficient next POI category recommendations for users.

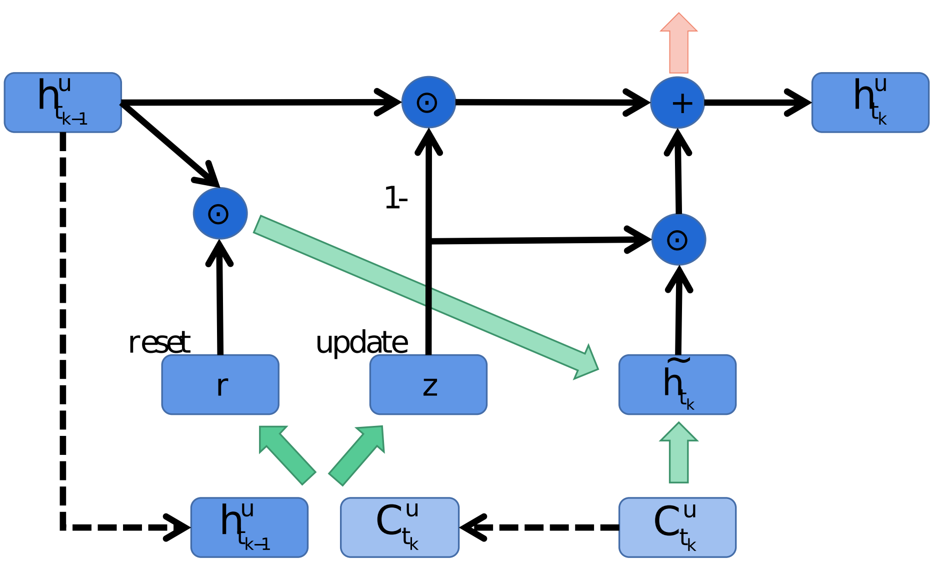


FIGURE 1 The structure of GRU unit.

3 | PRELIMINARIES TO THIS STUD

3.1 | Problem definition

Definition 1. **Cheak-in:** A user's check-in activity is a 4-tuple $Q^u_{t_k}$=(u, v_l, c_p, t_k) where $u \in U$, $v_l \in V$ and $c_p \in C$. It indicates that a user u visits POI v_l from category c_p at time point t_k.

Definition 2. **Cheak-in history:** A check-in history is a set of check-ins of a user. It can be denoted as $CH_u = \{(u, v^u_1, c^u_1, t^u_1), (u, v^u_2, c^u_2, t^u_2), ..., (u, v^u_n, c^u_n, t^u_n)\}$, where $u \in U$, $v_i \in V$ and $c_i \in C$ for $1 \leq i \leq n$. Thus, all users check-ins can be denoted by $ACH = \{CH_{u_1}, CH_{u_2}, ..., CH_{u_{|U|}}\}$, where U is the number of all users.

Definition 3. **POI category recommendation:** Given a user's check-in history CH_u, the user's current time t_k, the goal is to recommend to users the top-K POI categories that users may visit in the short term.

Table 1 shows the notations used in this paper.

TABLE 1 Notations in this paper

Symbol	Description
u, v, c, t	user, POI, POI category, time
$U = \{u_1, u_2, ..., u_M\}$	collection of all users
$V = \{v_1, v_2, ..., v_L\}$	collection of all POIs
$C = \{c_1, c_2, ..., c_P\}$	collection of all POI categories
$c^u_{t_k}$	user u visit POI category c at t_k
$Q^u_{t_k} = (u, v_l, c_p, t_k)$	check-in activity of user u
$C^u_{t_k}$	embedding vector of POI category
$t^u_{t_k}$	temporal feature vector of $c^u_{t_k}$
$h^u_{t_k}$	hidden vector of a GRU unit
$o^u_{t_{N+1},c_k}$	probability of user u
	visits POI category c_k at t_{N+1}
$r^u_{t_k}, z^u_{t_k}$	reset gate vector and
	update gate vector of GRU units
s^u	Context vector of u
α	attention weight vector of ATCA-GRU
σ	sigmoid function
$\{w\}$	set of weight matrices for a GRU model
$\{b\}$	set of bias vectors for a GRU model

3.2 | Gated Recurrent Unit

To make full use of the user's previous check-in trajectories to learn the dependency relationship between check-ins, some researchers choose to use RNN. However, RNN is limited by its structure to capture the long-range dependence between check-ins [38]. Therefore, some researchers choose to use the variant LSTM, which can mine the long-term dependence between check-ins better. LSTM has a complex gate structure, which will lead to long time consumption in the process of model training. As a variant of RNN, GRU has a simpler gate structure and better effect on training results [39]. In this study, we choose to use GRU to capture the long-term dependence between check-in, then mine the user's sequential patterns.

Fig. 1 illustrates the structure of GRU units. The reset gate and update gate control input vector and output hidden vector correspondingly. The calculation of reset gate and update gate as follows:

$$X = \begin{bmatrix} h^u_{t_{k-1}} \\ C^u_{t_k} \end{bmatrix} \tag{1}$$

$$r^u_{t_k} = \sigma\left(W_r \cdot X + b_r\right) \tag{2}$$

$$z^u_{t_k} = \sigma\left(W_z \cdot X + b_z\right) \tag{3}$$

Then, the candidate hidden layer $\tilde{h}_{t_k}$ is calculated:

$$\tilde{h}_{t_k} = \tanh\left(W_{\tilde{h}} \cdot \left[r^u_{t_k} \odot X\right]\right) \tag{4}$$

The reset gate $r^u_{t_k}$ is used to control how much previous memory needs to be retained. For example, if the $r^u_{t_k}$ is 0, then the $\tilde{h}_{t_k}$ only contains the information of the current input. Finally, the update gate $z^u_{t_k}$ controls how much information needs to be forgotten from the previous hidden layer $h^u_{t_{k-1}}$ and how much hidden layer information $\tilde{h}_{t_k}$ needs to be added to the current hidden layer formula. Finally, the final hidden layer information h_{t_k} is obtained directly.

$$h_{t_k} = \left(1 - z^u_{t_k}\right) \odot h_{t_{k-1}} + z^u_{t_k} \odot \tilde{h}_{t_k} \tag{5}$$

Specifically, r, z are the reset gate and update gate respectively. And σ represents the sigmoid function, $\odot$ represents the elements multiplication. Where W_r, W_z, $W_{\tilde{h}}$ are weight matrices and b_r, b_z are bias vectors of GRU units. Then, we can calculate the probability that the category of user u visits POI v_k at time point t_{N+1}, where Ck donates the embedding vector of POI category. And this method will offer top-K POI categories for users.

3.3 | Attention model

Attention model has been used in various fields in recent years[40 41]. It is easy to encounter attention model in different types of tasks. It can also improve the interpretability of neural network[42]. Therefore, reasonable use of attention mechanism is helpful to improve the performance of the model. In connection with the next POI category recommendation, it can be understood that the historical check-in records have different impacts on users' next check-in. Some of the check-in records have greater influence on users' next check-in.

In the attention mechanism, we choose the key-value pair attention mode. Inspired by[43], we connect matrix Q, K and V_{val} respectively and take them as the input of attention function. The core of attention mechanism is the calculation of attention value, and its processing process mainly consists of three parts.

1) Calculate the similarity between key and value: By calculating the similarity between query and key, attention score can be obtained. The most frequently used functions is dot-product attention due to its high efficiency and speed. It can be defined as follows:

$$f_{mul}(Q, K) = QK^T \tag{6}$$

2) Numerical conversion: The softmax function is used to normalize the attention score, which can highlight the weight of important elements.

3) Weighted sum of values: The value is weighted and summed according to the weight coefficient. The formula of the integrated process definition is expressed as follows:

$$Attention(Q, K, V_{val}) = softmax(f(Q, K))V_{val} \tag{7}$$

4 | SOLUTION

In this section, we present the ATCA-GRU model to address the next POI recommendation problem in the next 24 hours.

4.1 | Category-aware GRU model

We use GRU to capture non-linear dependence between users' check-in, and GRU is good at dealing with time series. Considering that the user's check-in records are very sparse, it is challenging to use the extremely sparse POI check-in data to train the model. Inspired by[44], we take the POI category as the user's preference and take the check-in time and POI category as the input of the model. Considering that we are trying to recommend the most likely POI categories that users are expected to visit in the next 24 hours, we have made some settings in the process of model training. If the user's check-in interval exceeds ten times, they will belong to different check-in tracks. Fig. 2 illustrates the architecture of CA-GRU.

We define a binary ($C^u_{t_i}$, t^u_i) as the input at each time point. And we can update each hidden vector $h^u_{t_k}$ by using current input and the previous hidden vector $h^u_{t_{i-1}}$. The definition is as follows:

$$h^u_{t_k} = GRU\left(W_C C^u_{t_i} + W_t t^u_i, h^u_{t_i-1}\right) \tag{8}$$

where $W_C \in R^{d\times d}$ and $W_t \in R^{d\times d}$ are transition matrices. Finally, we will use the output of GRU to get the category probabilities of the classification, these probabilities are the probability of the next check-in category that the user may visit. The category with a higher probability is selected as the next check-in category that the user may visit. The probability $o^u_{t_{N+1,C_k}}$ can be obtained by the following formula.

$$o^u_{t_{N+1,C_k}} = \left(W_N h^u_{t_N} + W_P \mathrm{P}_u\right)^T \left(W_C C^u_{t_i} + W_t t^u_i\right) \tag{9}$$

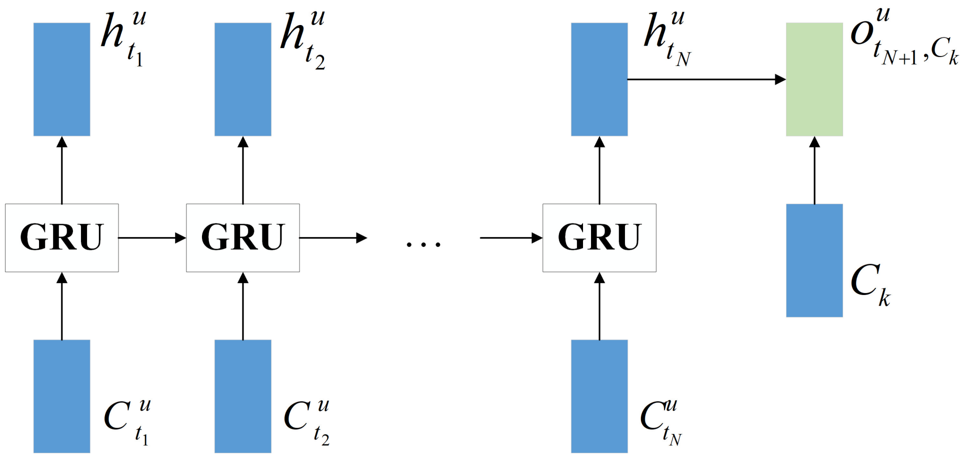


FIGURE 2 The architecture of CA-GRU.

4.2 | Attention-based category-aware GRU model

Not all check-in records have the same impact on the next check-in. (e.g., if a user's check-in records are school, shopping mall or cinema, it is obvious that the impact of a shopping mall on cinema is far greater than that of a school on cinema). However, a standard GRU cannot judge which historical check-in is critical to the next POI recommendation. This is an area that needs to be improved. Therefore, we propose an attention-based category-aware GRU (ATCA-GRU) model to address the above issue. Using attention mechanism, we can focus on the different history check-ins of users, assign different weights for historical check-ins, and then we can mine the users' preferences.

The ATCA-GRU model consists of two key steps. First, we design a category-aware GRU model. To overcome the negative impact of data sparsity on model training, we take POI category as users' preference and predict the most likely POI category to be accessed by users. In addition, we add attention mechanism in CA-GRU to better mine users' preferences inspired by NLP, called ATCA-GRU. Finally, the top-K POI categories in the remaining POI categories are recommended to users. Fig. 3 illustrates the architecture of ATCA-GRU, where N indicates the length of check-in sequence. Then, ATCA-GRU generates an attention weight vector α and produces a weighted hidden representation r. The process is as follows:

$$\mathrm{r}_{t_N}^u = \sum_{i=1}^{N} a_i h_{t_i}^u \tag{10}$$

Finally, we can get the probability that user will visit POI category C_k at time t_{N+1}. And calculation process is as follows:

$$o_{t_{N+1},c_k}^u = \left(W_N r_{t_N}^u + W_P \mathrm{P}_u\right)^T \left(W_C C_{t_i}^u + W_t t_i^u\right) \tag{11}$$

where the definitions of W_N and W_p refer to (9). In summary, we have proposed a model that can be used to predict the category of POI, called the ATCA-GRU model. We use the gating unit in GRU to better capture the long-term and short-term dependence between users' check-ins. Then, in order to pay more attention to the historical records that have a greater impact on the next check-in, we combine attention mechanism with LSTM. It is helpful to improve the interpretability of neural network model. Finally, in order to ensure the timeliness of the predicted POI categories, we set a sliding window with the size of 10. It can focus on the 10 users' check-in records that are close to the users' next check-in. Finally, we use the ATCA-GRU model to perform large-scale POI category prediction.

Our pseudocode is shown in Algorithm 1. We first construct the data needed for forwarding propagation training (see lines 1-7), and then train our ATCA-GRU model by using backpropagation, in which we use cross-entropy loss function (see lines 9-14).

5 | EXPERIMENTS

In this section, we evaluate the performance of the ATCA-GRU model and compare it with the advanced baselines on dataset from Foursquare. Through comparison, it can be seen from the comparison that the ATCA-GRU model is superior to all baselines in predicting POI category.

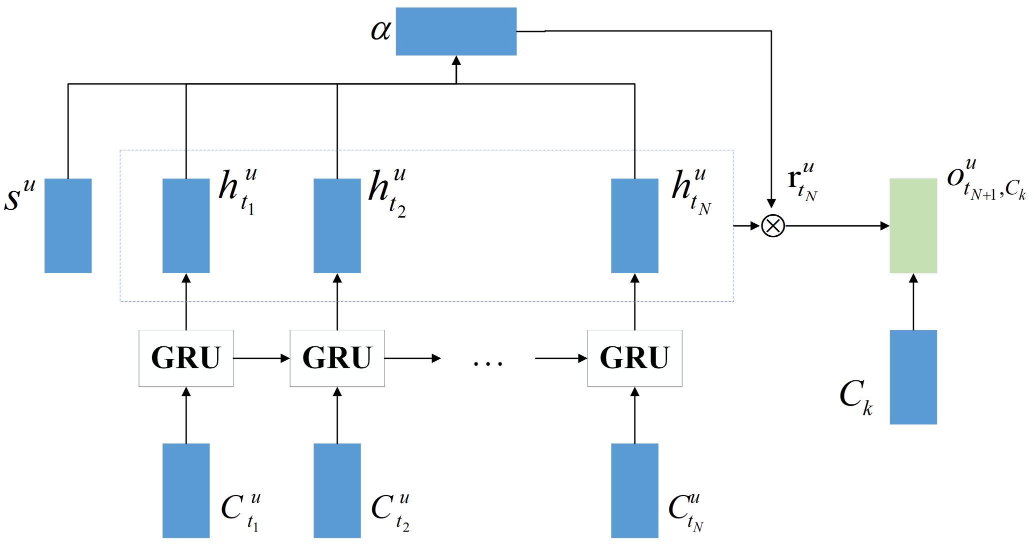


FIGURE 3 The architecture of ATCA-GRU.

Algorithm 1 ATCA-GRU

Require: users *U* and their previous check-in records *ACH*

Ensure: ATCA-GRU model $\{M\}_u$

//building training examples

$D=U_u D^u=\phi$;

for each user *u* in *U* **do**

  for each check in trajectory $Q^u_{t_k}$ in CH_U **do**

    Compute the embedding vector $C^u_{t_k}$ of category $c^u_{t_k}$

  end for

  Add a training instance$\{(C^u_{t_k}), (c'^u_{t_k})\}$ to D^u

end for

//model training

Initialize the parameter set Θ

while exceed (maximum number if iterations) == false **do**

  for each user *u* in *U* **do**

    Randomly select a batch of instances D^u_b from D^u

    Find Θ minimizing the objective with D^u_b

  end for

end while

return $\{M\}_u$

5.1 | Data collection and preprocessing

Location privacy protection is an important challenge in location-based social networks [45 46]. We have also considered this aspect. Our fine-grained POI category prediction method only focuses on the specific category of users' check-in, not the specific location of users' check-in. Therefore, in the training process of the model, we did not use the specific location information of the users' check-in. So the users' location privacy would not be exposed. We evaluate our model on a real-world Foursquare dataset collected from New York City (NYC) [47] from 12 April 2012 to 16 February 2013. This dataset is frequently-used in POI prediction field, which includes user ID, POI ID of user check-in, POI category, check-in location (longitude and latitude) and time of check-in. In the process of using the dataset, we first preprocess the dataset and delete the POIs with less than 5 users visits, because they are not recommended to users and hinder the training of the model. Then, we divide the dataset. we use the top 80 % of the dataset as the training set and the last 20 % as the test set. An overview of the dataset we used is shown in Table 2.

TABLE 2 Basic statistic of foursquare dataset

	Users	Check-in	Categories
NYC	1083	179,468	233

5.2 | Baseline methods

We select the following POI category prediction methods to compare with our proposed methods.

1. ATCA-LSTM: Using our idea but using the LSTM model to predict the POI category.
2. Frequency: The percentage of users' most frequent visits is taken as the probability of users visiting the POI category.

5.3 | Evaluation metrics

In the evolution, we use the following evaluation metrics: precision @K, recall @K and F1-score @K. Precision, recall and F1-score are several commonly used indicators to measure the performance of models. Besides, F1-score is the harmonic mean of precision and recall.

In this study, the three metrics are defined as follows:

$$P@k = \frac{1}{n}\sum_{u=1}^{n} P_u@k = \frac{1}{n}\sum_{u=1}^{n}\frac{|S_u(k)\cap C_u|}{k} \tag{12}$$

$$R@k = \frac{1}{n}\sum_{u=1}^{n} R_u@k = \frac{1}{n}\sum_{u=1}^{n}\frac{|S_u(k)\cap C_u|}{|c_u|} \tag{13}$$

$$F_1@k = \frac{1}{n}\sum_{u=1}^{n} F_{1_u}@k = \frac{1}{n}\sum_{u=1}^{n}\frac{2\cdot P@k\cdot R@k}{P@k+R@k} \tag{14}$$

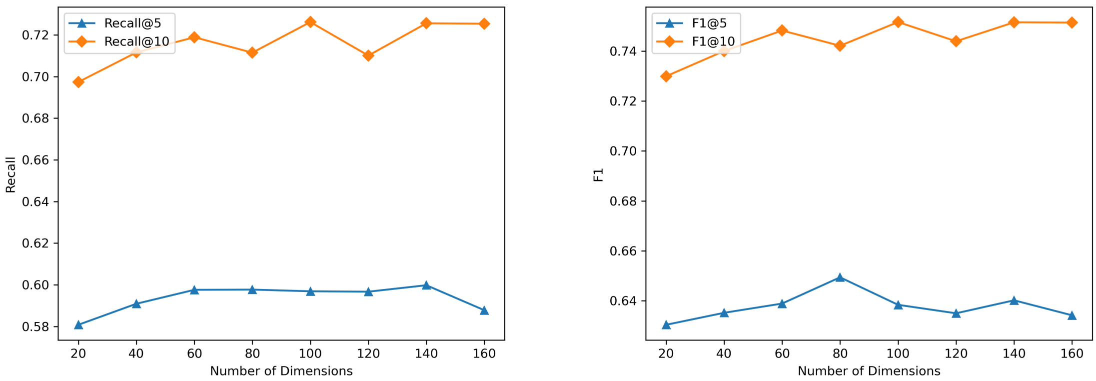
where $S_u(k)$ denotes the set of top-k POI categories will be recommended to user u and C_u denotes POI categories that users will visit in real world. For most users, it is not enough for them to predict the category with the highest probability. People tend to have the right to choose. Therefore, in order to provide users with selectivity and diversity, we take k = 5, k = 10 and k = 15 as examples to predict users' interest categories.

5.4 | Effect of parameters

The parameter d is responsible for determining the embedding dimension of POI category, and it impacts the performance of ATCA-GRU. To find the best performance of the model, we conduct several experiments. We use Rec@5 and F1@5 to measure the performance. Fig. 4 shows that the performance of our model rises steadily before d=60 and then fluctuates and the performance does not change much. And we set the embedding dimension to 60 in the experiments.

5.5 | Results on POI category recommendation

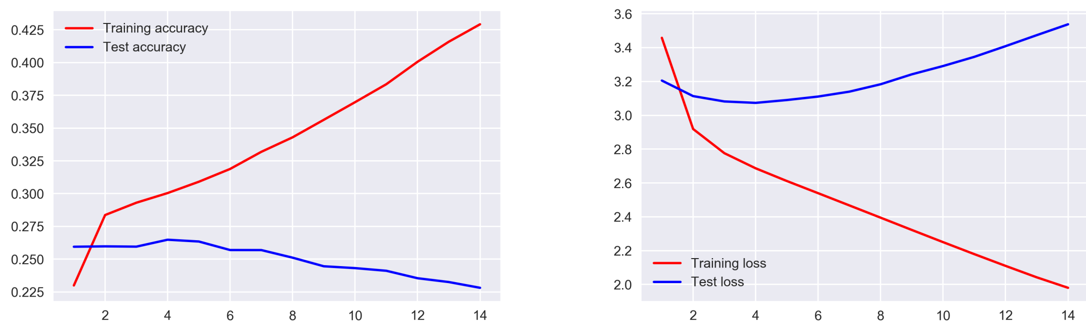
After completing the ATCA-GRU model, it is necessary to evaluate the effect of the model. At present, the evaluation indicators that are often used include precision, recall, F1-score, etc. The research has selected recall and F1-score to be visualized. The training process will be briefly described below. From Fig. 5(a), we can see that during the initial training process, the values of recall@5 and recall@10 are gradually increasing as the embedding dimension increases. It shows that the model's performance is gradually getting better. But when the embedding dimension=80, the value of recall@10 begins to decrease, and the value of recall@5 begins to be stabilized. It shows that if the embedding dimension is increased at this time, the model is difficult to achieve better performance. Increasing embedding dimension will increase the time cost of model training. In Fig. 5(b), the value of F1-score increases when the embedding dimension is less than 60. When the embedding dimension exceeds 60, the value of F1-score begins to float up and down, which indicates that it is reasonable for our model to select the embedding dimension of 60. Therefore, under the premise of ensuring that the model performance, the smaller of the embedding dimension, the better. We finally decided to choose the embedding dimension as 60. Since we predict the POI categories that users may visit in the next 24 hours, we need to set a sliding window to make GRU pay more attention to the user's check-in categories shortly soon. We finally determined that the size of the sliding window is 10 by analyzing the data. In other words, we use the POI categories of the previous 10 visits to mine users' short-term preferences. The reason why we don't select the specific duration as the sliding window standard is that users' check-in is random. It is difficult to have an appropriate time to measure all users'



(a) Recall@k

(b) F1-Score@k

FIGURE 4 Performance corresponding to different embedding dimensions.



(a) Accuracy of training process

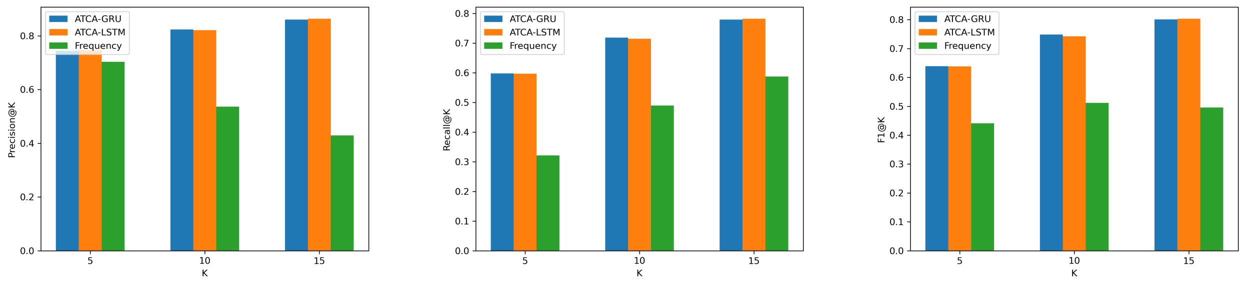
(b) Loss of training process

FIGURE 5 Performance corresponding to different training process.

check-in. Some users' check-in frequently and some users' check-in sparsely. We transform the time in dataset into sequential timestamps after unified preprocessing. And we finally chose the size of the sliding window to be 10.

As shown in Fig. 5(a), the accuracy of training set is increasing from the first iteration to the second iteration, which indicates that the iterative training effect is significant. But the accuracy of the verification set will not increase with the iteration number increasing again. When the iteration number reaches 5, the accuracy of the test set begins to decline. In Fig. 5(b), the loss value of model training also changes accordingly, because the loss value is closely related to the accuracy. The loss value of the training set decreases rapidly, which reflects effect of model training. But when the iteration times are 5, the loss value begins to rise. This is due to the overfitting phenomenon caused by small dataset and insufficient data. At the beginning of model training, the loss value of the training set is closely related to the accuracy, we set the initial iteration number to 50 times. In order to prevent the model from wasting too much time on overfitting, we set the condition of model stopping ahead of time: when the model accuracy is no longer rising twice in a row, the model training ends in advance.

When epoch=14, the model reaches the early stopping condition, No more training. The comparison chart of the experimental results is shown in the Fig. 6. It can be seen from Fig. 6 that the ATCA-GRU model we proposed is better than other baseline methods in three aspects: precision, recall and F1-score, but ATCA-GRU is consistent with the ATCA-LSTM model in terms of training ideas, so the efficiency of the GRU model is only a little higher than that of the LSTM. But our model is much more effective than the frequency-based method. Fig. 6(a) shows the accuracy and comparison of our ATCA-GRU model. When the recommended category is the top 5, the accuracy of the model is 0.7441, which is 5.42% higher than the frequency method and 0.21% lower than the ATCA-LSTM method. When the recommended category is the top 10, the accuracy of the model is 0.8236, which is 28.36% higher than the frequency method and 0.19% higher than the ATCA-LSTM method. When the recommended category is



(a) Precision@K (b) Recall@K (c) F1-Score@K

FIGURE 6 Performance compared with different training process.

the top 15, the accuracy of our model is 0.8606, which is 43.4% higher than the frequency method and 0.36% lower than the ATCA-LSTM method. Fig. 6(b) shows the recall and comparison of our ATCA-GRU model. When the recommended category is the top 5, the recall of the model is 0.5976, which is 25.76% higher than the frequency method and 0.8% higher than the ATCA-LSTM method. When the recommended category is the top 10, the recall of the ATCA-GRU model is 0.7189, which is 23.67% higher than the frequency method and 0.41% higher than the ATCA-LSTM method. When the recommended category is the top 15, our model recall is 0.7794, which is 18.95% higher than the frequency method and 0.33% lower than the ATCA-LSTM method. Fig. 6(c) shows the F1-score and comparison of our ATCA-GRU model. When the recommended category is the top 5, the F1-score of the ATCA-GRU model is 0.6389, which is 16.78% higher than the frequency method and 0.11% higher than the ATCA-LSTM method. When the recommended category is the top 10, the F1-score of our model is 0.7482, which is 20.98% higher than the frequency method and 0.56% higher than the ATCA-LSTM method. When the recommended category is the top 15, the F1-score of our model is 0.8008, which is 29.81% higher than the frequency method and 0.28% lower than the ATCA-LSTM method.

From these experimental results, it can be seen that the recommendation effect of our ATCA-GRU model far exceeds the frequency-based method, and the quality of the recommendation results has been greatly improved. However, the difference between ATCA-LSTM and ATCA-GRU model is very small. Sometimes ATCA-LSTM can even surpass the ATCA-GRU model occasionally. This is because ATCA-LSTM is consistent with our ideas, so the recommended results are similar, but the LSTM unit has three gated structures (forget gate, input gate and output gate), which will reduce the training speed of the model. However, there are only two gated structures (update gate and reset gate) in the GRU. In the case of similar recommendation results, faster GRU is more suitable to complete our POI category recommendation task.

6 | CONCLUTION

Recently, the next POI recommendation has attached considerable attention. However, it is hard to discover users' interests from extremely sparse check-in data. Therefore, we take the POI category as the user's preference to carry out the category aware POI recommendation. We propose an attention-based category-aware GRU model, which can not only alleviate the sparsity of users' check-in, but also capture the short-term and long-term dependence between user check-in. Finally, we will predict the probability of the user's next check-in category, and recommend top-K categories according to the probability. We deployed the experiment on a real data set and compared our method with the baseline methods. Experiments show that our ATCA-GRU is obviously better than other baseline methods.

For future work, we will provide users with specific POI recommendations based on category prediction. Time is still an important factor affecting the user's check-in. We hope to find a better way to capture the time pattern and find a more appropriate time window to analyze the user's behavior pattern and predict the user's preference. Although GRU can capture the user's historical check-in information well, there is still a room for improvement. While improving the accuracy, we should also consider dealing with the long tail effect and providing more diverse recommendation lists to better meet the needs of users.

7 | ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (No. 61872219), the Natural Science Foundation of Shandong Province (ZR2019MF001), the Open Project of State Key Laboratory for Novel Software Technology (No. KFKT2020B08), the Macao Science and

Technology Development Fund under Macao Funding Scheme for Key R & D Projects (0025/2019/AKP) and the Fundamental Research Funds for the Central Universities under Grant (No. 30919011282).

References

1. Huang H, Gartner G, Krisp JM, Raubal M, Weghe V. dN. Location based services: ongoing evolution and research agenda. *Journal of Location Based Services* 2018; 12(2): 63–93.

2. Liu X, Liu Y, Aberer K, Miao C. Personalized point-of-interest recommendation by mining users' preference transition. *In: ACM international conference on Information* 2013: 733–738.

3. Cai Z, Zheng X, Yu J. A differential-private framework for urban traffic flows estimation via taxi companies. *IEEE Transactions on Industrial Informatics* 2019; 15(12): 6492–6499.

4. Liang Y, Cai Z, Yu J, Han Q, Li Y. Deep learning based inference of private information using embedded sensors in smart devices. *IEEE Network* 2018; 32(4): 8–14.

5. Zheng X, Cai Z, Li J, Gao H. Location-privacy-aware review publication mechanism for local business service systems. *In: IEEE Conference on Computer Communications* 2017: 1–9.

6. Sun Z, Wang Y, Cai Z, Liu T, Tong X, Jiang N. Privacy Protection based on Stream Cipher for Spatio-temporal Data in IoT. *IEEE Internet of Things Journal* 2021. doi: 10.1002/int.22371

7. Zhao S, Zhao T, Yang H, Lyu M, King I. STELLAR: Spatial-temporal latent ranking for successive point-of-interest recommendation. *In: AAAI Conference on Artificial Intelligence* 2016; 30(1).

8. Zhou X, Liang W, Kevin I, Wang K, Wang H, Yang . Deep Learning Enhanced Human Activity Recognition for Internet of Healthcare Things. *IEEE Internet of Things Journal* 2020; 7(7): 6429-6438.

9. Zhou X, Hu Y, Liang W, Ma J, Jin Q. Variational LSTM enhanced anomaly detection for industrial big data. *IEEE Transactions on Industrial Informatics* 2020. doi: 10.1109/TII.2020.3022432

10. Cai Z, He Z. Trading private range counting over big IoT data. *In: International Conference on Distributed Computing Systems* 2019: 144–153.

11. Wang Y, Gao Y, Li Y, Tong X. A worker-selection incentive mechanism for optimizing platform-centric mobile crowdsourcing systems. *Computer Networks* 2020; 171: 107144.

12. Cai Z, He Z, Guan X, Li Y. Collective data-sanitization for preventing sensitive information inference attacks in social networks. *IEEE Transactions on Dependable and Secure Computing* 2016; 15(4): 577–590.

13. Liu T, Wang Y, Li Y, Tong X, Qi L, Jiang N. A Two-stage Privacy Protection Mechanism Based on Blockchain in Mobile Crowdsourcing. *IEEE Internet of Things Journal*; 7(9): 7928-7940.

14. Ye M, Yin P, Lee WC, Lee DL. Exploiting geographical influence for collaborative point-of-interest recommendation. *In: ACM SIGIR conference on Research and development in Information Retrieval* 2011: 325–334.

15. Yuan Q, Cong G, Ma Z, Sun A, Thalmann NM. Time-aware point-of-interest recommendation. *In: ACM SIGIR conference on Research and development in information retrieval* 2013: 363–372.

16. Zhang Z, Liu Y, Zhang Z, Shen B. Fused matrix factorization with multi-tag, social and geographical influences for POI recommendation. *World Wide Web* 2019; 22(3): 1135–1150.

17. Zhou X, Liang W, Huang S, Fu M. Social recommendation with large-scale group decision-making for cyber-enabled online service. *IEEE Transactions on Computational Social Systems* 2019; 6(5): 1073–1082.

18. Zhao X, Zhang Z, Bi X, Sun Y. A new point-of-interest group recommendation method in location-based social networks. *Neural Computing & Applications* 2020. doi: 10.1007/s00521-020-04979-4

19. Zhang DY, Wang D, Zheng H, Mu X, Li Q, Zhang Y. Large-scale point-of-interest category prediction using natural language processing models. *In: IEEE International Conference on Big Data* 2017: 1027–1032.

20. Jiao X, Xiao Y, Zheng W, Wang H, Hsu CH. A novel next new point-of-interest recommendation system based on simulated user travel decision-making process. *Future Generation Computer Systems* 2019; 100: 982–993.

21. Chen H, Wang S, Jiang N, Li Z, Yan N, Shi L. Trust-aware generative adversarial network with recurrent neural network for recommender systems. *International Journal of Intelligent Systems* 2020. doi: 10.1002/int.22320

22. Zhou X, Li Y, Liang W. CNN-RNN Based Intelligent Recommendation for Online Medical Pre-Diagnosis Support. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 2020. doi: 10.1109/TCBB.2020.2994780

23. Chen X, Xu H, Zhang Y, Tang J, Cao Y, Qin Z. Sequential recommendation with user memory networks. *In: ACM international conference on web search and data mining* 2018: 108–116.

24. Huang L, Ma Y, Wang S, Liu Y. An attention-based spatiotemporal lstm network for next poi recommendation. *IEEE Transactions on Services Computing* 2019. doi: 10.1109/TSC.2019.2918310

25. Li Y, Zeng J, Shan S, Chen X. Occlusion aware facial expression recognition using cnn with attention mechanism. *IEEE Transactions on Image Processing* 2018; 28(5): 2439–2450.

26. Chorowski JK, Bahdanau D, Serdyuk D, Cho K, Bengio Y. Attention-based models for speech recognition. *Advances in neural information processing systems* 2015; 28: 577–585.

27. Kirkham NZ, Slemmer JA, Johnson SP. Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition* 2002; 83(2): B35–B42.

28. Hu D. An introductory survey on attention mechanisms in NLP problems. *In: Proceedings of SAI Intelligent Systems Conference* 2019: 432–448.

29. Pang G, Wang X, Hao F, Wang L, Wang X. Efficient point-of-interest recommendation with hierarchical attention mechanism. *Applied Soft Computing* 2020; 96: 106536.

30. Wang MF, Lu YS, Huang JL. SPENT: A successive POI recommendation method using similarity-based POI embedding and recurrent neural network with temporal influence. *In: IEEE International Conference on Big Data and Smart Computing* 2019. doi: 10.1109/BIGCOMP.2019.8679431

31. Kong D, Wu F. HST-LSTM: A Hierarchical Spatial-Temporal Long-Short Term Memory Network for Location Prediction. *In: International Joint Conference on Artificial Intelligence* 2018; 18(7): 2341–2347.

32. He J, Li X, Liao L. Category-aware Next Point-of-Interest Recommendation via Listwise Bayesian Personalized Ranking. *In: International Joint Conference on Artificial Intelligence* 2017; 17: 1837–1843.

33. Zhang JD, Chow CY. GeoSoCa: Exploiting geographical, social and categorical correlations for point-of-interest recommendations. *In: ACM SIGIR conference on research and development in information retrieval* 2015: 443–452.

34. Chen J, Li X, Cheung WK, Li K. Effective successive POI recommendation inferred with individual behavior and group preference. *Neurocomputing* 2016; 210: 174–184.

35. Lin IC, Lu YS, Shih WY, Huang JL. Successive POI Recommendation with Category Transition and Temporal Influence. *In: Annual Computer Software and Applications Conference* 2018: 57–62.

36. Cai Z, Zheng X. A private and efficient mechanism for data uploading in smart cyber-physical systems. *IEEE Transactions on Network Science and Engineering* 2020; 7(2): 766–775.

37. Zhong W, Yin X, Zhang X, Li S, Dou W, Qi L. Multi-dimensional quality-driven service recommendation with privacy-preservation in mobile edge environment. *Computer Communications* 2020; 157: 116-123.

38. Jozefowicz R, Zaremba W, Sutskever I. An empirical exploration of recurrent network architectures. *In: International conference on machine learning* 2015: 2342–2350.

39. Zhang Z, Robinson D, Tepper J. Detecting hate speech on twitter using a convolution-gru based deep neural network. *In: European semantic web conference* 2018: 745–760.

40. Wang Q, Hao Y. ALSTM: An Attention-Based Long Short-Term Memory framework for Knowledge Base Reasoning. *Neurocomputing* 2020; 399: 342–351.

41. Duan Z, Li W, Zheng X, Cai Z. Mutual-preference driven truthful auction mechanism in mobile crowdsensing. *In: International Conference on Distributed Computing Systems* 2019: 1233–1242.

42. Sardianos C, Varlamis I, Chronis C, Dimitrakopoulos G, Alsalemi A, Himeur Y. The emergence of explainability of intelligent systems: Delivering explainable and personalized recommendations for energy efficiency. *International Journal of Intelligent Systems* 2020. doi: 10.1002/int.22314

43. Zhao X, Chen Y, Guo J, Zhao D. A spatial-temporal attention model for human trajectory prediction. *IEEE/CAA Journal of Automatica Sinica* 2020; 7(4): 965–974.

44. Likhyani A, Padmanabhan D, Bedathur S, Mehta S. Inferring and exploiting categories for next location prediction. *In: International Conference on World Wide Web* 2015: 65–66.

45. Qi L, Hu C, Zhang X, Khosravi MR, Sharma S, Pang S. Privacy-aware data fusion and prediction with spatial-temporal context for smart city industrial environment. *IEEE Transactions on Industrial Informatics* 2020. doi: 10.1109/TII.2020.3012157

46. Li J, Ye H, Li T, Wang W, Lou T, Liu J. Efficient and Secure Outsourcing of Differentially Private Data Publishing with Multiple Evaluators. *IEEE Transactions on Dependable and Secure Computing* 2020. doi: 10.1109/TDSC.2020.3015886

47. Yu F, Cui L, Guo W, Lu X, Li Q, Lu H. A Category-Aware Deep Model for Successive POI Recommendation on Sparse Check-in Data. *In: The Web Conference* 2020: 1264–1274.