

Article

An Attention-Based Spatiotemporal Gated Recurrent Unit Network for Point-of-Interest Recommendation

Chunyang Liu ¹, Jiping Liu ^{2,*}, Jian Wang ³, Shenghua Xu ², Houzeng Han ³ and Yang Chen ⁴

¹ School of Environment Science and Spatial Informatics, China University of Mining and Technology, Xuzhou 221116, China

² Chinese Academy of Surveying and Mapping, Beijing 100830, China

³ School of Geomatics and Urban Spatial Informatics, Beijing University of Civil Engineering and Architecture, Beijing 100044, China

⁴ State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China

* Correspondence: liujp@casm.ac.cn; Tel.: +86-138-0105-1915

Received: 4 July 2019; Accepted: 7 August 2019; Published: 13 August 2019



Abstract: Point-of-interest (POI) recommendation is one of the fundamental tasks for location-based social networks (LBSNs). Some existing methods are mostly based on collaborative filtering (CF), Markov chain (MC) and recurrent neural network (RNN). However, it is difficult to capture dynamic user's preferences using CF based methods. MC based methods suffer from strong independence assumptions. RNN based methods are still in the early stage of incorporating spatiotemporal context information, and the user's main behavioral intention in the current sequence is not emphasized. To solve these problems, we proposed an attention-based spatiotemporal gated recurrent unit (ATST-GRU) network model for POI recommendation in this paper. We first designed a novel variant of GRU, which acquired the user's sequential preference and spatiotemporal preference by feeding the continuous geographical distance and time interval information into the GRU network in each time step. Then, we integrated an attention model into our network, which is a personalized process and can capture the user's main behavioral intention in the user's check-in history. Moreover, we conducted an extensive performance evaluation on two real-world datasets: Foursquare and Gowalla. The experimental results demonstrated that the proposed ATST-GRU network outperforms the existing state-of-the-art POI recommendation methods significantly regarding two commonly-used evaluation metrics.

Keywords: point-of-interest recommendation; spatiotemporal context; recurrent neural networks; gated recurrent unit; attention model

1. Introduction

With the prevalence of smart devices and location-based social networks (LBSNs) services, people can easily share their locations and check-in information with others in LBSNs [1,2]. The huge volume of users' history check-in data brings opportunities for researching human mobility behavior, and point-of-interest (POI) recommendation become one of the important tasks in LBSNs. As shown in Figure 1, POI recommendation may help a user find a place of interest after dinner, or provide a user with discount information about nearby shopping malls. Therefore, POI recommendation can not only meet user's personalized preferences for visiting new places but also help LBSNs service providers implement intelligent location-aware online advertising services [3,4].

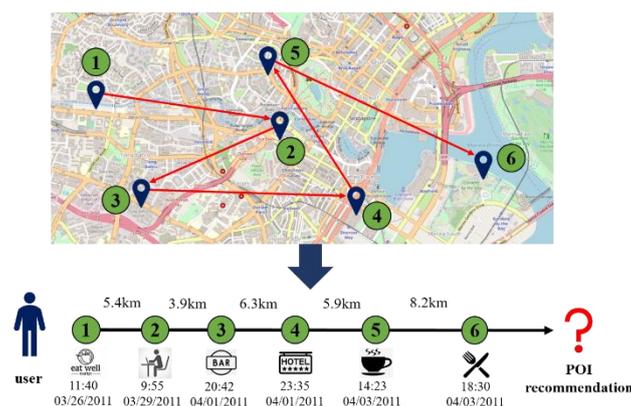


Figure 1. Diagram of a user's check-in sequence.

The task of POI recommendations is investigated in different settings, such as general POI recommendation [5,6], out-of-town recommendation [7], next POI recommendation [3,8,9], tour recommendation for groups [10,11], and so on. In this work, we focus on next POI recommendation by modeling check-in sequences and incorporating spatiotemporal influence in a personalized way, this task is more significant since it can predict users' next movement behaviors. As illustrated in Figure 1, a user's historical check-in sequence contains geographical information and temporal information, similarly, given all users' check-in sequences, the task of POI recommendation is to recommend the next POI that the user is likely interested in visiting. Specifically, according to prediction scores, we can recommend top- k POIs to a user, and a higher predicted score indicates that the user is more likely to go.

Nowadays, POI recommendation has been extensively studied from the academic and industrial fields [5–13]. POI recommendation is different from the general recommendation tasks (e.g., goods, movies, news recommendation) because there is highly spatiotemporal dependence. The spatial and temporal contextual information is precisely the basis for modeling user movement behaviors [8]. Tobler's first law of geography [14] states that "everything is related to everything else, but near things are more related than distant things". For example, people visit nearby places more often in real life, such as cinemas, restaurants, and so on. That is, adjacent POIs are more geographically relevant than distant POIs. Geographic factors significantly impact the user's movement behavior, and geographic location information can effectively improve the recommendation quality [12].

Due to the time sensitivity of the POI recommendation, the temporal information also plays a critical role [2], because physical constraints on check-in activities can lead to specific patterns. For instance, some people always like to go to a cinema on weekend nights to see a movie. The temporal influences of a POI recommender system typically show in the three aspects: periodicity, non-uniformness, and consecutiveness [15]. Besides, sequential influence is also essential in POI recommendation since individual behaviors of movement commonly exhibit sequential patterns [16].

In brief, the influences of spatial, temporal and sequential factors are crucial to analyze individual behaviors for personalized POI recommendation. So far, many researches have considered these factors to improve the performance of POI recommendation algorithms, such as the collaborative filtering (CF) [4] and Markov chain (MC) [3]. However, it is difficult to process sequence data and to capture dynamic user's preferences using CF based methods; and MC based methods depend on strong independent assumptions among different factors. Therefore, there are still many challenges regarding how to integrate information of various features to accurately model users' complex behavioral preferences and to recommend reliable POIs to users.

Recently, recurrent neural network (RNN) [17] and its variants (e.g., long-short term memory (LSTM) [18] and gated recurrent unit (GRU) [19,20]) have been successfully applied to sequential recommender systems [21]. The hidden states of RNN methods have both characteristics in nature and are adequate for modeling sequential correlations and temporal dynamics in POI recommender systems [8,22], and they can better capture the long-term dependency. However, the existing RNN

based POI recommendation methods have difficulty in alleviating the cold-start problem. An excellent choice is to apply RNN and incorporate additional spatiotemporal contextual information, such as continuous geographical distance and time interval. Most RNN methods rely on the last hidden layer activation vector when calculating the output of the network, and this limits the ability to understand and learn the main intention of user check-in behavior from the hidden states [21]. In other words, the historical check-in behaviors of a user are not equally important for predicting the next behavior, and we need to focus on the main information.

In view of the above analysis, we propose a novel attention-based spatiotemporal GRU network (ATST-GRU) for POI recommendation. Figure 2 delineates the architecture of the network. First, considering the GRU network is a more robust variant of RNN which work better in capturing long term dependencies and alleviating the exploding or vanishing gradients problems [23], we attempted to use an extended GRU network to model check-in sequences by considering geographical distances between continuous POIs and time intervals between continuous check-in behaviors. Such a network structure can work better to explore the spatiotemporal influence and sequential influence and alleviate the problems of data heterogeneity and sparsity. Then, inspired by the attention mechanism in neural networks [24,25], we further improved our method by introducing an attention model, which can explore the most pertinent piece of user's check-in behavior. Next, the parameters of ATST-GRU were learned by the Bayesian personalized ranking (BPR) [26] framework and back propagation through time (BPTT) algorithm [27]. Finally, extensive experiments were conducted on two public datasets (Foursquare and Gowalla) and the results were compared with several state-of-the-art POI recommendation methods to evaluate the model. The main contributions of this work are as follows:

- A novel spatiotemporal gated recurrent unit (ST-GRU) network model is proposed in this paper, which combines continuous values of spatial and temporal contexts information into GRU network naturally to capture the user's spatiotemporal preferences and alleviate the problems of data heterogeneity and sparsity;
- An attention-based method is introduced to ST-GRU network for POI recommendation, named ATST-GRU model. This can automatically pay more attention to critical information and extract the user's main purpose, which significantly strengthens the user's long-term interest;
- Extensive experiments on two real-world datasets show that ATST-GRU is effective and outperforms state-of-the-art methods significantly.

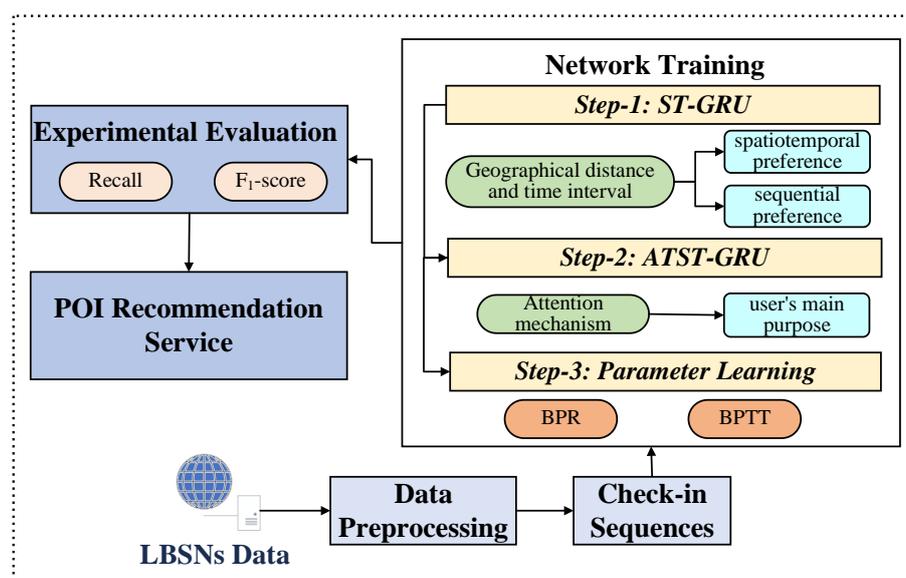


Figure 2. Architecture of the proposed method for point-of-interest (POI) recommendation.

The rest of this paper is organized as follows: the related methods are briefly reviewed in Section 2. The details of our ATST-GRU network are described in Section 3. Experiments and results of the proposed method are illustrated in Section 4. Finally, conclusions are summarized in Section 5.

2. Related Work

In this section, we review several methods for POI recommendation, including CF based methods, MC based methods and deep learning (DL) based methods.

2.1. Collaborative Filtering Based Methods

Personalized POI recommendation has been extensively studied in the related research field. CF is one of the widely used technique [28]. Some previous POI recommendation methods are user-based CF or item-based CF, which take advantage of check-ins of similar users or POIs [2,3,5,29]. A cutting-edge CF method is matrix factorization (MF) [4,6,12,30], which mining potential location preferences of a user by factorizing the observed user-POI matrix. To the best of our knowledge, many POI recommender systems integrate the geographical information [4–6,12,31], temporal information [2,32–35], social information [1,4,36] or other POI characteristic information (reviews, categories, labels, etc.) [37–40] into traditional recommendation algorithms.

Geographical influence is one of the important factors for POI recommendation. Most researchers have modeled the geographical influences by considering distance as a penalty [41] or building a distance distribution model, such as power-law distribution [2,42], multi-center Gaussian distribution [4] or personalized nonparametric distribution [43]. Ye et al. [5] proposed a user-based CF framework for POI recommendation which models geographical influence by power law distribution. Cheng et al. [4] proposed a multi-center Gaussian model to capture the spatial clustering phenomenon. In addition, Zhang et al. [44] developed a kernel density estimation approach to capture the personalized geographical influence. Lian et al. [12] proposed MF-based POI recommendation method which captures the spatial clustering phenomenon from the aspect of two-dimensional kernel density estimation. Li et al. [35] proposed a ranking based geographical factorization method, which exploits both geographical and temporal contexts for POI recommendation.

Temporal information has been proved as another important type of context for POI recommendation and has attracted significant attention from some researchers. For example, Yuan et al. [2] incorporated temporal information into a user-based CF recommender by dividing time into 24 time slots. Furthermore, Gao et al. [32] proposed MF-based location recommendation framework which investigated the temporal properties of users' check-in behavior.

Besides, social information and other POI characteristic information have also been studied for POI recommendation. For instance, Gao et al. [1] proposed a social-historical model which integrated the social and historical effects and assessed the role of social correlation in user's check-in behavior. Yang et al. [37] fused the spatial and social information, and user tips with a location-based social matrix factorization algorithm.

However, most of these methods fail to study the spatiotemporal sequential influence of the user's check-ins history, which are very important for mining dynamic user's behavior and preferences.

2.2. Markov Chain Based Methods

Since historical check-in information in different time periods and spatial locations have different effects on users' behavior, sequential influence should be considered for POI recommendation. Most of the existing studies usually employ the properties of a Markov chain to model the sequential influence [45–51]. For instance, Rendle et al. [45] first proposed a state-of-the-art personalized Markov chain model, namely FPMC, which implemented the recommended task for sequence data in an MF-based approach. Rather than merely modeling temporal information, Cheng et al. [3] employed FPMC to model the personalized POI transition and considers users' movement constraint. Mathew et al. [46] proposed a hybrid method to predict human movement by a hidden Markov model

(HMM). Chen et al. [47] proposed a POI recommendation with Markov modeling, which considered both individual and collective movement patterns in making prediction. Ye et al. [16] attempted to model the underlying user movement pattern by using check-in category information and proposed a mixed hidden Markov model to predict the most likely next location. Zhang et al. [48] proposed a novel HMM based group-level mobility modeling framework. Similarly, some other POI recommendation methods based on Markov chain have also been proposed [49–51]. However, the drawbacks of MC based methods are there strong Markov assumptions among different components, and they are challenging to model long-term dependency. Although Personalized Ranking Metric Embedding (PRME) method [52] learns a personalized metric embedding and models the sequential POI transition, it merely models short-term transition patterns within users' movements.

2.3. Deep Learning Based Methods

Deep learning has been successfully applied to the POI recommendation system in recent years. Many such methods have been introduced or used in POI recommendation, such as Word2vec [53], multilayer perceptron (MLP) [54,55], convolutional neural network (CNN) [56,57] and deep neural network (DNN) [58]. Zhao et al. [34] proposed a temporal POI embedding model by introducing the word2vec framework, which incorporated both sequential and spatial-temporal context influence. Yang et al. [54] developed a deep neural architecture called Preference and Context Embedding (PACE) to bridge collaborative filtering and semi-supervised learning for POI recommendation. Wang et al. [56] proposed a novel POI recommender system, which used CNN to learn user and POI latent features from images. Ding et al. [58] proposed a DNN-based POI recommendation framework, and incorporated co-visiting pattern, geographical influence, and categorical correlation to alleviate the data sparsity issue.

Recently, recurrent neural networks (RNNs) [17] have become more and more powerful in modeling sequential history and transition of the user's movement. Moreover, they have been successfully applied to many fields like sequential click prediction [59], session-based recommendation [23] and mobility prediction [8], and so on. With the help of gated activation function like gated recurrent unit (GRU) [19] and long-short term memory (LSTM) [18], they can better capture the long-term dependencies. Hidasi et al. [23] first applied recurrent neural network with GRU for sequence recommendation, and their experimental results have shown a significant improvement over traditional methods. Spatial and temporal contextual information has shown its importance on different tasks. Liu et al. [8] employed a spatial and temporal recurrent neural network (ST-RNN) to model spatiotemporal contextual information with continuous values for location prediction. However, it ignores long-term dependencies, and the standard RNN method may suffer from the exploding or vanishing gradients problem. Zhao et al. [60] proposed a new variant of LSTM, which implemented time gates and distance gates into LSTM to capture the spatiotemporal relation between successive check-ins. Cui et al. [22] proposed a Distance-to-Preference (Distance2Pre) network for the next POI prediction, which modeled check-in sequences and successive distances to acquire the user's sequential preference and spatial preference. However, the above methods were unable to capture different contributions of each POIs in the history check-in sequence. In other words, it is difficult for them to extract the user's main intentions in the current sequence.

A large amount of research has benefited from the attention mechanism model in recent years [61–63], which does not only enhance the ability of the neural network to capture long-term dependencies but also enhances the interpretability of neural networks. Based on the seq2seq model [64], Bahdanau et al. [65] introduced an attention mechanism into the neural machine translation task. Vaswani et al. [61] proposed a new simple network architecture to encode an input sequence into an output sequence using the attention mechanism. Feng et al. [66] proposed an attentional GRU network for user's movement prediction from sparse data. Unlike existing studies, our work combines geographical distances and time intervals contextual information into a more robust GRU network to

capture user's sequential preference and spatiotemporal preference. Besides, an attention model is introduced to capture the user's main intentions.

3. Proposed ATST-GRU Model

In this section, we first address the problem and introduce the basic GRU model. Then we present proposed ST-GRU and ATST-GRU network. Finally, we train our model with the BPR framework and the BPTT algorithm.

3.1. Problem Statement

For convenience of expression, we give several important definitions, and the essential notations are listed in Table 1.

Definition 1. *POI.* A point-of-interest (POI) is a uniquely identified spatial location. In this paper, we use v to represent a POI and $\mathcal{Q} = \{v_1, v_2, \dots\}$ represents the set of POIs. Each POI v has a unique identifier and geographical coordinates, which include geographical latitude and geographical longitude.

Definition 2. *Check-in sequence.* A check-in sequence represents that a user's history check-ins are arranged in chronological order, denoted by $\mathcal{C}_u = \{c_{t_1}^u, c_{t_2}^u, \dots, c_{t_N}^u\}$.

Definition 3. *POI recommendation.* Given a set of users' check-in sequences \mathcal{C}^U and a set of POIs \mathcal{Q} , the POI recommendation task is to recommend top- k POIs that user u would be interested in.

Table 1. Notations in this paper.

Notation	Explanation
u, v, l, t	user, POI, location, time
$\mathcal{P}, \mathcal{Q}, \mathcal{L}$	set of users, set of POIs, set of locations
p_u	latent representation of user u
$v_{t_k}^u$	POI visited by user u at time point t_k
$h_{t_k}^u$	the hidden vector of GRU units
$\tilde{h}_{t_k}^u$	candidate state of GRU units
$r_{t_k}^u, z_{t_k}^u$	reset gate vector and update gate vector of GRU units
$\mathcal{C}_u = \{c_{t_1}^u, c_{t_2}^u, \dots, c_{t_N}^u\}$	check-in sequence of user u
$\{\mathbf{U}\}, \{\mathbf{W}\}$	set of weight matrices for a GRU network
$\{\mathbf{b}\}$	set of bias vectors for a GRU network
o_{u, t_{N+1}, v_k}	predicted probability that u visits POI v_k at t_{N+1}
σ	sigmoid function
$\alpha_{t_N t_k}$	attention weight factors
$s_{t_k}^u, \delta_{t_k}^u$	vector representations of spatial and temporal intervals

3.2. Gated Recurrent Unit

The primary challenge of the POI recommendation task is modelling the user's sequential preference and the spatiotemporal preference, this can be considered as a sequence modelling problem. As we know, a good choice is RNN architectures. In this paper, we choose the GRU rather than a standard RNN because it can work better to deal with the gradient vanishing and gradient exploding problem. Hidasi et al. [23] demonstrate that GRU outperformed LSTM in the sequence-based recommendation. The hidden unit of GRU contains a reset gate $r_{t_k}^u$ and an update gate $z_{t_k}^u$ to control the flow of information. The formulas are as follows:

$$\begin{aligned}
z_{t_k}^u &= \sigma(\mathbf{U}_z \mathbf{v}_{t_k}^u + \mathbf{W}_z \mathbf{h}_{t_{k-1}}^u + \mathbf{b}_z) \\
r_{t_k}^u &= \sigma(\mathbf{U}_r \mathbf{v}_{t_k}^u + \mathbf{W}_r \mathbf{h}_{t_{k-1}}^u + \mathbf{b}_r) \\
\tilde{h}_{t_k}^u &= \tanh(\mathbf{U}_c \mathbf{v}_{t_k}^u + \mathbf{W}_c (r_{t_k}^u \odot \mathbf{h}_{t_{k-1}}^u) + \mathbf{b}_c) \\
\mathbf{h}_{t_k}^u &= (1 - z_{t_k}^u) \odot \mathbf{h}_{t_{k-1}}^u + z_{t_k}^u \odot \tilde{h}_{t_k}^u
\end{aligned} \tag{1}$$

where $\mathbf{U}_z, \mathbf{U}_r, \mathbf{U}_c, \mathbf{W}_z, \mathbf{W}_r, \mathbf{W}_c$ are transition matrices and $\mathbf{b}_z, \mathbf{b}_r, \mathbf{b}_c$ are the biases. $\mathbf{v}_{t_k}^u \in R^d$ is the input vector of user u , and t_k is the time step. $\tilde{h}_{t_k}^u$ is the candidate state activated by element-wise $\tanh(\cdot)$, $\mathbf{h}_{t_k}^u$ is the hidden vector. σ denotes the sigmoid function $\sigma(x) = 1/(1 + e^{-x})$ \odot represents the element-wise multiplication between two vectors.

In GRU network, the prediction of next POI can be calculated the inner product of user and POI representations. In our network, we regard the last hidden vectors $\mathbf{h}_{t_N}^u$ as the representation of the user. Like MF-based POI recommendation approaches [4,12,35], a user's preference for a POI by considering sequential preference is denoted as:

$$o_{u,t_{N+1},v_k} = (\mathbf{h}_{t_N}^u)^T \mathbf{v}_k \tag{2}$$

where o_{u,t_{N+1},v_k} represents the predicted probability that user u visits POI v_k at time point t_{N+1} .

3.3. Spatiotemporal GRU Network

Spatial and temporal contextual information are the basis for mining user movement behavior patterns, which help us to understand the behavior background more precisely to improve user behavior modeling. General sequence modeling only considers the order relationship between check-in behavior, ignoring the continuous geographical distances and time intervals information. However, these continuous time intervals and geographical distance values are crucial for modeling user behavior and mining user's preference in personalized POI recommendation systems.

We argue that spatial and temporal context can work as implicit information to guide the learning of gate mechanism. We propose to add the continuous geographical distances and time intervals information into the basic GRU network, which more naturally captures personalized spatiotemporal preferences for POI recommendation. Figure 3 illustrates the architecture of ST-GRU network, at each time step, each ST-GRU unit takes an embedded vector $\mathbf{v}_{t_k}^u$, a spatial contexts vector $\mathbf{s}_{t_k}^u$ and a temporal contexts vector $\mathbf{g}_{t_k}^u$ as inputs. In this way, the output of ST-GRU is a hidden layer vector $\mathbf{h}_{t_k}^u$, which indicates the combined influence of POIs and spatiotemporal contexts information. The formulas are:

$$\begin{aligned}
z_{t_k}^u &= \sigma(\mathbf{U}_z \mathbf{v}_{t_k}^u + \mathbf{W}_z \mathbf{h}_{t_{k-1}}^u + \mathbf{W}_{sz} \mathbf{s}_{t_k}^u + \mathbf{W}_{gz} \mathbf{g}_{t_k}^u + \mathbf{b}_z) \\
r_{t_k}^u &= \sigma(\mathbf{U}_r \mathbf{v}_{t_k}^u + \mathbf{W}_r \mathbf{h}_{t_{k-1}}^u + \mathbf{W}_{sr} \mathbf{s}_{t_k}^u + \mathbf{W}_{gr} \mathbf{g}_{t_k}^u + \mathbf{b}_r) \\
\tilde{h}_{t_k}^u &= \tanh(\mathbf{U}_c \mathbf{v}_{t_k}^u + \mathbf{W}_{sh} \mathbf{s}_{t_k}^u + \mathbf{W}_{gh} \mathbf{g}_{t_k}^u + \mathbf{W}_c (r_{t_k}^u \odot \mathbf{h}_{t_{k-1}}^u) + \mathbf{b}_c) \\
\mathbf{h}_{t_k}^u &= (1 - z_{t_k}^u) \odot \mathbf{h}_{t_{k-1}}^u + z_{t_k}^u \odot \tilde{h}_{t_k}^u
\end{aligned} \tag{3}$$

where $\mathbf{s}_{t_k}^u$ and $\mathbf{g}_{t_k}^u$ denote the vector representations of geographical distances and time intervals between $\mathbf{v}_{t_{k-1}}^u$ and $\mathbf{v}_{t_k}^u$. $\mathbf{W}_{sz}, \mathbf{W}_{sr}, \mathbf{W}_{sh}$ and $\mathbf{W}_{gz}, \mathbf{W}_{gr}, \mathbf{W}_{gh}$ are transition matrices for $\mathbf{s}_{t_k}^u$ and $\mathbf{g}_{t_k}^u$, respectively. Therefore, $\mathbf{h}_{t_k}^u$ contains not only information of original input $\mathbf{v}_{t_k}^u$ but also important information of spatial contexts $\mathbf{s}_{t_k}^u$ and temporal contexts $\mathbf{g}_{t_k}^u$. The user's preferences in each hidden state are greatly enhanced.

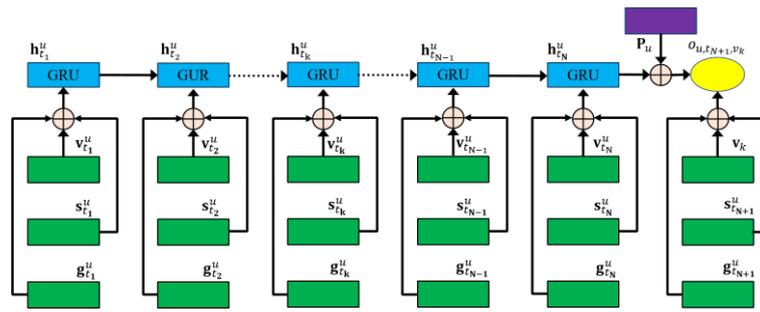


Figure 3. The architecture of ST-GRU network.

However, if we learn a distinct matrix for each continuous geographical distance and time interval, the ST-GRU network will face the data sparsity problem. We partition continuous geographical distances and time interval values into discrete bins respectively and utilize a linear interpolation to acquire their transition matrices as follows:

$$s = \frac{W_{U(\delta s)}[U(\delta s) - \delta s] + W_{L(\delta s)}[\delta s - L(\delta s)]}{U(\delta s) - L(\delta s)} \tag{4}$$

$$g = \frac{W_{U(\delta g)}[U(\delta g) - \delta g] + W_{L(\delta g)}[\delta g - L(\delta g)]}{U(\delta g) - L(\delta g)} \tag{5}$$

where $U(\delta s)$ and $L(\delta s)$ denote the upper bound and lower bound values of a specific geographical distance δs . Similarly, $U(\delta g)$ and $L(\delta g)$ indicate the upper bound and lower bound values of a specific time interval δg . $W_{U(\delta s)}$ and $W_{L(\delta s)}$ are the spatial factor matrix, and $W_{U(\delta g)}$ and $W_{L(\delta g)}$ are the temporal factor matrix.

Finally, POI recommendation for target users can be calculated the dot-product of user and POI representations, which is similar to Formula (2). And the prediction of whether user u would go to a location v_k at time t_{N+1} can be calculated as:

$$o_{u,t_{N+1},v_k} = (W_N h_{t_N}^u + W_p p_u)^T (W_v v_k + W_s s_{t_{N+1}}^u + W_g g_{t_{N+1}}^u) \tag{6}$$

where p_u is the permanent representation of a user, and it is specifically designed to indicate a user’s profile and long-term preference. $h_{t_N}^u$ is the dynamic representation of a user, which captures a user’s dynamic interests under a specific spatial and temporal contexts. W_N and W_p are the parameters of the output layer. W_v , W_s , and W_g are transition matrices.

3.4. Attention-Based Spatiotemporal GRU Network

Intuitively speaking, when we predict the user’s next behavior, all the users’ history check-in behavior does not contribute equally. Moreover, previous methods have not been able to capture the user’s main intention adequately. Therefore, in our model, we involve an attention mechanism to capture the user’s main purpose in the current sequence, which allows the different parts of the sequence of the past check-in behavior’ to be dynamically selected and linearly combined by the decoder. In other words, attention mechanism helps us select only the relevant and important POIs for next POI recommendation at each time step, and all the previous hidden states can be utilized by a weighted sum of visited POIs. Figure 4 illustrates the architecture of ATST-GRU network, in this encoding scheme we use ST-GRU as the basic component, where the weighted sum of hidden states is interpreted as the user’s main intention feature.

$$c_{t_N}^u = \sum_{k=1}^N \alpha_{t_N t_k} h_{t_k}^u \tag{7}$$

where the weighted factors $\alpha_{t_N t_k}$ determine which part of the history check-in sequence should be emphasized or ignored when making next behavior predictions in the POI recommendation model, which in turn is a function of hidden states as follows:

$$\alpha_{t_N t_k} = m(h_{t_N}^u, h_{t_k}^u) \tag{8}$$

$$m(h_{t_N}^u, h_{t_k}^u) = A_0^T \sigma(A_1 h_{t_N}^u + A_2 h_{t_k}^u) \tag{9}$$

where the function $m(h_{t_N}^u, h_{t_k}^u)$ is used to calculate the similarity between the final hidden state $h_{t_N}^u$ and the representation of the previously visited POI $h_{t_k}^u$. σ denotes the sigmoid function $\sigma(x) = 1/(1 + e^{-x})$. A_0, A_1 and A_2 are used to transform $h_{t_N}^u$ and $h_{t_k}^u$ into a common latent space.

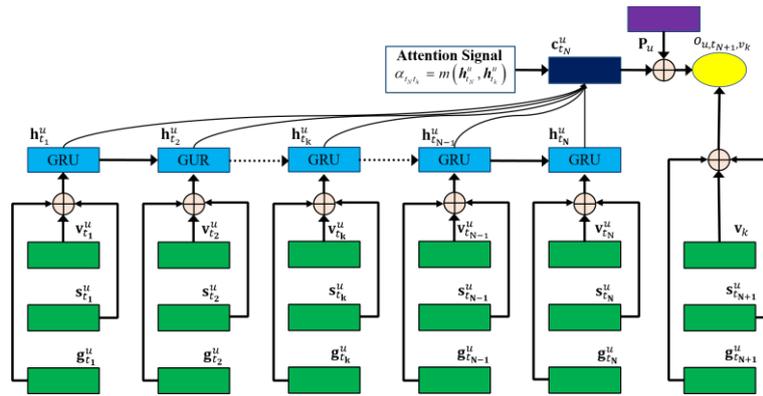


Figure 4. The architecture of attention-based spatiotemporal gated recurrent unit (ATST-GRU) network.

Similarly to Formula (6), the prediction of whether user u would go to a location v_k at time t_{N+1} can be computed as:

$$o_{u,t_{N+1},v_k} = (W_N c_{t_N}^u + W_p p_u)^T (W_v v_k + W_s s_{t_{N+1}}^u + W_g g_{t_{N+1}}^u) \tag{10}$$

where W_N and W_p are the parameters of the output layer. W_v, W_s , and W_g are transition matrices.

3.5. Network Learning

In this subsection, we train our ATST-GRU network under the Bayesian Personalized Ranking (BPR) framework by using the backpropagation through time (BPTT) algorithm. These methods have been successfully used for network training of RNN based recommendation models [8,22]. BPR is a pairwise ranking framework that is widely used to process implicit feedback data. The basic assumption of BPR is that a user prefers previously visited POIs than negative ones. In the BPR framework, at each sequential position k , the objective of ATST-GRU is to maximize the following probability:

$$p(u, t, v > v') = g(o_{u,t,v} - o_{u,t,v'}) \tag{11}$$

where v and v' denote a positive location and a negative location, respectively. Additionally, a negative location is randomly chosen from location sets that users have not visited. $g(\cdot)$ is a nonlinear sigmoid function $g(x) = 1/(1 + e^{-x})$.

Finally, by incorporating the negative log-likelihood, we can solve the objective function for POI recommendation as follows:

$$\begin{aligned} J &= -\sum p(u, t, v > v') + \frac{\lambda}{2} \|\Theta\|^2 \\ &= \sum \ln(1 + e^{-(o_{u,t,v} - o_{u,t,v'})}) + \frac{\lambda}{2} \|\Theta\|^2 \end{aligned} \tag{12}$$

where $\Theta = \{\mathbf{U}, \mathbf{W}, \mathbf{b}, A_0, A_1, A_2\}$ is the set of parameters, \mathbf{U} represent the set of weight matrices which include \mathbf{U}_z , \mathbf{U}_r and \mathbf{U}_c , which is similar to \mathbf{W} and \mathbf{b} . λ is the regularization parameter. Then, we use stochastic gradient descent (SGD) and BPTT to optimize the network parameters in this study. Additionally, the range of initialization parameters was $(-0.5$ to $0.5)$. The size of each batch was set to 100. The regularization and the initial learning rate were set to 0.001 and 0.01, respectively. Moreover, our model is trained on a GeForce GTX TitanX GPU, the code used in our experiments was written by using Theano and Python 3.5. The learning algorithm of ATST-GRU is summarized in Algorithm 1.

Algorithm 1: ATST-GRU

Input: historical check-in sequences C_u .
Output: model parameters Θ .
 //construct training instances
 $D \leftarrow \phi$
For each user u in P **do**
For each check-in sequence $C_u = \{c_{t_1}^u, c_{t_2}^u, \dots, c_{t_N}^u\}$ **do**
 Get the set of negative samples v' ;
For each check-in activity in C_u **do**
 Compute the embedded vector $v_{t_k}^u$;
 Compute the spatial contexts vector $s_{t_k}^u$;
 Compute the temporal contexts vector $g_{t_k}^u$;
End for
 Add a training instance $(\{(v_{t_k}^u, s_{t_k}^u, g_{t_k}^u)\}, \{v'\})$ into D ;
End for
End for
 Initialize the parameter set Θ ;
repeat
For each user u in P **do**
 Randomly select a batch of instances D_m^u from D ;
 Update Θ by minimizing the object with D_m^u
End for
until convergence;
return $\Theta = \{\mathbf{U}, \mathbf{W}, \mathbf{b}, A_0, A_1, A_2\}$

4. Experimental Results and Analysis

In this section, we conduct empirical experiments on two publicly-available datasets to validate the effectiveness of the proposed method. First, we introduce the datasets, baseline methods and evaluation metrics. Then we compare ATST-GRU with some state-of-the-art POI recommendation methods. Finally, we study the effects of different model parameters.

4.1. Experimental Settings

4.1.1. Datasets

We applied two widely-used publicly-available LBSNs datasets called Foursquare and Gowalla to evaluate the performance of different methods, and the datasets were preprocessed in [2]. In the Foursquare dataset, all check-ins data were collected at Singapore, from August 2010–July 2011. In the Gowalla dataset, all check-ins data were collected in California and Nevada, from February 2019–October 2010. Figure 5 presents the check-in distribution in the Foursquare and Gowalla datasets, where the locations are concentrated in some geographical regions. Moreover, we randomly selected several different user's check-in sequences on the two datasets and visualized them on the map. As shown in Figure 6, the results of map visualization have shown that people prefer to visit nearby POIs and the visited POIs often form spatial clusters. In particular, we can observe that

people may have different moving patterns and different users have different preferences for travel distance. This further supports that considering the spatial impact can effectively improve the POI recommendation performance.

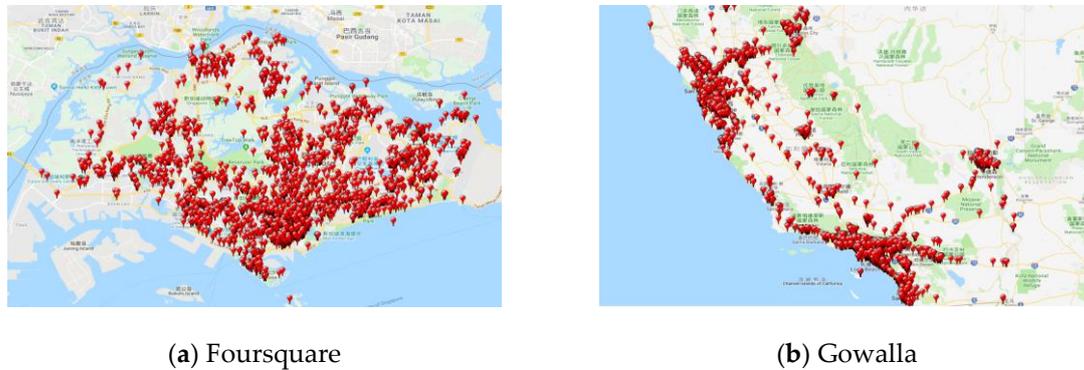


Figure 5. All users' check-in distribution of POIs in the two datasets.

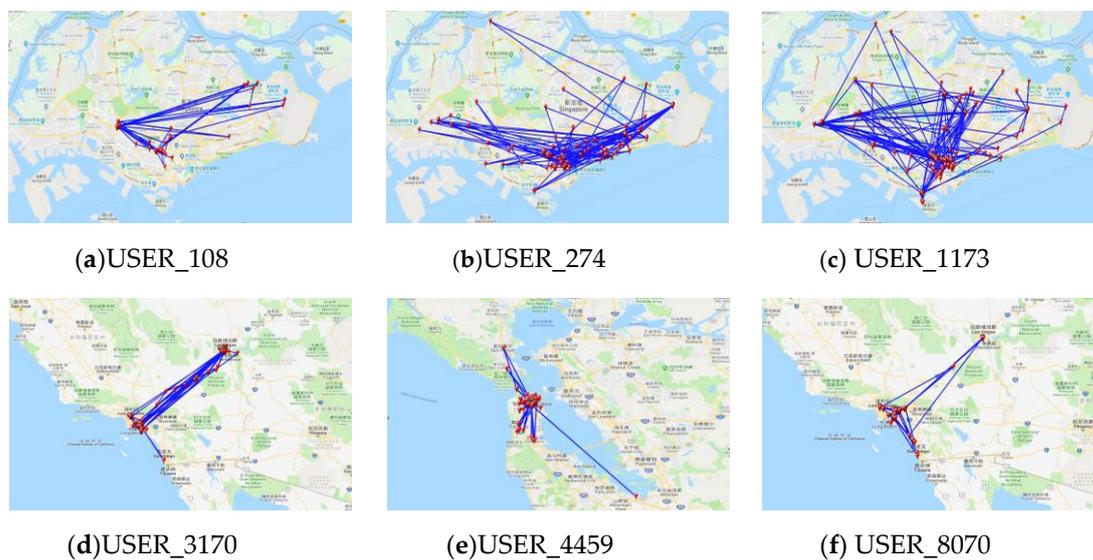


Figure 6. Different users' check-in sequences on the map. (a–c) and (d–f) are from the Foursquare and Gowalla datasets, respectively.

Also, to alleviate data sparsity and cold start problems, we removed POIs checked in by less than five users and users who have checked in fewer than five POIs. After pre-processing, the basic statistics of two datasets are summarized in Table 2. Inspired by previous studies [22,26], we employed the leave-one-out evaluation. We used the last POI of each user's check-in sequence as the test data and the remaining POI as the training data.

Table 2. Basic statistics of Foursquare and Gowalla datasets.

Dataset	#Users	#POIs	#Check-ins	Sparsity
Foursquare	2321	5596	194,108	99.18%
Gowalla	10,162	24,237	456,967	99.88%

4.1.2. Baseline Methods

We compared the effectiveness of our proposed ST-GRU and ATST-GRU model with the following state-of-the-art POI recommendation approaches.

- BPR [26]. This method is a generic optimization criterion and learning algorithm for personalized ranking, and we applied BPR in POI recommendation;
- GRU [19]. GRU network is a more robust variant of RNN which work better in capturing long term dependencies, and we applied GRU in POI recommendation;
- FPMC-LR [3]. This method is a state-of-the-art Markov chain method that models personalized sequential transitions for POI recommendation;
- PRME-G [52]. This method is a state-of-the-art metric embedding method for POI recommendation, which integrates geographical influence and sequential information;
- Rank-GeoFM [35]. It is a state-of-the-art ranking-based factorization model for POI recommendation, which incorporates the geographical influence and temporal influence;
- ST-RNN [8]. This is a state-of-the-art RNN-based model for successive POI recommendation. It incorporates both local temporal and spatial transition context within the RNN architecture;
- DeepMove [66]: It is a state-of-the-art attentional RNN model which capture the sequential transitions by jointly embedding the multiple factors.

4.1.3. Evaluation Metrics

To evaluate the performance of the above methods, we applied two popular evaluation metrics called $Recall@k$ and $F_1-score@k$ as follows:

$$R@k = \frac{1}{N} \sum_{u=1}^N \frac{|R(u) \cap T(u)|}{T(u)} \quad (13)$$

$$F_1@k = \frac{1}{N} \sum_{u=1}^N \frac{2 \cdot (|R(u) \cap T(u)|/k) \cdot (|R(u) \cap T(u)|/T(u))}{(|R(u) \cap T(u)|/k) + (|R(u) \cap T(u)|/T(u))} \quad (14)$$

where k indicates the number of POIs recommended to the user, we reported $R@k$ and $F_1@k$ with $k = 5, 10, 15$ and 20 in our experiments. $R(u)$ indicates the Top- k list recommended to the user. $T(u)$ represents the number of POIs the user actually visited.

4.2. Comparison and Results

Figure 7 shows the performance of all methods on the Foursquare and the Gowalla datasets. We made the following observations:

First, we explored traditional baselines BPR, GRU, FPMC-LR, PRME-G and Rank-GeoFM. For both the two datasets, we can see that BPR and GRU dropped behind other algorithms since they did not take into account other useful information such as geographical influence and temporal information. Besides, GRU performed slightly better than BPR, indicating that modeling sequential influence is effective for POI recommendation. Compared with BPR and GRU, FPMC-LR and PRME-G employed both geographical and sequential information in LBSNs, and their performance on two datasets were better. Specifically, we observed that Rank-GeoFM performed obviously better than FPMC-LR and PRME-G. There are two possible reasons for this result: Rank-GeoFM incorporated both geographical influence and temporal context which could work better to capture spatiotemporal preference and to deal with the data sparsity problems. Therefore, the above analysis also supports that spatial influence, temporal influence, and sequential influence are critical factors in improving POI recommendation performance.

In the following, we compare the above traditional methods with RNN-based methods (i.e., ST-RNN, ST-GRU, DeepMove). Firstly, compared with the above traditional methods, ST-RNN and ST-GRU outperformed them significantly, this was because the RNN-based structure combined with spatiotemporal context information can better capture the user's sequence preferences and spatiotemporal behavior preferences. We can see that ST-GRU greatly outperformed GRU, indicating that ST-GRU benefits from considering the spatiotemporal features. Besides, we observe that the

ST-GRU outperformed the ST-RNN, which may be due to the advantage of GRUs over RNNs, i.e., GRU is a more robust network structure which works better in capturing long term dependencies and alleviating the exploding or vanishing gradients problem. Specifically, DeepMove obtained much better performance than ST-RNN and ST-GRU as it introduces the attention mechanisms and considers period influence. The above discussion indicates that using RNN-based network structure and attention model can effectively improve the performance of POI recommendation.

Finally, we find that the proposed ATST-GRU model performed significantly better than all existing state-of-the-art methods evaluated here on the two datasets. Specifically, ATST-GRU outperformed the conventional matrix factorization methods significantly by a large margin. For instance, on the two datasets, compared with Rank-GeoFM, the ATST-GRU improved $R@5$, $R@10$, $R@15$, $R@20$, $F_1 @5$, $F_1 @10$, $F_1 @15$ and $F_1 @20$ by 35.25%, 33.58%, 42.04%, 52.61%, 35.20%, 33.76%, 42.08% and 52.70% on average, respectively. Moreover, ATST-GRU consistently outperforms three RNN-based methods: ST-RNN, ST-GRU and DeepMove. For example, on the two datasets, compared with DeepMove, the ATST-GRU improved $R@5$, $R@10$, $R@15$, $R@20$, $F_1 @5$, $F_1 @10$, $F_1 @15$ and $F_1 @20$ by 8.07%, 7.55%, 6.70%, 9.12%, 8.10%, 7.58%, 6.78%, 9.19% on average, respectively.

In summary, the experimental results suggest that the proposed ATST-GRU can successfully capture the user's sequential preference, spatiotemporal preference, and main behavioral intention, leading to a superior performance for POI recommendation.

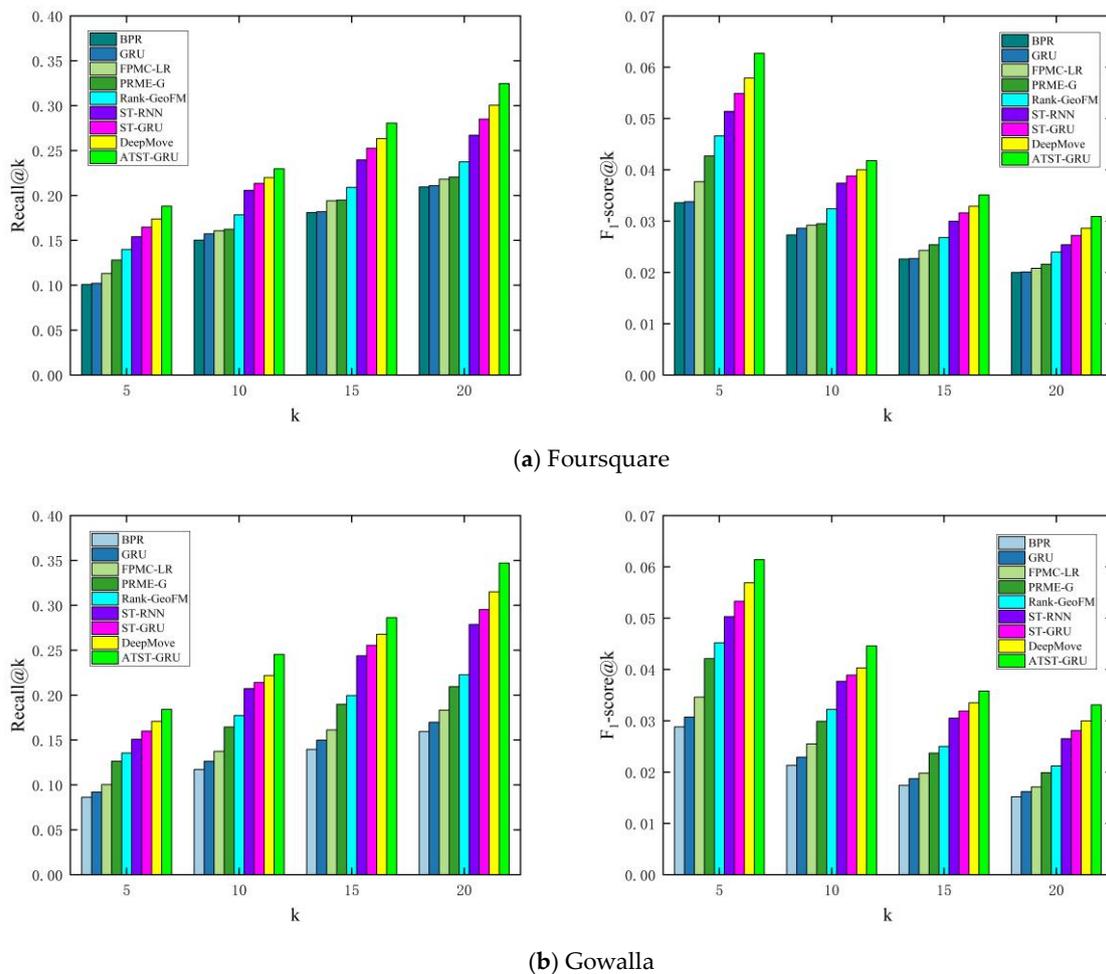


Figure 7. Recommendation performance comparison on Foursquare and Gowalla.

4.3. Influence of Embedding Dimension Size

We further studied the effect of the embedding dimension size on the performance of our ATST-GRU network. In general, a higher number of embedding dimensions may enhance the performance of the model. However, it also leads to over-fitting. Here, we varied the dimension number from 10 to 150 and computed the network's generalization using Recall@ k and F₁-score@ k with $k = 10, 20$ for each case. Figure 8 shows the Recall@ k and F₁-score@ k values for different dimension numbers on the two datasets. We observed that our ATST-GRU network achieved stable performance in the range of 70–150 and 90–150 on the Foursquare and Gowalla datasets, respectively. Therefore, on the Foursquare dataset and Gowalla dataset, we could set the number of embedding dimensions to 70 and 90 in our experiments, respectively.

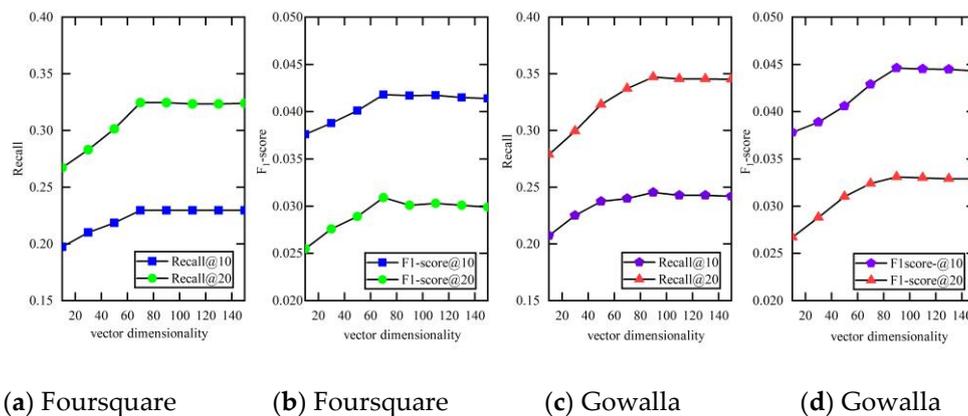


Figure 8. Effect of the number of dimensions in ATST-GRU.

4.4. Influence of Different Spatial and Temporal Window Widths

Spatial and temporal window widths are essential factors to affect the performance of ATST-GRU, and we also did a batch of experiments on the two datasets with different spatial and temporal window width settings. Table 3 illustrates the performance of ATST-GRU evaluated by Recall@20 with varying window widths. On the Foursquare dataset, we can observe that we achieved best prediction performance when using a spatial window width of 0.3 km and temporal window width of 12 h. On the Gowalla dataset, the best prediction performance of recall@20 was obtained with a spatial window width of 0.1 km and temporal window width of 48 h. Moreover, we observe that ATAT-GRU outperformed state-of-the-art methods even when the spatial and temporal window's width were not optimal. The results further suggest the superiority of the proposed ATST-GRU for POI recommendation.

Table 3. Performance of AST-GRU with varying window width by Recall@20.

Dataset	Window	6 h	12 h	24 h	48 h	60 h
Foursquare	0.1 km	0.3215	0.3205	0.3106	0.3032	0.3011
	0.3 km	0.3201	0.3247	0.3009	0.3105	0.3093
	0.5 km	0.3126	0.3133	0.3211	0.3119	0.3042
Gowalla	0.1 km	0.3158	0.3301	0.3325	0.3471	0.3412
	0.3 km	0.3192	0.3215	0.3298	0.3392	0.3317
	0.5 km	0.3204	0.3287	0.3153	0.3266	0.3255

5. Conclusions

In recent years, POI recommendation based on deep learning has widely attracted attention in academia and industry. Compared to general POI recommendation, we focus on next POI

recommendation task in this work, and comprehensively utilize user's check-in sequence information and spatiotemporal information to mine user's movement behavior rules and preferences. Hence, an innovative attention-based spatiotemporal GRU (ATST-GRU) network is proposed to tackle the POI recommendation problem in this paper. ATST-GRU introduces spatial-temporal factors into the gate mechanism of GRU to model the spatiotemporal contextual information and sequential nature. Such a network structure can better capture the user's spatiotemporal preferences and alleviate the sparsity of the data. More specifically, the contextual attention-based modeling can capture the important information of the user's historical behavior, and this greatly enhances the modeling of user's main interest. Besides, we validate the effectiveness of ATST-GRU by using two real-life mobility datasets (i.e., Foursquare and Gowalla). The experimental results show that ATST-GRU outperforms other state-of-the-art methods for POI recommendation.

In the future, we will focus on extending the current study by considering other check-in features (e.g., the semantic context of POI, user comments, social relations) or other more advanced neural networks (e.g., graph neural networks). These may motivate the model to improve performance of POI recommendation.

Author Contributions: Conceptualization and Methodology, Chunyang Liu and Jiping Liu; Validation and Formal analysis, Chunyang Liu, Shenghua Xu and Jian Wang; Writing-original draft preparation, Chunyang Liu and Houzeng Han; Writing-review and editing, Chunyang Liu and Yang Chen.

Funding: This research is partially supported by the National Key Research and Development Program of China (2016YFC0803101 and 2016YFC0803108).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Gao, H.; Tang, J.; Liu, H. Exploring social-historical ties on location-based social networks. In Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media, Dublin, Ireland, 4–7 June 2012.
2. Yuan, Q.; Cong, G.; Ma, Z.; Sun, A.; Thalmann, N.M. Time-aware point-of-interest recommendation. In Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval, Dublin, Ireland, 28 July–1 August 2013; pp. 363–372.
3. Cheng, C.; Yang, H.; Lyu, M.R.; King, I. Where you like to go next: Successive point-of-interest recommendation. In Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence, Beijing, China, 3–19 August 2013.
4. Cheng, C.; Yang, H.; King, I.; Lyu, M.R. Fused matrix factorization with geographical and social influence in location-based social networks. In Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence, Toronto, ON, Canada, 22–26 July 2012.
5. Ye, M.; Yin, P.; Lee, W.C.; Lee, D.L. Exploiting geographical influence for collaborative point-of-interest recommendation. In Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval, Beijing, China, 24–28 July 2011; pp. 325–334.
6. Cai, L.; Xu, J.; Liu, J.; Pei, T. Integrating spatial and temporal contexts into a factorization model for POI recommendation. *Int. J. Geogr. Inf. Sci.* **2018**, *32*, 524–546. [[CrossRef](#)]
7. Yin, H.; Sun, Y.; Cui, B.; Hu, Z.; Chen, L. LCARS: A location-content-aware recommender system. In Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining, Chicago, IL, USA, 11–14 August 2013; pp. 221–229.
8. Liu, Q.; Wu, S.; Wang, L.; Tan, T. Predicting the next location: A recurrent model with spatial and temporal contexts. In Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, Phoenix, AZ, USA, 12–17 February 2016.
9. He, J.; Li, X.; Liao, L.; Song, D.; Cheung, W.K. Inferring a personalized next point-of-interest recommendation model with latent behavior patterns. In Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, Phoenix, Arizona USA, 12–17 February 2016.
10. Anagnostopoulos, A.; Atassi, R.; Becchetti, L.; Fazzino, A.; Silvestri, F. Tour recommendation for groups. *Data Min. Knowl. Discov.* **2017**, *31*, 1157–1188. [[CrossRef](#)]

11. Ayala-Gómez, F.; Daróczy, B.Z.; Mathioudakis, M.; Benczúr, A.; Gionis, A. Where could we go? Recommendations for groups in location-based social networks. In Proceedings of the 2017 ACM on Web Science Conference, Troy, New York, NY, USA, 25–28 June 2017; pp. 93–102.
12. Lian, D.; Zhao, C.; Xie, X.; Sun, G.; Chen, E.; Rui, Y. GeoMF: Joint geographical modeling and matrix factorization for point-of-interest recommendation. In Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 24–27 August 2014; pp. 831–840.
13. Gan, M.; Gao, L. Discovering Memory-Based Preferences for POI Recommendation in Location-Based Social Networks. *ISPRS Int. J. Geo Inf.* **2019**, *8*, 279. [[CrossRef](#)]
14. Tobler, W.R. A computer movie simulating urban growth in the Detroit region. *Econ. Geogr.* **1970**, *46*, 234–240. [[CrossRef](#)]
15. Cho, E.; Myers, S.A.; Leskovec, J. Friendship and mobility: User movement in location-based social networks. In Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, CA, USA, 21–24 August 2011; pp. 1082–1090.
16. Ye, J.; Zhu, Z.; Cheng, H. What's your next move: User activity prediction in location-based social networks. In Proceedings of the 2013 SIAM International Conference on Data Mining, Austin, TX, USA, 2–4 May 2013; pp. 171–179.
17. Williams, R.J.; Zipser, D. A learning algorithm for continually running fully recurrent neural networks. *Neural Comput.* **1989**, *1*, 270–280. [[CrossRef](#)]
18. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)] [[PubMed](#)]
19. Cho, K.; Van Merriënboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv* **2014**, arXiv:1406.1078.
20. Duives, D.C.; Wang, G.; Kim, J. Forecasting pedestrian movements using recurrent neural networks: An application of crowd monitoring data. *Sensors* **2019**, *19*, 382. [[CrossRef](#)]
21. Quadrana, M.; Cremonesi, P.; Jannach, D. Sequence-aware recommender systems. *ACM Comput. Surv. (CSUR)* **2018**, *51*, 66. [[CrossRef](#)]
22. Cui, Q.; Tang, Y.; Wu, S.; Wang, L. Distance2Pre: Personalized Spatial Preference for Next Point-of-Interest Prediction. In Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining, Macau, China, 14–17 April 2019; pp. 289–301.
23. Hidasi, B.; Karatzoglou, A.; Baltrunas, L.; Tikk, D. Session-based recommendations with recurrent neural networks. *arXiv* **2015**, arXiv:1511.06939, 2015.
24. Mnih, V.; Heess, N.; Graves, A. Recurrent models of visual attention. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 8–13 December 2014; pp. 2204–2212.
25. Wang, S.; Hu, L.; Cao, L.; Huang, X.; Lian, D.; Liu, W. Attention-based transactional context embedding for next-item recommendation. In Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018.
26. Rendle, S.; Freudenthaler, C.; Gantner, Z.; Schmidt-Thieme, L. BPR: Bayesian personalized ranking from implicit feedback. In Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence, Montreal, QC, Canada, 18–21 June 2009; pp. 452–461.
27. Werbos, P.J. Backpropagation through time: What it does and how to do it. *Proc. IEEE* **1990**, *78*, 1550–1560. [[CrossRef](#)]
28. Hu, Y.; Koren, Y.; Volinsky, C. Collaborative Filtering for Implicit Feedback Datasets. In Proceedings of the 8th IEEE International Conference on Data Mining, Pisa, Italy, 15–19 December 2008; pp. 263–272.
29. Wang, H.; Terrovitis, M.; Mamoulis, N. Location recommendation in location-based social networks using user check-in data. In Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Orlando, FL, USA, 5–8 November 2013; pp. 374–383.
30. Koren, Y. Factorization meets the neighborhood: A multifaceted collaborative filtering model. In Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Las Vegas, NV, USA, 24–27 August 2008; pp. 426–434.
31. Liu, B.; Fu, Y.; Yao, Z.; Xiong, H. Learning geographical preferences for point-of-interest recommendation. In Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Chicago, IL, USA, 11–14 August 2013; pp. 1043–1051.

32. Gao, H.; Tang, J.; Hu, X.; Liu, H. Exploring temporal effects for location recommendation on location-based social networks. In Proceedings of the 7th ACM Conference on Recommender Systems, Hong Kong, China, 12–16 October 2013; pp. 93–100.
33. Zhao, S.; Zhao, T.; Yang, H.; Lyu, M.R.; King, I. STELLAR: Spatial-temporal latent ranking for successive point-of-interest recommendation. In Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, Phoenix, AZ, USA, 12–17 February 2016.
34. Zhao, S.; Zhao, T.; King, I.; Lyu, M.R. Geo-teaser: Geo-temporal sequential embedding rank for point-of-interest recommendation. In Proceedings of the 26th International Conference on World Wide Web Companion. International World Wide Web Conferences Steering Committee, Perth, Australia, 3–7 April 2017; pp. 153–162.
35. Li, X.; Cong, G.; Li, X.L.; Pham, T.A.N.; Krishnaswamy, S. 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, 9–13 August 2015; pp. 433–442.
36. Huang, L.; Ma, Y.; Liu, Y.; Sangaiah, A.K. Multi-modal Bayesian embedding for point-of-interest recommendation on location-based cyber-physical-social networks. *Future Gener. Comput. Syst.* **2017**. [[CrossRef](#)]
37. Yang, D.; Zhang, D.; Yu, Z.; Wang, Z. A sentiment-enhanced personalized location recommendation system. In Proceedings of the 24th ACM Conference on Hypertext and Social Media, Paris, France, 1–3 May 2013; pp. 119–128.
38. Liu, B.; Xiong, H.; Papadimitriou, S.; Fu, Y.; Yao, Z. A general geographical probabilistic factor model for point of interest recommendation. *IEEE Trans. Knowl. Data Eng.* **2014**, *27*, 1167–1179. [[CrossRef](#)]
39. Kefalas, P.; Manolopoulos, Y. A time-aware spatio-textual recommender system. *Expert Syst. Appl.* **2017**, *78*, 396–406. [[CrossRef](#)]
40. Ren, X.; Song, M.; Haihong, E.; Song, J. Context-aware probabilistic matrix factorization modeling for point-of-interest recommendation. *Neurocomputing* **2017**, *241*, 38–55. [[CrossRef](#)]
41. Levandoski, J.J.; Sarwat, M.; Eldawy, A.; Mokbel, M.F. Lars: A location-aware recommender system. In Proceedings of the 2012 IEEE 28th International Conference on Data Engineering, Arlington, VA, USA, 1–5 April 2012; pp. 450–461.
42. Kurashima, T.; Iwata, T.; Hoshide, T.; Takaya, N.; Fujimura, K. Geo topic model: joint modeling of user's activity area and interests for location recommendation. In Proceedings of the Sixth ACM International Conference on Web Search and Data Mining, Rome, Italy, 4–8 February 2013; pp. 375–384.
43. Zhang, J.D.; Chow, C.Y. TICRec: A probabilistic framework to utilize temporal influence correlations for time-aware location recommendations. *IEEE Trans. Serv. Comput.* **2015**, *9*, 633–646. [[CrossRef](#)]
44. Zhang, J.D.; Chow, C.Y. iGSLR: Personalized geo-social location recommendation: A kernel density estimation approach. In Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Orlando, FL, USA, 5–8 November 2013; pp. 334–343.
45. Rendle, S.; Freudenthaler, C.; Schmidt-Thieme, L. Factorizing personalized markov chains for next-basket recommendation. In Proceedings of the 19th International Conference on World Wide Web, Raleigh, NC, USA, 26–30 April 2010; pp. 811–820.
46. Mathew, W.; Raposo, R.; Martins, B. Predicting future locations with hidden Markov models. In Proceedings of the 2012 ACM Conference on Ubiquitous Computing, Pittsburgh, PA, USA, 5–8 September 2012; pp. 911–918.
47. Chen, M.; Liu, Y.; Yu, X. NLPMM: A next location predictor with markov modeling. In Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining, Tainan, Taiwan, 13–16 May 2014; pp. 186–197.
48. Zhang, C.; Zhang, K.; Yuan, Q.; Zhang, L.; Hanratty, T.; Han, J. Gmove: Group-level mobility modeling using geo-tagged social media. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 1305–1314.
49. Natarajan, N.; Shin, D.; Dhillon, I.S. Which app will you use next: collaborative filtering with interactional context. In Proceedings of the 7th ACM Conference on Recommender Systems, Hong Kong, China, 12–16 October 2013; pp. 201–208.

50. Chen, J.; Wang, C.; Wang, J. A personalized interest-forgetting markov model for recommendations. Proceedings of Twenty-Ninth AAAI Conference on Artificial Intelligence, Austin, TX, USA, 25–30 January 2015.
51. Wang, H.; Shen, H.; Ouyang, W.; Cheng, X. Exploiting POI-Specific Geographical Influence for Point-of-Interest Recommendation. In Proceedings of the IJCAI, Stockholm, Sweden, 13–19 July 2018; pp. 3877–3883.
52. Feng, S.; Li, X.; Zeng, Y.; Cong, G.; Chee, Y.M.; Yuan, Q. Personalized ranking metric embedding for next new POI recommendation. In Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, Buenos Aires, Argentina, 25–31 July 2015.
53. Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G.S.; Dean, J. Distributed representations of words and phrases and their compositionality. In Proceedings of the Advances in Neural Information Processing Systems, Lake Tahoe, NV, USA, 5–10 December 2013; pp. 3111–3119.
54. Yang, C.; Bai, L.; Zhang, C.; Yuan, Q.; Han, J. 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, 13–17 August 2017; pp. 1245–1254.
55. Chen, Y.; Fan, R.; Yang, X.; Wang, J.; Latif, A. Extraction of urban water bodies from high-resolution remote-sensing imagery using deep learning. *Water* **2018**, *10*, 585. [[CrossRef](#)]
56. Wang, S.; Wang, Y.; Tang, J.; Shu, K.; Ranganath, S.; Liu, H. What your images reveal: Exploiting visual contents for point-of-interest recommendation. In Proceedings of the 26th International Conference on World Wide Web, Perth, Australia, 3–7 April 2017; pp. 391–400.
57. Chen, Y.; Fan, R.; Bilal, M.; Yang, X.; Wang, J.; Li, W. Multilevel cloud detection for high-resolution remote sensing imagery using multiple convolutional neural networks. *ISPRS Int. J. Geo Inf.* **2018**, *7*, 181. [[CrossRef](#)]
58. Ding, R.; Chen, Z. RecNet: A deep neural network for personalized POI recommendation in location-based social networks. *Int. J. Geogr. Inf. Sci.* **2018**, *32*, 1631–1648. [[CrossRef](#)]
59. Zhang, Y.; Dai, H.; Xu, C.; Feng, J.; Wang, T.; Bian, J.; Liu, T.Y. Sequential click prediction for sponsored search with recurrent neural networks. In Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, Québec, QC, Canada, 27–31 July 2014.
60. Zhao, P.; Zhu, H.; Liu, Y.; Li, Z.; Xu, J.; Sheng, V.S. Where to Go Next: A Spatio-temporal LSTM model for Next POI Recommendation. *arXiv* **2018**, arXiv:1806.06671.
61. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Polosukhin, I. Attention is all you need. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 5998–6008.
62. Xu, K.; Ba, J.; Kiros, R.; Cho, K.; Courville, A.; Salakhudinov, R.; Bengio, Y. Show, attend and tell: Neural image caption generation with visual attention. In Proceedings of the International Conference on Machine Learning, Lille, France, 6–11 July 2015; pp. 2048–2057.
63. Chen, J.; Zhang, H.; He, X.; Nie, L.; Liu, W.; Chua, T.S. Attentive collaborative filtering: Multimedia recommendation with item-and component-level attention. In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, Shinjuku, Tokyo, Japan, 7–11 August 2017; pp. 335–344.
64. Sutskever, I.; Vinyals, O.; Le, Q.V. Sequence to sequence learning with neural networks. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 8–13 December 2014; pp. 3104–3112.
65. Bahdanau, D.; Cho, K.; Bengio, Y. Neural machine translation by jointly learning to align and translate. *arXiv* **2014**, arXiv:1409.0473.
66. Feng, J.; Li, Y.; Zhang, C.; Sun, F.; Meng, F.; Guo, A.; Jin, D. Deepmove: Predicting human mobility with attentional recurrent networks. In Proceedings of the 2018 World Wide Web Conference on World Wide Web, Lyon, France, 23–27 April 2018; pp. 1459–1468.

