

Article

A Spatiotemporal Dilated Convolutional Generative Network for Point-Of-Interest Recommendation

Chunyang Liu ¹, Jiping Liu ^{2,*}, Shenghua Xu ², Jian Wang ³, Chao Liu ⁴, Tianyang Chen ⁵ and Tao Jiang ²

¹ School of Environment Science and Spatial Informatics, China University of Mining and Technology (CUMT), Xuzhou 221116, China; liuchunyang@cumt.edu.cn

² Chinese Academy of Surveying and Mapping, Beijing 100830, China; xushh@casm.ac.cn (S.X.); taojiang0402@gmail.com (T.J.)

³ School of Geomatics and Urban Spatial Informatics, Beijing University of Civil Engineering and Architecture, Beijing 100044, China; wangjian@bucea.edu.cn

⁴ School of Geodesy and Geomatics, Anhui University of Science and Technology, Huainan 232001, China; chliu1@aust.edu.cn

⁵ Department of Geography and Earth Science, University of North Carolina at Charlotte, Charlotte, NC 28223, USA; tchen19@uncc.edu

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

Received: 18 January 2020; Accepted: 18 February 2020; Published: 19 February 2020



Abstract: With the growing popularity of location-based social media applications, point-of-interest (POI) recommendation has become important in recent years. Several techniques, especially the collaborative filtering (CF), Markov chain (MC), and recurrent neural network (RNN) based methods, have been recently proposed for the POI recommendation service. However, CF-based methods and MC-based methods are ineffective to represent complicated interaction relations in the historical check-in sequences. Although recurrent neural networks (RNNs) and its variants have been successfully employed in POI recommendation, they depend on a hidden state of the entire past that cannot fully utilize parallel computation within a check-in sequence. To address these above limitations, we propose a spatiotemporal dilated convolutional generative network (ST-DCGN) for POI recommendation in this study. Firstly, inspired by the Google DeepMind' WaveNet model, we introduce a simple but very effective dilated convolutional generative network as a solution to POI recommendation, which can efficiently model the user's complicated short- and long-range check-in sequence by using a stack of dilated causal convolution layers and residual block structure. Then, we propose to acquire user's spatial preference by modeling continuous geographical distances, and to capture user's temporal preference by considering two types of time periodic patterns (i.e., hours in a day and days in a week). Moreover, we conducted an extensive performance evaluation using two large-scale real-world datasets, namely Foursquare and Instagram. Experimental results show that the proposed ST-DCGN model is well-suited for POI recommendation problems and can effectively learn dependencies in and between the check-in sequences. The proposed model attains state-of-the-art accuracy with less training time in the POI recommendation task.

Keywords: point-of-interest recommendation; dilated causal convolution; residual block; spatial preference; temporal preference

1. Introduction

During the past few years, with the rapid growth of mobile devices and location-based social networks (LBSNs) services, these services have attracted many users to share their locations and experiences with massive amounts of check-in data accumulated. The huge volume of check-in data

and contextual information brought opportunities for researching human mobility behavior in a large scale [1,2]. Point-of-interest (POI) recommendation plays an important role in LBSNs because it can predict users' preferences to provide users valuable suggestions and assist them to make adequate decisions in their daily routines and trip planning [3,4]. Figure 1 illustrates an example of POI recommendation, given all users' check-in sequences data; the task is to predict the POI of a user, who will visit at a specific time point, by mining user's location preferences and movement patterns. This task is meaningful and important, as it not only helps users discover interesting locations to increase their engagement with location-based services, but also creates the opportunities for LBSN service providers to increase their revenue through personalized advertising [5]. Therefore, the research on POI recommendation has attracted widespread attention from the academic and industrial fields [6–9].

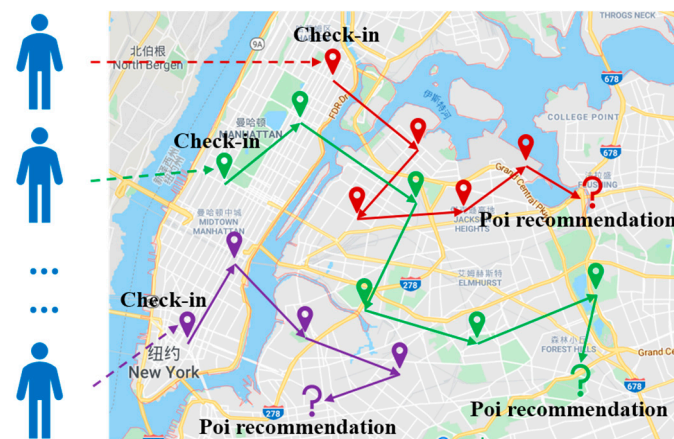


Figure 1. An example of point-of-interest recommendation.

Unlike items such as news, videos, and music in traditional context-free recommender systems, the user's history check-in data implies the interactions between a user and POIs in a physical world [10]. Thus, geospatial information, such as geographical distance, would have a significant effect on user's daily activities and check-in behaviors. For example, people prefer to go to nearby malls or gyms because such a decision is more time-efficient than attending similar places in a further distance. As per Tobler's First Law of Geography [11] that "Everything is related to everything else, but near things are more related than distant things", adjacent POIs are more geographically relevant than distant POIs. In the literature, spatial influence has been mostly modeled by utilizing the distance between two POIs; moreover, many existing studies have shown that there is a strong relationship between user's check-in activities and geographical distances [12,13]. Besides, temporal context and sequential relations are also crucial factors that affect human real-life check-in activities [7,14–16] due to the time sensitivity of the POI recommendation. For example, people would repeatedly go to the gym after work on weekdays, and they could also prefer to visit cinemas at night on weekends. This also reflects the periodic characteristics of users' check-in behaviors, e.g., different hours in a day or different days in a week. In addition, sequential relations of the check-in also need to be considered. For instance, most people may want to find a hotel instead of a gym after arriving at the airport. Therefore, how to effectively capture user's short- and long-range dependencies from a given check-in sequence is also an interesting problem to be investigation. However, how to accurately predict user's movement behavior preference according to complex spatiotemporal contextual features and sequential patterns is still a challenging issue.

The POI recommendation methods have been applied in the numerous studies, most of which are based on collaborative filtering (CF) [17] and Markov chains (MC) [18]. However, traditional user-based CF methods, item-based CF methods, and matrix factorization (MF)-based CF methods find it difficult to handle long-range sequences and incorporate various features effectively because they only learn linear or low-order interactions between features. Moreover, MC-based methods assume

strong independence among different components and only utilize the last POI when modeling check-in sequences. Recently, deep learning-based methods, especially RNNs-based methods, have been applied in POI recommendation and were assumed to be effective [10,13,19]. RNNs-based methods outperform other POI recommendation methods since they can learn long-range dependencies effectively. Moreover, some studies consider integrating spatiotemporal contextual information into RNN structure to enhance the performance of POI recommendation [20,21]. While RNNs and its variants have shown an impressive capability in modeling check-in sequences, these RNN-based methods depend on a hidden state of the entire past that cannot effectively utilize parallel computation within a check-in sequence and fully learn high-level interactions between features [22]. Consequently, these issues inevitably affect RNNs to further improve their performance when applying to POI recommendation.

To address the identified issues in existing studies, inspired by the WaveNet model [23], we propose a spatiotemporal dilated convolutional generative network, or ST-DCGN for short, as a solution to POI recommendation. The framework of the proposed method is depicted in Figure 2. This model not only considers modeling complex long-range sequential relations to acquire the user's sequential preference, but also modeling continuous geographic movement and temporal periodic patterns to acquire the user's personalized spatiotemporal preference. From our experiments, we observe that our model outperforms state-of-the-art algorithms on two publicly available datasets, namely Foursquare [24] and Instagram [25]. In conclusion, our contributions are summarized as follows:

- We proposed a novel POI recommendation framework based on WaveNet model, where the conditional generative model and dilated causal convolutions are used to enable much larger receptive fields and model complex long-range check-in sequence. The framework not only achieves higher recommendation performance, but also appears to have a lower level of model complexity compared to the identified state-of-the-art POI recommendation methods.
- Considering the importance of spatiotemporal contextual information, we acquire the user's personalized spatial preference by modeling continuous geographical distances, and capture the user's personalized temporal preference by modeling specific continuous time IDs, which integrated patterns in two time scales (e.g., hours in a day and days in a week).
- We conducted experiments to study the spatiotemporal characteristics of users' check-in behavior on two real-world datasets, and we compared ST-DCGN with seven baseline approaches of POI recommendation, and extensive experiments showed that ST-DCGN was effective and outperforms state-of-the-art methods significantly.

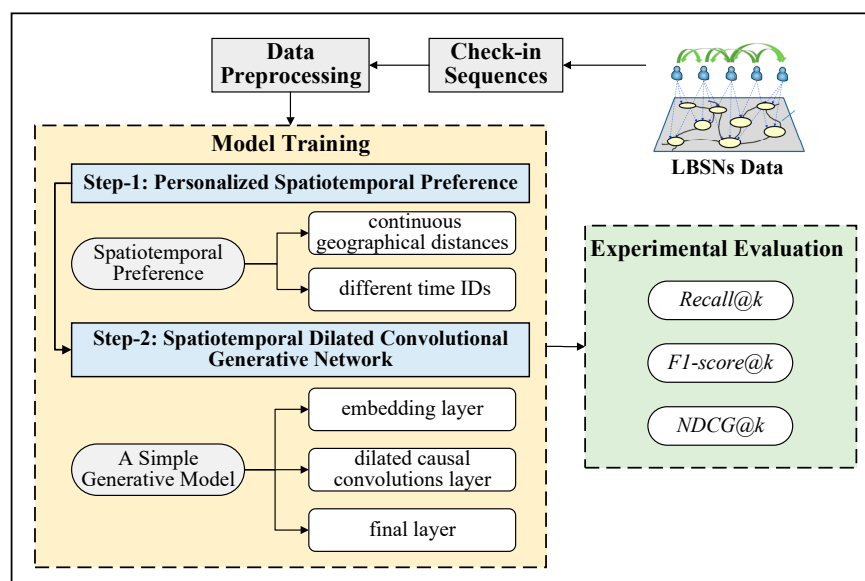


Figure 2. Framework of the proposed model for point-of-interest recommendation.

The rest of this paper is organized as follows: The existing related studies are briefly reviewed in Section 2. The details of our ST-DCGN method are delivered in Section 3. Experiments and results of the proposed method are illustrated in Section 4. Finally, conclusions and future work are drawn in Section 5.

2. Related Work

In this section, we review related work from two stream of methods, conventional and deep learning-based POI recommendation methods.

2.1. Conventional POI Recommendation Methods

POI recommendation has been widely investigated in the field of LBSNs. Most previous solutions learned user preference for POIs using CF-based methods. User-based CF and item-based CF techniques are widely exploited for POI recommendation [6,7]. For example, Ye et al. [6] firstly proposed user-based and item-based approaches for POI recommendation by using CF techniques, which assumed that similar users had similar tastes for locations and users were interested in similar POIs. Furthermore, other researchers employed the model-based CF technique such as MF for POI recommendation in LBSNs [5,8,17,26], which searched for potential location preferences of users by factorizing a user-POI matrix into two low rank matrices, each of which represented the latent factors of users or POIs.

Differing from traditional recommender systems, POI recommender systems need to consider geographical influence, temporal influence, sequential influence, or other characteristics (e.g., social relationship, reviews, categories, etc.) [8,21]. The geographical influence has been proven to be a significant factor in POI recommendation [13], where many existing studies mainly focus on integrating the geographical information due to the well-known strong correlation between users' activities and geographical distance. Existing methods of modeling geographical influence mainly use several types of spatial distribution functions, such as power law function, multi-center Gaussian distribution, or kernel density estimation model [17,26–28]. For example, Cheng et al. [17] explained that users always visited nearby POIs around several centers (i.e., the most popular POIs), thus they capture the geographical influence via modeling the probability of a user's check-in on a location as a multi-center Gaussian model (MGM). In addition, Zhang et al. [28] capture the personalized geographical influence by using a kernel density estimation approach. Lian et al. [26] proposed a GeoMF model to incorporate geographical information into MF, and used a two-dimensional kernel density estimation to characterize geographical influence over distance. The results of these works demonstrated the effectiveness of incorporating spatial context in POI recommendations.

Temporal influence has been proved effective for modeling users' check-in behavior by recent studies [5,7,14]. For example, Yuan et al. [7] argued that users' visiting preferences for some locations exhibited time periodicity. Thus, they split time into hourly based slots and proposed time-aware point-of-interest recommendation method. Gao et al. [14] proposed four temporal aggregation strategies to integrate a user's check-in preferences of different temporal states. Furthermore, some studies focus on the application of content information such as social information and other characteristics in LBSNs for POI recommendation as well. For example, Li et al. [29] presented a unified POI recommendation approach, which exploited geographical, social, and categorical associations between users and POIs. Yang et al. [30] considered both check-ins and comments of venues in location recommendation, and proposed a fusion framework to get a unified preference model from both check-ins and tips. However, most approaches fail to model complicated relations in the check-in sequence data.

In addition to traditional CF methods, sequential methods have been considered for POI recommendation and they mostly rely on Markov chains. Mathew et al. [31] proposed a hybrid approach based on hidden Markov models, which clusters location histories according to their characteristics, and later trains an HMM for each cluster. Cheng et al. [18] proposed a matrix factorization model, namely FPMC-LR, to include both personalized Markov chain and localized regions solving the POI

recommendation task. However, the underlying strong Markov assumption of these methods has difficulty in constructing more effective relationship among different components.

2.2. Deep Learning-Based POI Recommendation Methods

Deep learning, developed in computer science, has been widely applied in many research fields, such as computer vision [32,33], natural language processing [34,35], and speech recognition [36,37]. Also, many deep learning techniques have recently been applied to POI recommendation systems, which may change the architectures of traditional recommendation and brings new opportunities to improve the recommended accuracy [38]. For example, a few previous works utilized Word2vec [39] to model human mobility behavior [40,41].

Recently, RNNs-based methods have gained remarkable attention and become more powerful in modeling user's sequential history and transition. For example, Liu et al. [19] firstly brought RNN to next location prediction, where they employed a temporal and spatial recurrent neural network (ST-RNN) to model local temporal and spatial contexts in each layer with time-specific transition matrices for different time intervals and distance-specific transition matrices for different geographical distances. Kong et al. [42] built a hierarchical spatial-temporal long-short term memory (HST-LSTM) model, which naturally combined spatial-temporal influence into LSTM to mitigate the problem of data sparsity. Zhao et al. [20] proposed a ST-LSTM network for the next POI recommendation, which modeled spatiotemporal intervals between check-ins under LSTM architecture to learn user's visiting behavior. Cui et al. [13] proposed a Distance2Pre network for the next POI prediction, and it can mine spatial preference to model the correlation of the user distance. Moreover, some researchers have integrated attention models into RNNs and achieved better performance. For example, Huang et al. [10] developed an attention-based spatiotemporal LSTM (ATST-LSTM) network for the next POI recommendation, which considered the relevant historical check-in records in a check-in sequence selectively using the spatiotemporal contextual information. Feng et al. [43] proposed an attentional mobility model, namely DeepMove, which predicted human mobility from lengthy and sparse trajectories. However, the above RNNs-based methods depend on a hidden state of the entire past that cannot effectively utilize parallel computing within a check-in sequence. This also results in a speed limit on the model's training and evaluation process [22].

By contrast, the structure using convolutional neural network (CNN) does not depend on the calculation of each time step in the sequence history, but little work exists for POI recommendation by using CNN structure. Wang et al. [44] proposed a novel CNN-based visual content enhanced POI recommendation (VPOI), which incorporated visual contents into a probabilistic model for learning user and POI latent features, but they only used CNN framework when extracting features from images. Furthermore, Tang et al. [45] proposed a convolutional sequence embedding recommendation model by modeling recent actions as an "image" among time, latent dimensions, and learning sequential patterns using convolutional filters. It abandoned RNN structures and demonstrated that this CNN-based recommender can achieve superior performance to the popular RNN model in the Top-N sequential recommendation task. Yuan et al. [22] proposed a simple, efficient, and highly effective convolutional generative network for next-item recommendation, which was capable of learning high-level representation from both short- and long-range item dependencies. However, the above two sequence recommendation methods do not consider the spatiotemporal contextual information, and they are not specialized solutions to POI recommendations. Unlike existing studies, our work considers geographical influence and temporal influence in a personalized way into a spatiotemporal dilated convolutional generative network to capture user's sequential preference and spatiotemporal preference.

3. Proposed Method

In this section, we firstly addressed the identified problem of POI recommendation and then described our approach to obtain personalized spatiotemporal preference and components of ST-DCGN,

which included personalized spatiotemporal preference processing, a simple generative model under spatiotemporal conditions, an embedding layer, dilated causal convolution layers, and a final layer.

3.1. Problem Formulation

Let $\mathbf{U} = \{u_1, u_2, \dots, u_m\}$ and $\mathbf{X} = \{x_1, x_2, \dots, x_n\}$ be the sets of m users and n POIs, respectively. Each POI has a unique identifier and geographical coordinates, which include geographical latitude and longitude. For user u , a check-in sequence that represents that user's history check-ins are arranged in chronological order, denoted by $\mathbf{X}^u = \{x_1^u, x_2^u, \dots, x_T^u\}$. Given each user's check-in sequence \mathbf{X}^u , the goal of POI recommendation is to predict the most likely POI x_{T+1} that the user u will visit at next time point $T + 1$.

3.2. Personalized Spatiotemporal Preference

In this part, we model check-in sequences and capture personalized spatiotemporal preference by considering geographical influences and temporal periodic patterns. Recent studies show that continuous geographic movement and temporal periodic patterns are important for POI recommendations [10,13,16,19].

3.2.1. Personalized Spatial Preference

Previous works show that power law distribution and multi-center Gaussian distribution can represent the geographical information by using the users' overall historical check-in record [7,17]. Although they reflect geographical differences of user's check-in behavior, they ignore the user's personalized differences in check-in behavior. In order to better model the user's personalized check-in behavior, we use geographical distances of continuous user's check-in to model the personalized spatial preference. More specifically, we calculate the distances between two successive POIs that all users' check-in and map these distances to discrete bins, for example, as shown in Figure 3, where Δs_1 is mapped to the interval Δd to $2\Delta d$, and Δs_3 is mapped to the interval $2\Delta d$ to $3\Delta d$, so every other distance value can be similarly mapped to a specific interval. In our scheme, we need to define one value Δd to represent the interval of discrete bins, as for the effects of parameter settings, we will discuss them in the experiments.

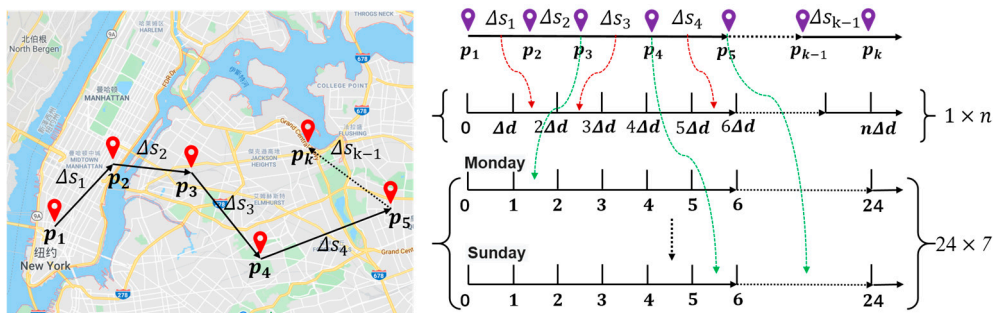


Figure 3. User's check-in behavior and its spatiotemporal context division.

We transform each user's check-in sequence $\mathbf{X}^u = \{x_1^u, x_2^u, \dots, x_T^u\}$ into a fixed-length sequence $\mathbf{E}_X^u = \{x_1^u, x_2^u, \dots, x_k^u\}$, where k represents the maximum length that we consider. If the sequence length was greater than k , we would only consider the most recent k check-in records. If the sequence length was less than k , we would add padding items to the left until the length became k . Therefore, we can further obtain fixed-length continuous geographic distance sequences $\mathbf{E}_S^u = \{r_1^u, r_2^u, \dots, r_k^u\}$, and the continuous geographic distance matrix for all m users is provided as follows.

$$E_S = \begin{bmatrix} r_1^1 & r_2^1 & \dots & r_k^1 \\ r_1^2 & r_2^2 & \dots & r_k^2 \\ \dots & \dots & \dots & \dots \\ r_1^m & r_2^m & \dots & r_k^m \end{bmatrix} \quad (1)$$

3.2.2. Personalized Temporal Preference

Previous works have shown that users' check-in behavior exhibits periodic characteristics [7,16]. For example, users tend to check in around the gym from 18:00 to 20:00 on Tuesday and Thursday evenings, but prefer to go to the market for shopping on Saturday from 15:00 to 17:00. Therefore, we can divide the time periodic pattern into two scales: Different hours in a day and different days in a week. To capture two periodic patterns of users' check-in behaviors, we introduce a two-slice time indexing scheme [16]. As shown in Figure 3, we firstly obtain the timestamp sequence $\{T_1^u, T_2^u, \dots, T_k^u\}$ corresponding to the user's check-in sequence $\{x_1^u, x_2^u, \dots, x_k^u\}$, and then divide each timestamp T_i^u into the specific time interval of a week and a day. To be specific, a timestamp is divided into two slices in terms of day of week, and hour slot. Furthermore, we split a week into seven days (i.e., Sunday to Saturday) and a day into 24 h (i.e., 1 to 24). Then, we use 3 bits to denote the day in one week and 5 bits to define the hour in one day. Finally, we convert the binary code into a unique decimal digit as the time ID. In this time indexing scheme, we can obtain $T=7 \times 24 = 168$ time slices. Figure 4 demonstrates the procedure of encoding an exemplary time stamp, "2016-08-29 23:29:12". Therefore, we can further obtain fixed-length continuous time ID sequences $E_T^u = \{t_1^u, t_2^u, \dots, t_k^u\}$, and the continuous time ID matrix for all m users are provided as follows.

$$E_T = \begin{bmatrix} t_1^1 & t_2^1 & \dots & t_k^1 \\ t_1^2 & t_2^2 & \dots & t_k^2 \\ \dots & \dots & \dots & \dots \\ t_1^m & t_2^m & \dots & t_k^m \end{bmatrix} \quad (2)$$

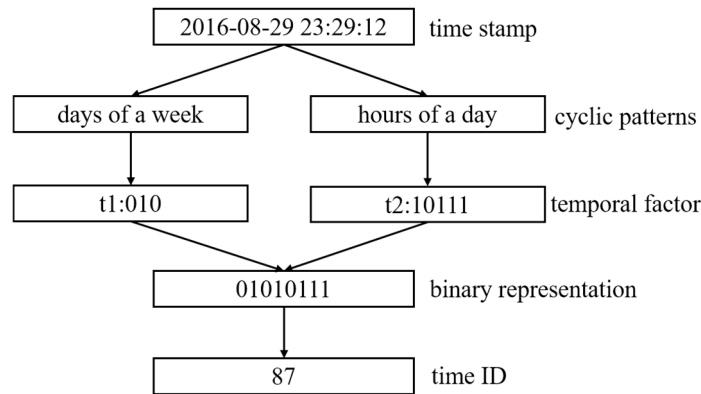


Figure 4. Time indexing scheme demonstration.

3.3. A Generative Model under Spatiotemporal Conditions

In this section, we introduce a novel generative model that is operated directly on the user's check-in sequence. The solution proposed here is inspired by the idea of WaveNet [23], a generative model for raw audio based on the PixelCNN [46] architecture. WaveNet provides a generic and flexible framework for tackling many applications that rely on audio generation (e.g., text-to-speech, music, speech enhancement, voice conversion, source separation). Similarly, we consider a user's history check-in sequence $E_X^u = \{x_1^u, x_2^u, \dots, x_k^u\}$, given a model with parameter θ . We aim to output the next value \hat{x}_{k+1}^u conditional on the check-in sequence history. Let $p(E_X^u | \theta)$ be the joint probability of check-in

sequence $\{x_1^u, x_2^u, \dots, x_k^u\}$; moreover, we can factorize $p(E_X^u|\theta)$ as a product of conditional probabilities by chain rule as follows:

$$p(E_X^u|\theta) = \prod_{i=1}^k p(x_{i+1}^u|x_1^u, \dots, x_i^u, \theta) \quad (3)$$

where the POI sample x_{k+1}^u is therefore conditioned on the samples of all the previous POIs $\{x_1^u, x_2^u, \dots, x_k^u\}$.

As mentioned, we considered the spatial and temporal contextual information in the POI recommendation. Therefore, we also consider continuous geographic distance sequences $E_S^u = \{r_1^u, r_2^u, \dots, r_k^u\}$ and continuous time ID sequences $E_T^u = \{t_1^u, t_2^u, \dots, t_k^u\}$ as conditional inputs, when predicting the user's check-in sequence $E_X^u = \{x_1^u, x_2^u, \dots, x_k^u\}$. Further, we can model the conditional distribution $p(E_X^u|\theta)$ of the check-in sequence given these inputs. Equation (3) now becomes

$$p(E_X^u|\theta) = \prod_{i=1}^k p(x_{i+1}^u|x_1^u, \dots, x_i^u, r_1^u, r_2^u, \dots, r_i^u, t_1^u, t_2^u, \dots, t_i^u, \theta) \quad (4)$$

where the conditional probability distribution is modelled by using stacked layers of dilated convolutions, which we will describe later.

3.4. Embedding Look-Up Layer

Given a user's continuous check-in sequence, the model retrieves each of the first k POIs $E_X^u = \{x_1^u, x_2^u, \dots, x_k^u\}$ via a look-up table, and stacks these POI embeddings together. Similarly, we deal with the user's continuous geographic distance sequences $E_S^u = \{r_1^u, r_2^u, \dots, r_k^u\}$ and time ID sequence $E_T^u = \{t_1^u, t_2^u, \dots, t_k^u\}$ simultaneously. Assuming the embedding dimension is $2d$, where d can be set as the number of inner channels in the convolutional network, we create three embedding matrices $E_X^u \in \mathbb{R}^{k \times 2d}$, $E_S^u \in \mathbb{R}^{k \times 2d}$, and $E_T^u \in \mathbb{R}^{k \times 2d}$ for POIs, geographic distances, and time IDs, respectively. Inspired by previous work [22], our proposed method will learn the embedding layer through one-dimensional convolution filters. To be specific, the 2D matrix (i.e., E_X^u , E_S^u and E_T^u) is reshaped from $k \times 2d$ to a $1 \times k \times 2d$ three-dimensional tensor. Figure 5 illustrates the reshaping process.

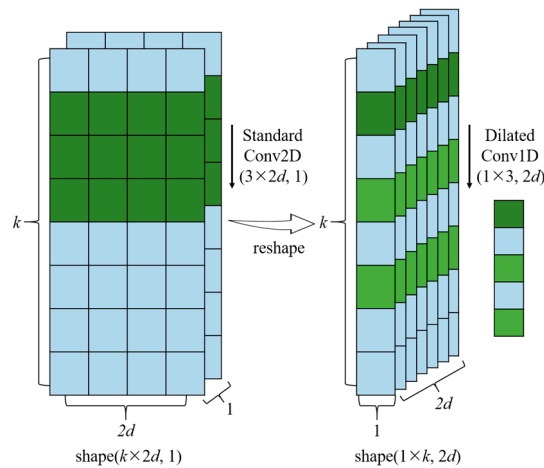


Figure 5. Transformation from the standard 2D filter to the 1D 2-dilated filter.

3.5. Dilated Causal Convolutions Layer

There are several obvious drawbacks of traditional convolution operation process for processing sequence prediction problems, e.g., (1) some sequential information will be lost during the pooling process; (2) a simple standard causal convolution is only able to increase the receptive field with

size linear in the depth of the network. This makes it challenging to handle long-range dependence of check-in history sequence, as shown in Figure 6. Therefore, inspired by early work on speech modeling [23], our solution here is to construct the proposed generative model by using dilated causal convolution algorithm enabling an exponentially large receptive field. Figure 7 depicts a dilated causal convolution with filter size $g = 3$ and dilation factors $l = 1, 2, 4, 8$. We can see that a dilated convolution is a convolution where the filter is applied over an area larger than its length by skipping input values with a certain step. It is equivalent to a convolution with a larger filter derived from the original filter by dilating it with zeros, but is significantly more efficient since it utilizes fewer parameters. Thus, the dilated convolutional operation can better handle long-term users' check-in sequences without using more network layers.

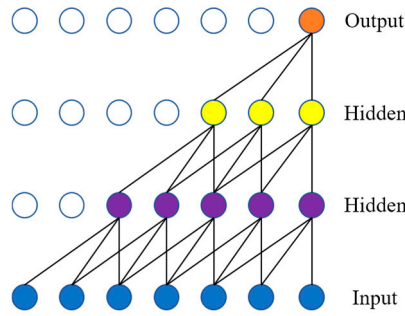


Figure 6. A figure showing a stack of causal convolutional layers.

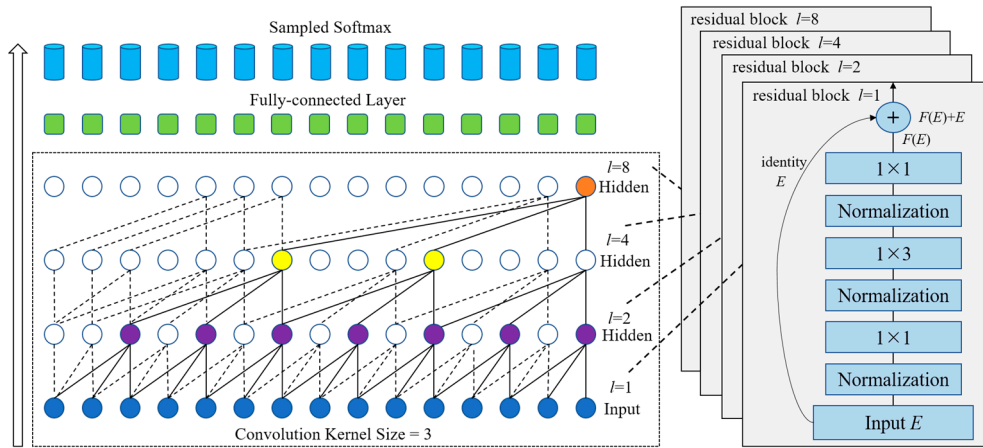


Figure 7. The proposed generative architecture with dilated causal convolutional network.

In addition, at training time, the conditional probabilities for all timesteps can be calculated in parallel because all timesteps of check-in sequences are known. Note that unlike RNN-based models that depend on a hidden state of the entire check-in history, it cannot fully utilize a parallel mechanism. As a result, the computing advantage of CNN models are more preferred by POI recommendation systems.

More formally, given a one-dimensional sequence input $X \in \mathbb{R}^k$ and a filter $f : \{0, 1, \dots, g-1\} \rightarrow \mathbb{R}$, the one-dimensional dilated convolution F on element s of the sequence is defined as

$$F(s) = (X *_l f)(s) = \sum_{i=0}^{g-1} f(i) \cdot X_{s-l \cdot i} \quad (5)$$

where f is the filter function, g is the filter size, l is the dilation factor, and $s - l \cdot i$ accounts for the direction of the past. Clearly, dilated causal convolution algorithm can better capture long-term

check-in sequence dependencies without using more network layers and larger filters. In practice, to further increase the receptive fields and model capacity, we just need to repeat the dilated convolution structure in Figure 7 by stacking (e.g., 1, 2, 4, 8, 1, 2, 4, 8).

As discussed in [22], in order to learn higher-level feature representations from long-range sequence dependencies, an intuitive method is to increase the number of layers in our network. However, in practice, it also easily results in the degradation problem, which makes the training process much harder. To solve this problem, we introduce residual connections [33,47] in our method. As shown in Figures 7 and 8b, a residual block contains two branches. One branch is to convert the input layer E to F through a series of network layers, including the dilated causal convolution with the layer-normalization [48], activation (e.g., $ReLU$ [49]), and 1×1 convolutional in a specific order. The other branch is a direct projection of the input E . The residual mapping $F(E)$ can be computed as follows:

$$E_1 = W_1(ReLU(\phi(E))) + b_1 \quad (6)$$

$$E_2 = W_2(ReLU(\phi(E_1))) + b_2 \quad (7)$$

$$F(E) = W_3(ReLU(\phi(E_2))) + b_3 \quad (8)$$

where ϕ denotes the layer-normalization. W_1, W_2, W_3, b_1, b_2 , and b_3 are a set of weights and biases for the residual block. Specifically, W_2 denotes the dilated causal convolution weight function with filter size $g = 3$ and dilation factors $l = 1, 2, 4, 8$. W_1 and W_3 denote standard 1×1 convolution weight function.

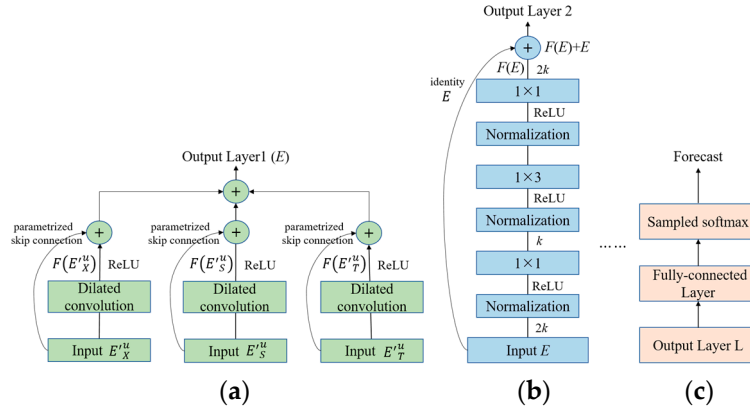


Figure 8. Fusion input layer (a), dilated residual blocks (b), and final output layer (c).

The desired mapping is now recast into $F(E) + E$ by element-wise addition. This effectively allows layers to learn modifications to the identity mapping rather than the entire transformation, which has been proven beneficial in deeper networks by previous literature [22,33,47]. In our framework, we capture the geographical influences and temporal periodic patterns by modeling specific spatiotemporal information. Therefore, we need to integrate continuous geographic distance sequence and specific time ID sequence into our network. As shown in Figure 8a, the check-in sequence input E_X^u and specific spatiotemporal conditions (i.e., E_S^u and E_T^u) are fused through the dilated causal convolutional and summed with the parametrized skip connections in the first layer. The result of the first layer is the input in the subsequent dilated convolution layer with a residual connection from the input to the output of the convolution (see Figure 8b). Instead of the standard residual connection, we use parametrized skip connection in the first layer, dynamically adjusting the weight parameters to ensure our model correctly extracting the necessary relations between the forecast and both the check-in sequence input and specific spatiotemporal conditions. The conditioning on the continuous geographic distance sequence E_S^u and specific time ID sequence E_T^u are done by computing the activation function of the convolution in the first layer as:

$$E = \text{ReLU}(w_x *_l E_X^u + b_x) + \text{ReLU}(w_r *_l E_S^u + b_r) + \text{ReLU}(w_t *_l E_T^u + b_t) \quad (9)$$

where w_x , w_r , and w_t are learnable convolution filter, $*_l$ denotes a convolution operator, and E denotes the result of multivariate sequence fusion.

3.6. Final Layer and Network Training

We have already mentioned the matrix in the last layer of the dilated causal convolution architecture has the same dimensional size of the input embedding E (i.e., $E \in \mathbb{R}^{k \times 2d}$), but the result we need should be a probability distribution that includes all POIs in the output sequence, where the probability distribution is the desired one that generates top- k POI recommendation list. In such a view, we use a fully connected layer with weight matrix $W^g \in \mathbb{R}^{2d \times n}$. As mentioned, we aim to maximize the conditional likelihood (equation 4). Clearly, maximizing $\log p(E_X^u | \theta)$ is mathematically equivalent to minimizing the sum of the binary cross-entropy loss for each item in $\{x_1^u, x_2^u, \dots, x_k^u\}$. Furthermore, we use negative sampling strategy (e.g., sampled softmax [50]) to avoid the calculation of the full softmax distributions for network training.

4. Experimental Results and Analysis

In this section, extensive experiments are conducted to compare our proposed ST-DCGN model with several state-of-the-art POI recommendation approaches. Firstly, two publicly accessible datasets are described and analyzed in detail. Then, baseline methods and evaluation metrics are introduced. Finally, experimental results are fully demonstrated, which include the recommendation performance and influence of hyper-parameters. In summary, our work attempts to answer the following research questions:

RQ1: Can our proposed method perform better than state-of-the-art baselines in accuracy for POI recommendation tasks?

RQ2: Does ST-DCGN outperform other deep neural networks (i.e., GRU, Distance2Pre, ST-RNN) in efficiency for POI recommendation tasks?

RQ3: How do the parameters affect our model performance, such as the embedding size, spatial windows widths, and sequence length?

4.1. Datasets Description and Analysis

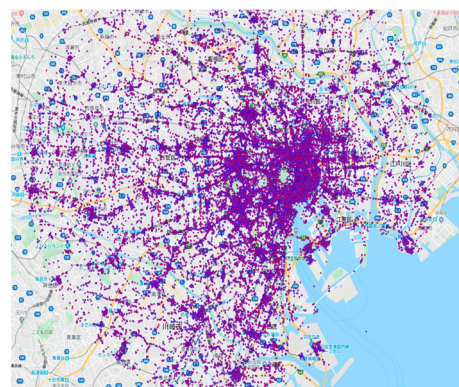
Our experiments were conducted on the two publicly accessible LBSNs check-in datasets. The first one is the Foursquare check-ins, which were collected in Tokyo City from April 2012 to February 2013 [24]. The second one is the Instagram check-ins, which were collected in New York City from June 2011 to November 2016 [25]. Both the two datasets provide sufficient richness of user check-ins. Each check-in contains user ID, POI ID, and timestamp. For both two datasets, we removed POIs checked in by less than five users and users who have checked in fewer than five POIs to reduce noise and alleviate data sparsity problems. Furthermore, we also removed check-in data without time stamps in the original Instagram dataset and extracted data from October 2015 to September 2016 as our experimental dataset. After pre-processing, statistics of the two datasets are shown in Table 1. Similar to some previous work [12,18], we further analyzed the geographic influence and temporal periodic patterns of the two datasets.

Figure 9 presents all users' check-in distribution in the two datasets, and we can find that the check-in distributions in the two datasets were significantly different. More specifically, for both datasets, the check-in distribution of users was concentrated in some hot areas, but Foursquare check-in distribution was more scattered than Instagram, which may be due to the different distribution of hot spots. This phenomenon further revealed the spatial patterns across different cities. Moreover, we further investigated the geographical influence on users' successive check-in behavior. In order to more intuitively explain the impact of geographical distance in users' check-in behaviors, we calculated the cumulative distribution function (CDF) of geographical distance between any two check-ins and

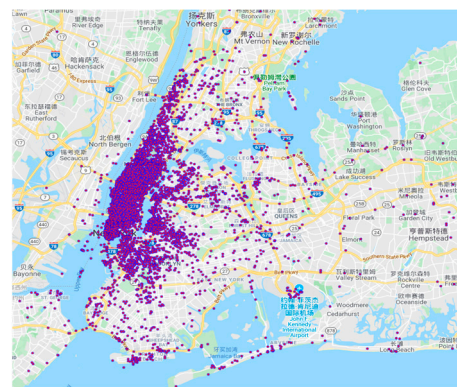
two consecutive check-ins of the same user in the Foursquare and Instagram datasets, respectively, as shown in Figure 10a,b. The results in Figure 10a indicate that users' check-in behaviors have highly geographic relevance since both the CDF curves for the two datasets increase fast when the distance is small. Specifically, this phenomenon is more apparent in Figure 10b because it considers the user's two consecutive check-ins. The above analysis suggests that it is necessary to consider the distance effect of continuous check-in behaviors in the POI recommendation algorithm. Thus, we attempted to utilize continuous geographical distance to capture user's personalized spatial preferences and movement patterns.

Table 1. Basic statistics of Foursquare and Instagram dataset.

Statistics	Foursquare	Instagram
#Users	2293	16,889
#POIs	6870	3961
#Check-ins	385,914	278,735
Avg. #check-ins per user	168.3	16.5
Avg. #visited POIs per user	56.2	70.4
sparsity	97.550%	99.583%
Time span	April 2012–February 2013	October 2015–September 2016

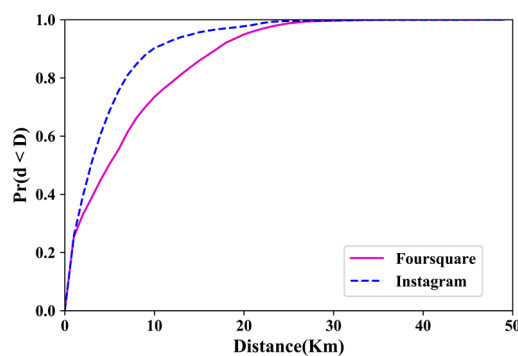


(a) Foursquare (Tokyo)

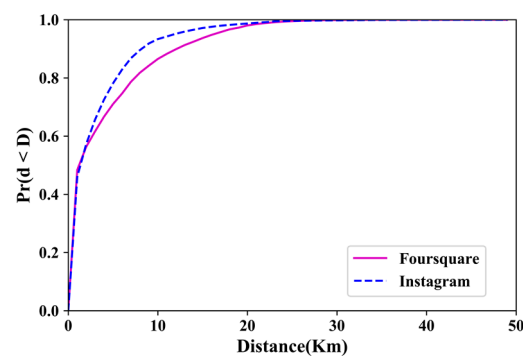


(b) Instagram (New York)

Figure 9. Distribution of all users' check-in in the two datasets.



(a)



(b)

Figure 10. Geographical neighbor influence in users' check-in behaviors. (a) Cumulative distribution function of geographical distance between users' any two check-ins; (b) cumulative distribution function of geographical distance between users' consecutive check-ins.

We further explored two temporal periodic patterns of users' check-in behaviors. More specifically, for the two datasets, we compared users' check-in probabilities at different time in a day and different

days in a week by calculating the check-in frequencies in the corresponding time slots, respectively, as shown in Figure 11. Based on the results in Figure 11, we found that the two datasets exhibited different temporal patterns, and different living habits in different regions. More specifically, for the Foursquare dataset, Figure 11a shows that check-ins on weekdays were mainly concentrated between 8:00–9:00 and 19:00–20:00, while the weekends were mainly concentrated on 17:00–18:00, which also reflects the periodic characteristics of users' check-in behavior. For the Instagram dataset, the difference in check-in time pattern was relatively small on weekdays and weekends but there were still differences in the check-in patterns at different time periods. In summary, there are significant time periodic characteristics of user's check-in behavior. Therefore, we attempted to use specific time ID coding to capture the users' personalized temporal preferences and periodic patterns.

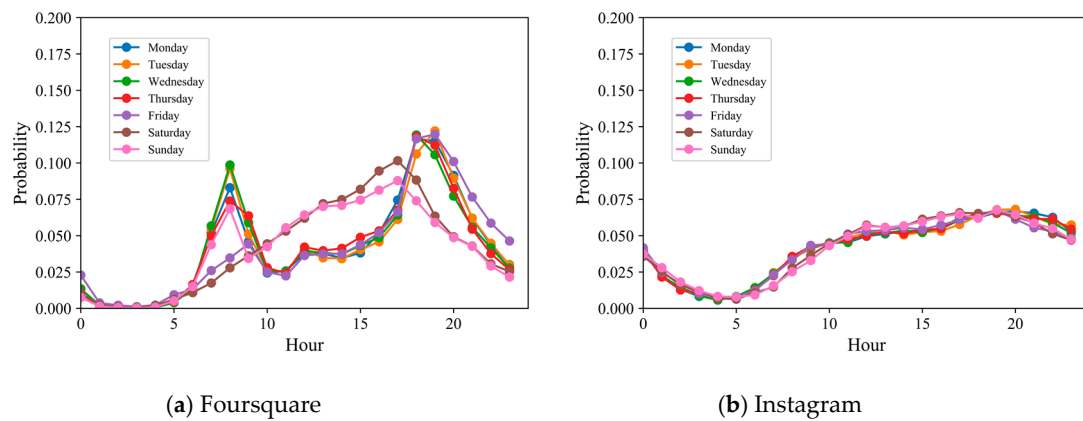


Figure 11. Periodic pattern in users' check-in behaviors.

4.2. Baseline Approaches

To evaluate the effectiveness of our proposed method, we compared ST-DCGN with the following representative baseline approaches for POI recommendation.

- **Bayesian Personalized Ranking (BPR):** This work presents the generic optimization criterion BPR-OPT derived from the maximum posterior estimator for optimal personalized ranking [51]. BPR is a classic baseline method for general POI recommendation.
- **GRU:** RNN is effective for POI recommendation task, and we applied an extension of RNN called GRU for capturing the long-term dependency [52].
- **FPMC-LR:** A state-of-the-art Markov chain method for POI recommendation. This method is designed based on first-order Markov chain and uses neighbors as negative samples [18].
- **PRME-G:** A state-of-the-art metric embedding method for POI recommendation, and the spatial distance is considered as the weight [12].
- **Caser:** A state-of-the-art standard 2D CNN-based method for personalized top-N sequential recommendation [45], and we applied Caser in POI recommendation.
- **Distance2Pre:** A state-of-the-art GRU-based model for POI prediction, which acquires the spatial preference by modeling distances between successive POIs [13].
- **ST-RNN:** A state-of-the-art RNN-based model for POI recommendation [19], which incorporates both local temporal and spatial transition context.

4.3. Evaluation Metrics and Experiment Setup

To our best knowledge, Recall@ k , F1-score@ k , and NDCG@ k (denoted by R@ k , F1@ k , and NDCG@ k , respectively) are three popular top- k metrics used for evaluating POI recommendation results, such as [2,8,13,19]. In this study, the three metrics are formulated as follows:

$$R@k = \frac{1}{N} \sum_{u=1}^N \frac{|R_u(k) \cap T_u|}{|T_u|} \quad (10)$$

$$F1@k = \frac{1}{N} \sum_{u=1}^N \frac{2 \cdot (|R_u(k) \cap T_u|/k) \cdot (|R_u(k) \cap T_u|/|T_u|)}{(|R_u(k) \cap T_u|/k) + (|R_u(k) \cap T_u|/|T_u|)} \quad (11)$$

$$NDCG@k = \frac{1}{N} \sum_{u=1}^N \frac{1}{Y_u} \sum_{n=1}^k \frac{2^{rel_n} - 1}{\log_2(k+1)} \quad (12)$$

where k indicates the number of POIs recommended to the user. We report R@ k , F1@ k , and NDCG@ k with $k = 5, 10$, and 20 in our experiments. $R_u(k)$ indicates the Top- k list recommended to the user. T_u represents the number of POIs the user visited. rel_n indicates the relevance of the n th POI to the user. Y_u represents the maximum DCG value of user u .

Additionally, all experiments were implemented through Python 3.5 and TensorFlow on one graphic processing unit (GPU), NVIDIA GeForce RTX 2080Ti. For the Foursquare dataset, the learning rate and batch size were set as 0.001 and 30, respectively. For the Instagram dataset, the learning rate and batch size were set as 0.001 and 40, respectively. Inspired by previous studies [13,21,22], we evaluated the POI recommendation results by using the leave-one-out evaluation. More specifically, we used the last (i.e., next) POI of each check-in sequence as the test data and the remaining POI as the training data. Furthermore, all baseline methods were reimplemented in the two datasets mentioned, and the relevant parameters were set according to the optimal configuration in the original paper.

4.4. Recommendation Performance

The performances of our proposed model ST-DCGN and six baselines on the Foursquare and the Instagram datasets evaluated by R@ k , F1@ k and NDCG@ k are shown in Figures 12 and 13, respectively (RQ1). We listed several findings as follows: (1) It is obvious that that our proposed ST-DCGN outperformed all identified baseline approaches on the Foursquare and Instagram datasets, showing ST-DCGN is effective for POI recommendation task. (2) Both BPR and GRU dropped behind other methods as they only model user–POI interactions without considering any contextual information to model users' check-in behavior. Furthermore, it is worthy to note that GRU did not always achieve better performance than BPR, especially on the Foursquare dataset. This result indicates that a good neural network architecture (i.e., RNN cell) is not enough to obtain excellent accuracy in the POI recommendation task, so we should consider more spatial and temporal contexts. (3) In comparison to BPR and GRU, FPMC-LR and PRME-G incorporated geographical and sequential information, and they took advantage of different ranking-based optimization strategies. Therefore, their performance on the two datasets were obviously better, indicating that modeling spatial contexts is indeed useful for POI recommendation. (4) Caser obtained much better performance than GRU, and this result demonstrates the advantage of using CNN architecture. Although Caser does not integrate any spatiotemporal context information, it still outperforms FPMC-LR, since FPMC-LR only modeled the first-order Markov chain while Caser captured high-order relations. (5) Distance2Pre had obviously better performance than FPMC-LR and PRME-G due to its capability in modeling user's sequential preference and spatial preference using RNN architecture. ST-RNN achieved further improvement by incorporating temporal contextual information. These great improvements indicate that neural network with spatiotemporal contextual information can obtain very promising performance in the POI recommendation task. (6) We firstly observed that ST-DCGN greatly outperformed Distance2Pre and ST-RNN on both datasets.

Compared with ST-RNN, the ST-DCGN improved $R@5$, $R@10$, $R@20$, $NDCG@5$, $NDCG@10$, and $NDCG@20$ by 14.62%, 12.13%, 7.95%, 17.92%, 15.50%, and 13.39%, respectively, on the Foursquare dataset. Also, for the Instagram dataset, the performance improvements in the evaluation metrics were 11.03%, 22.29%, 26.91%, 7.18%, 14.40%, and 16.12%, respectively.

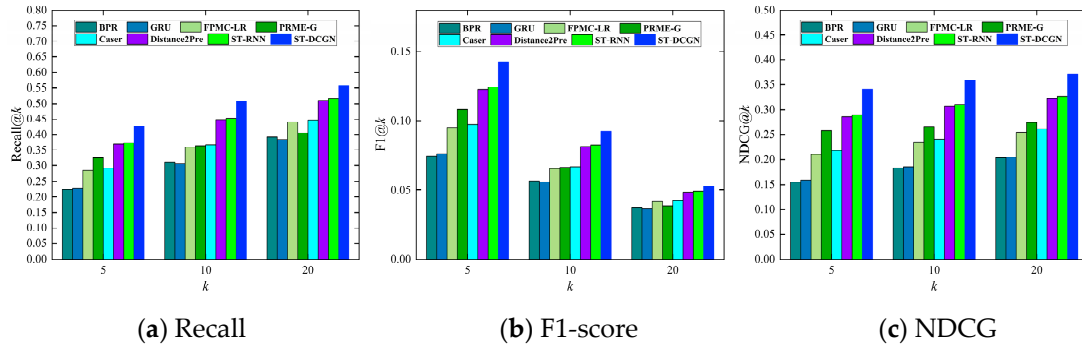


Figure 12. Performance comparison with state-of-the-art approaches on Foursquare.

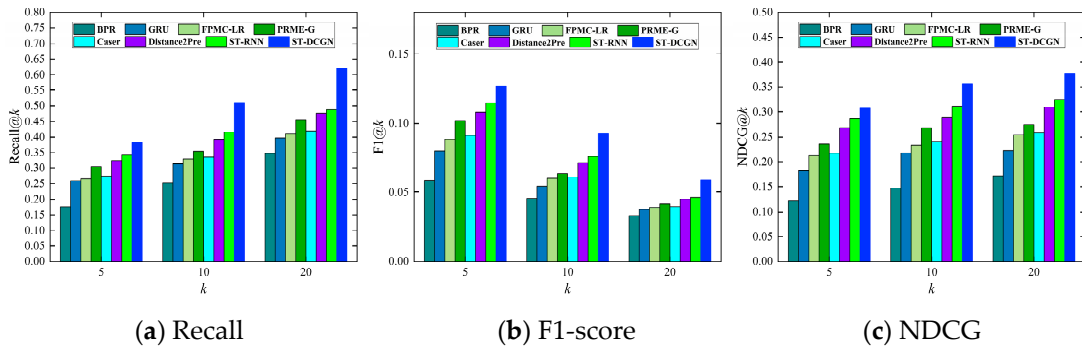


Figure 13. Performance comparison with state-of-the-art approaches on Instagram.

In addition to verifying the accuracy of our proposed model, we also evaluated the efficiency of ST-DCGN in Table 2 (RQ2). It is clear that our proposed ST-DCGN required less training time than other neural network models (i.e., GRU, Distance2Pre, ST-RNN). The reason is that CNN-based methods can effectively save training time through the full parallel mechanism of convolutions. For example, we can adopt parallelism when calculating the product of conditional probabilities. It is worth noting that although Caser achieved higher efficiency by using CNN structure and parallel computing compared with RNN-based methods, ST-DCGN achieved further improvements in training time compared with Caser, confirming the advantage of considering using dilated convolutional generative network.

Table 2. Overall training time (hours).

	GRU	Caser	Distance2Pre	ST-RNN	ST-DCGN
Foursquare	1.595	1.227	2.309	2.958	1.157
Instagram	2.116	1.892	4.236	5.156	1.793

In summary, ST-DCGN improved over the best baseline approaches on the two datasets with respect to the three metrics. On one hand, our model took advantage of 1D dilated causal convolutions network and residual learning to increase the receptive fields and enable training of much deeper networks, which greatly enhances the modeling of user's long-term dependency and short-term interest. Moreover, such a CNN-based network structure can fully utilize parallel computation to improve training efficiency. On the other hand, ST-DCGN took advantage of the personalized spatiotemporal information, and it can effectively acquire the user's spatial preference and temporal preference.

4.5. Sensitive Analysis of Parameters

In this part, we explored the effects of several key hyper-parameters on the performance of ST-DCGN. Here, we focused on analyzing the impacts of embedding size, spatial window widths, and sequence length (RQ3). Experiments were conducted on both the Foursquare and Instagram dataset.

Figure 14 presents the effects of embedding size on the performance. We analyzed the performance of the proposed ST-DCGN model on both datasets with different embedding sizes (i.e., 20, 40, 60, 80, 100, and 120) and use R@5 and R@10 as the measure metrics. It is apparent from this figure that the performance of ST-DCGN gradually increased with the embedding sizes, because high dimension representation can learn more latent features and capture more complex interactions. We notice that the performance of our model became robust when the embedding size reached 60 and 80 on the Foursquare and Instagram datasets, respectively. However, a larger embedding size may result in model performance degradation due to overfitting. Therefore, we chose the embedding size $2d = 60$ for the Foursquare dataset and $2d = 80$ for the Instagram dataset.

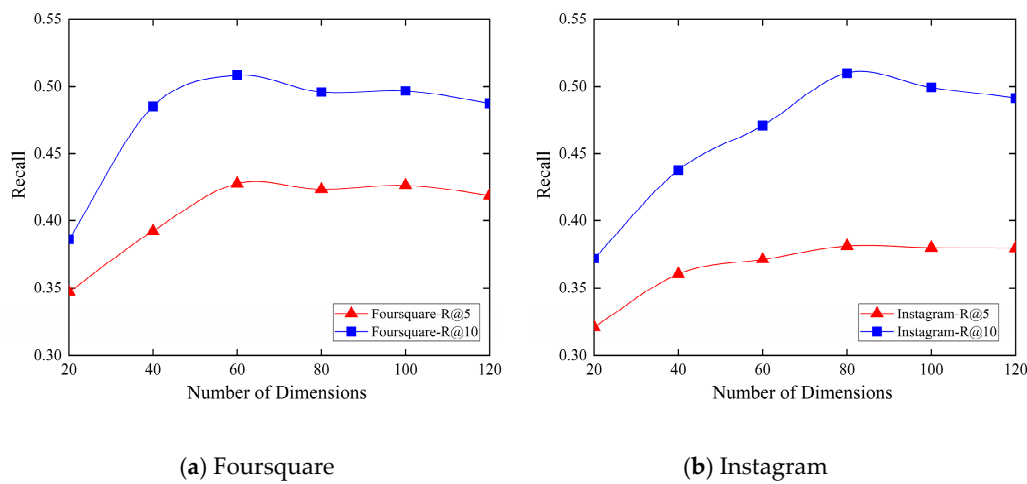


Figure 14. Effect of embedding size in ST-DCGN.

Table 3 shows the impact of different spatial windows widths. We analyzed the performance of the proposed ST-DCGN model on both datasets with different spatial window widths (i.e., 0.1 km, 0.3 km, 0.5 km, and 0.7 km) regarding R@5 and F1@5. It is obviously seen from Table 3 that ST-DCGN achieved the best performance on the Foursquare dataset when the spatial window width Δd was set to 0.5 km while the best performance was achieved on the Instagram dataset when Δd was set to 0.3 km. An explanation is that the distances distributions of consecutive check-ins are different on two datasets. For example, for the Foursquare and Instagram dataset, 85% and 93% consecutive check-ins were less than 10 km, respectively, as shown in Figure 10b. Therefore, we can see that a larger Δd value may be more suitable when dataset covers more longer distances.

Table 3. Performance of ST-DCGN with varying window width by R@5 and F1@5.

		0.1 km	0.3 km	0.5 km	0.7 km
Foursquare	R@5	0.3920	0.4248	0.4277	0.4121
	F1@5	0.1276	0.1367	0.1426	0.1251
Instagram	R@5	0.3752	0.3810	0.3793	0.3681
	F1@5	0.1189	0.1270	0.1207	0.1186

Figure 15 presents the performance of the proposed ST-DCGN with different sequence length while keeping other optimal hyperparameters unchanged. We can observe that the best POI recommendation performance is achieved, respectively, when maximum sequence length $k = 80$ and $k = 30$ on the

Foursquare and Instagram datasets. This result further suggests that our method can learn both short-term and long-term sequence dependencies well.

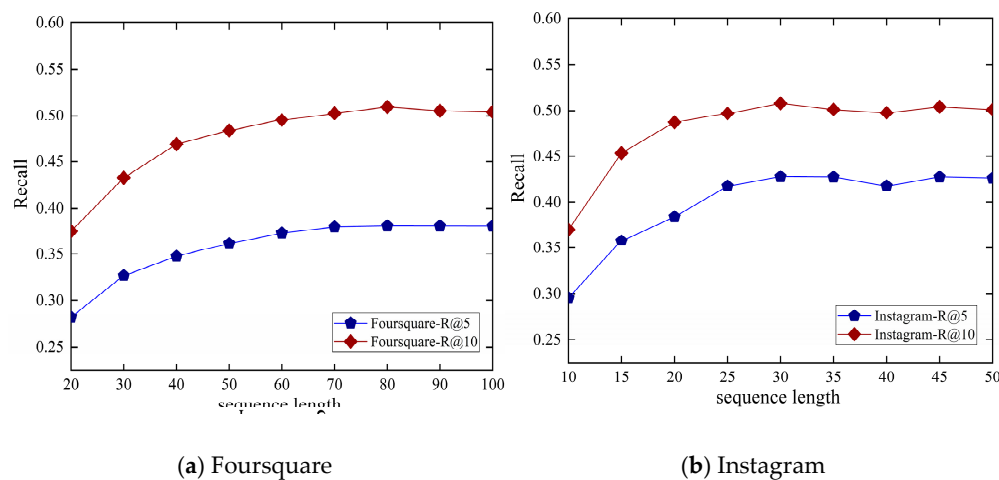


Figure 15. Performance of ST-DCGN with different sequence lengths by R@5 and R@10.

5. Conclusions and Future Work

In this work, we presented a spatiotemporal dilated convolutional generative network (i.e., ST-DCGN) for POI recommendation based on a deep neural network known as the WaveNet architecture [23]. The proposed method introduces a conditional generative model and dilated causal convolutions network to model users' check-in sequences, which are very effective to model the short- and long-range dependencies. Compared with the RNNs based methods, such a network structure can fully utilize parallel computation within a check-in sequence and greatly reduce the training and evaluation time of the model. In addition, we acquired the user's personalized spatial preference and personalized temporal preference by using the continuous geographical distance and encoded specific time ID in each time step. Extensive experiments were conducted to evaluate the performance of ST-DCGN and other comparative methods. The experimental results showed that our proposed ST-DCGN model can achieve better performance than state-of-the-art methods for POI recommendation.

In the future, we will incorporate more check-in features to improve performance of POI recommendation, like users' activities, comment text, and picture information. On the other hand, we will explore more advanced neural networks, like graph convolutional neural network. Moreover, recent studies show that some conventional methods based on matrix factorization could generalize better [53,54]. Therefore, these methods are also worth exploring in the future.

Author Contributions: Conceptualization and methodology, Chunyang Liu; validation and formal analysis, Chunyang Liu, Jiping Liu, Shenghua Xu, and Jian Wang; writing—original draft preparation, Chunyang Liu, Chao Liu, and Tianyang Chen; writing—review and editing, Chunyang Liu and Tao Jiang. All authors have read and agreed to the published version of the manuscript.

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. 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](#)]

3. Majid, A.; Chen, L.; Chen, G.; Mirza, H.T.; Hussain, I.; Woodward, J. A context-aware personalized travel recommendation system based on geotagged social media data mining. *Int. J. Geogr. Inf. Sci.* **2018**, *27*, 662–684. [\[CrossRef\]](#)
4. Wan, L.; Hong, Y.; Huang, Z.; Peng, X.; Li, R. A hybrid ensemble learning method for tourist route recommendations based on geo-tagged social networks. *Int. J. Geogr. Inf. Sci.* **2018**, *32*, 2225–2246. [\[CrossRef\]](#)
5. 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.
6. 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.
7. 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.
8. 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. [\[CrossRef\]](#)
9. Gan, M.; Gao, L. Discovering Memory-Based Preferences for POI Recommendation in Location-Based Social Networks. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 279. [\[CrossRef\]](#)
10. Huang, L.; Ma, Y.; Wang, S.; Liu, Y. An Attention-based Spatiotemporal LSTM Network for Next POI Recommendation. *IEEE Trans. Serv. Comput.* **2019**, *99*. [\[CrossRef\]](#)
11. Tobler, W.R. A computer movie simulating urban growth in the Detroit region. *Econ. Geogr.* **1970**, *46*, 234–240. [\[CrossRef\]](#)
12. 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.
13. 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.
14. 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.
15. Kefalas, P.; Manolopoulos, Y. A time-aware spatio-textual recommender system. *Expert. Syst. Appl.* **2017**, *78*, 396–406. [\[CrossRef\]](#)
16. Wang, W.; Yin, H.; Du, X.; Nguyen, Q.V.H.; Zhou, X. TPM: A temporal personalized model for spatial item recommendation. *ACM Trans. Intell. Syst. Technol.* **2018**, *9*, 1–25. [\[CrossRef\]](#)
17. 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.
18. 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.
19. 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.
20. 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.
21. Liu, C.; Liu, J.; Wang, J.; Xu, S.; Han, H.; Chen, Y. An Attention-Based Spatiotemporal Gated Recurrent Unit Network for Point-of-Interest Recommendation. *ISPRS Int. J. Geo Inf.* **2019**, *8*, 355. [\[CrossRef\]](#)
22. Yuan, F.; Karatzoglou, A.; Arapakis, I.; Jose, J.M.; He, X. A Simple Convolutional Generative Network for Next Item Recommendation. In Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining, Melbourne, Australia, 11–15 February 2019; pp. 582–590.

23. Oord, A.V.D.; Dieleman, S.; Zen, H.; Simonyan, K.; Vinyals, O.; Graves, A.; Kalchbrenner, N.; Senior, A.; Kavukcuoglu, K. Wavenet: A generative model for raw audio. *arXiv* **2016**, arXiv:1609.03499.
24. Yang, D.; Zhang, D.; Zheng, V.W.; Yu, Z. Modeling user activity preference by leveraging user spatial temporal characteristics in LBSNs. *IEEE Trans. Syst. Man Cybern. Syst.* **2014**, *45*, 129–142. [\[CrossRef\]](#)
25. Chang, B.; Park, Y.; Park, D.; Kim, S.; Kang, J. Content-Aware Hierarchical Point-of-Interest Embedding Model for Successive POI Recommendation. In Proceedings of the Program Committee of the 27th International Joint Conference on Artificial Intelligence, Stockholm, Sweden, 13–19 July 2019; pp. 3301–3307.
26. 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.
27. Kurashima, T.; Iwata, T.; Hoshida, 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.
28. 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.
29. Li, H.; Ge, Y.; Hong, R.; Zhu, H. Point-of-interest recommendations: Learning potential check-ins from friends. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Hilton Hotel, San Francisco, CA, USA, 13–17 August 2016; pp. 975–984.
30. 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.
31. 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.
32. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Adv. Neural Inf. Proc. Syst.* **2012**, *60*, 1097–1105. [\[CrossRef\]](#)
33. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
34. Irsoy, O.; Cardie, C. Deep recursive neural networks for compositionality in language. In Proceedings of the 27th International Conference on Neural Information Processing Systems, Montreal, QC, Canada, 8–13 December 2014; pp. 2096–2104.
35. Dong, L.; Yang, N.; Wang, W.; Wei, F.; Liu, X.; Wang, Y.; Hon, H.W. Unified Language Model Pre-training for Natural Language Understanding and Generation. *arXiv* **2019**, arXiv:1905.03197.
36. Hinton, G.; Deng, L.; Yu, D.; Dahl, G.; Mohamed, A.R.; Jaitly, N.; Sainath, T. Deep neural networks for acoustic modeling in speech recognition. *IEEE Signal. Proc. Mag.* **2012**, *29*, 82–97. [\[CrossRef\]](#)
37. Amodei, D.; Ananthanarayanan, S.; Anubhai, R.; Bai, J.; Battenberg, E.; Case, C.; Chen, J. Deep speech 2: End-to-end speech recognition in English and Mandarin. In Proceedings of the 33rd International Conference on Machine Learning, New York, NY, USA, 19–24 June 2016; pp. 173–182.
38. Zhang, S.; Yao, L.; Sun, A.; Tay, Y. Deep learning based recommender system: A survey and new perspectives. *ACM Comput. Surv.* **2019**, *52*, 5. [\[CrossRef\]](#)
39. Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G.S.; Dean, J. Distributed representations of words and phrases and their compositionality. In Proceedings of the 26th International Conference on Neural Information Processing Systems, Lake Tahoe, NV, USA, 5–10 December 2013; pp. 3111–3119.
40. Liu, X.; Liu, Y.; Li, X. Exploring the Context of Locations for Personalized Location Recommendations. In Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, New York, NY, USA, 9–15 July 2016; pp. 1188–1194.
41. Feng, S.; Cong, G.; An, B.; Chee, Y.M. Poi2vec: Geographical latent representation for predicting future visitors. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017.
42. Kong, D.; Wu, F. HST-LSTM: A Hierarchical Spatial-Temporal Long-Short Term Memory Network for Location Prediction. In Proceedings of the 27th International Joint Conference on Artificial Intelligence, Stockholm, Sweden, 13–19 July 2018; pp. 2341–2347.

43. 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.
44. 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.
45. Tang, J.; Wang, K. Personalized top-n sequential recommendation via convolutional sequence embedding. In Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, Los Angeles, CA, USA, 5–9 February 2018; pp. 565–573.
46. Oord, A.V.D.; Kalchbrenner, N.; Kavukcuoglu, K. Pixel recurrent neural networks. *arXiv* **2016**, arXiv:1601.06759.
47. He, K.; Zhang, X.; Ren, S.; Sun, J. Identity mappings in deep residual networks. In European conference on computer vision. In Proceedings of the European Conference on Computer Vision; Amsterdam, The Netherlands, 11–14 October 2016; pp. 630–645.
48. Ba, J.L.; Kiros, J.R.; Hinton, G.E. Layer normalization. *arXiv* **2016**, arXiv:1607.06450.
49. Nair, V.; Hinton, G.E. Rectified linear units improve restricted Boltzmann machines. In Proceedings of the 27th International Conference on Machine Learning, Haifa, Israel, 21–24 June 2010; pp. 807–814.
50. Jean, S.; Cho, K.; Memisevic, R.; Bengio, Y. On using very large target vocabulary for neural machine translation. *arXiv* **2014**, arXiv:1412.2007.
51. 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.
52. 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.
53. 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, NY, USA, 25–28 June 2017; pp. 93–102.
54. Rendle, S.; Zhang, L.; Koren, Y. On the difficulty of evaluating baselines: A study on recommender systems. *arXiv* **2019**, arXiv:1905.01395.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).