

## Article

# Understanding User Preferences in Location-Based Social Networks via a Novel Self-Attention Mechanism

Lei Shi <sup>1,2</sup> , Jia Luo <sup>3,4,\*</sup>, Peiying Zhang <sup>5</sup> , Hongqi Han <sup>2</sup>, Didier El Baz <sup>4</sup> , Gang Cheng <sup>6</sup> and Zeyu Liang <sup>7</sup>

<sup>1</sup> State Key Laboratory of Media Convergence and Communication, Communication University of China, Beijing 100024, China

<sup>2</sup> Institute of Scientific and Technical Information of China, Beijing 100038, China

<sup>3</sup> College of Economics and Management, Beijing University of Technology, Beijing 100021, China

<sup>4</sup> LAAS-CNRS, Université de Toulouse, CNRS, 31031 Toulouse, France

<sup>5</sup> College of Computer Science and Technology, China University of Petroleum (East China), Qingdao 266580, China

<sup>6</sup> School of Computer Science, North China Institute of Science and Technology, Beijing 065201, China

<sup>7</sup> School of Data Science and Intelligent Media, Communication University of China, Beijing 100024, China

\* Correspondence: jialuo@bjut.edu.cn

**Abstract:** The check-in behaviors of users are ubiquitous in location-based social networks in urban living. Understanding user preferences is critical to improving the recommendation services of social platforms. In addition, great quality of recommendation is also beneficial to sustainable urban living since the user can easily find the point of interest (POI) to visit, which avoids unnecessary consumption, such as a longer time taken for searching or driving. To capture user preferences from their check-in behaviors, advanced methods transform historical records into graph structure data and further leverage graph deep learning-based techniques to learn user preferences. Despite their effectiveness, existing graph deep learning-based methods are limited to the capture of the deep graph's structural information due to inherent limitations, such as the over-smoothing problem in graph neural networks, further leading to suboptimal performance. To address the above issues, we propose a novel method built on Transformer architecture named spatiotemporal aware transformer (STAT) via a novel graphically aware attention mechanism. In addition, a new temporally aware sampling strategy is developed to reduce the computational cost and enable STAT to deal with large graphs. Extensive experiments on real-world datasets have demonstrated the superiority of the STAT compared to state-of-the-art POI recommendation methods.

**Keywords:** user preference; social network; POI recommendation; deep learning; attention mechanism



**Citation:** Shi, L.; Luo, J.; Zhang, P.; Han, H.; El Baz, D.; Cheng, G.; Liang, Z. Understanding User Preferences in Location-Based Social Networks via a Novel Self-Attention Mechanism.

*Sustainability* **2022**, *14*, 16414.

<https://doi.org/10.3390/su142416414>

su142416414

Academic Editors: Haoran Wei, Zhendong Wang, Yuchao Chang and Zhenghua Huang

Received: 15 November 2022

Accepted: 3 December 2022

Published: 8 December 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 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 (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

With the popularization of intelligent device applications and the continuous progress of mobile positioning service technology, users are more likely to share their living experiences on social platforms compared with the past decade. For instance, a user could share a photo about a scenic spot or a comment about a restaurant with their friends on location-based social networks [1] (LBSNs). A place that a user is interested in visiting is called a point of interest (POI).

Various types of urban life-related interest points (shopping malls, restaurants, parks, museums, tourist attractions, entertainment venues, etc.) have emerged in large numbers on the Internet in a sustainable manner. All kinds of urban life and daily clothes not only enrich peoples' lives, but also bring about the problem of "choice paralysis". Point of interest recommendation is a personalized recommendation based on contextual information and location awareness. A point of interest recommendation relates users and points of interest, aiming to recommend new, interesting locations to users. Point of interest recommendations based on location-based social networks play an important role in providing a better life

and services for people. Hence, point of interest (POI) recommendation is one of the most important services in LBSNs, whose goal is to recommend POIs to users according to the check-in records of users [2–4]. Great POI recommendation services can improve the lives of urban users since the recommendation system can help users easily find their POIs, avoiding a long time searching, which is also beneficial to sustainable urban living.

However, the complex contextual information in the POI recommendation scenario, including geographical context and temporal context, makes the POI recommendation task more challenging than other conventional recommendation tasks, such as movie recommendation and music recommendation.

Such a challenging task has attracted great attention in the field of recommendation systems. Various techniques [5–9] have been proposed to enhance the performance of this task, such as matrix factorization-based methods and collaborative filtering-based methods. Among a diversity of techniques, deep learning-based methods have shown remarkable performance and exhibited high flexibility, which has attracted recent attention. Graph embedding-based models and graph neural network-based models are two representative approaches of deep learning-based methods.

The graph embedding-based models [6] aim to learn the low-dimension representations of users and POIs based on the graphs constructed according to the check-in records of users. By introducing different contextual information to generate the corresponding graphs, graph embedding-based models can flexibly integrate various information to learn the representation of nodes.

The goal of graph neural network-based models [8] is to leverage advanced graph deep learning techniques to learn the representations of users and POIs from both topology features and semantic features. Similar to the former, methods in this category also utilize generated graphs to capture the influences of different contextual factors. Benefiting from a message-passing mechanism, graph neural network-based models can extract more deep information from a graph's structural data, resulting in superior performance.

Despite effectiveness, the above two types of methods share a common weakness in that they are unable to capture the long-range dependencies of users and POIs. Since these methods treat users and POIs as nodes of graphs, they only leverage the topology information for learning node representations. For instance, most graph embedding-based methods regard immediate neighbors as the positive nodes, which indicates that the information of immediate neighbors is meaningful, while graph neural network-based methods depend on a message-passing mechanism, which has been proven to suffer from over-smoothing problems when obtaining remote nodes.

In light of the above limitations, in this paper we propose a novel recommendation model built on a Transformer architecture named spatiotemporal aware transformer (STAT). By introducing a self-attention mechanism, the STAT can carefully capture long-range dependencies via a full attention matrix. To enhance the expressiveness of the original self-attention mechanism, we propose a novel geographically aware attention mechanism that integrates geographical information into an attention matrix, which better preserves interactions between POIs. Moreover, for generalizing the transformer to large-scale social networks, we develop a temporally aware node sampling strategy that utilizes the temporal factor to sample relevant POIs for model training. To validate the effectiveness of the STAT, we conduct extensive experiments on widely used real-world datasets. The experimental results demonstrate the superiority of our proposed STAT compared to the representative graph-embedding methods and graph neural network-based methods.

The main contributions of this paper are summarized as follows:

- We propose a Transformer-based model named STAT, which leverages the novel geographically aware attention mechanism to learn the representations of POIs for the recommendation of urban life-related services.
- We develop a temporally aware sampling strategy that samples relevant POIs according to the check-in timestamp, which carefully preserves the influence of the temporal factor.

- We conduct extensive experiments on real-world urban life-related datasets, demonstrating our proposed STAT's effectiveness.

## 2. Related Work

In this section, we briefly review recent related works from the perspective of utilizing temporal factors and geographical factors.

### 2.1. Temporal Factor

The check-in behaviors of users are influenced by time factors to a large extent. For example, users tend to visit a restaurant, not a bar, to have a meal and a drink at twelve noon. In addition, users may share similar check-in habits in the same temporal context. Thus, exploiting the influence of temporal factors is crucial to capturing user preferences.

A general idea is to divide time into several timestamps to learn the representations of users based on different time patterns [10]. Christoforidis et al. [2] constructed two types of directed bipartite graphs to represent the interaction of a user or a POI and a specific timestamp and further leverages a graph embedding method to learn the representation vectors of users and POIs. Yuan et al. [4] proposed a collaborative recommendation model, which incorporates temporal information for generating the recommended POIs. Kefalas et al. [5] considered the temporal dimension and measured the impact of time on various time intervals for learning user preferences. Xie et al. [6] proposed a unified graph model to explore semantic vectors from a temporal context. Different from the above studies, Doan et al. [7] utilized the long short-term memory (LSTM) recurrent neural network to simultaneously capture both the sequential and temporal features of users' representations. Gao et al. [9] proposed a temporal state to represent the specific hour of the day and further utilize different temporal states to learn the check-in habits of users. Dai et al. [11] developed a spatiotemporal neural network framework to utilize the check-in history and social ties of users for recommending personalized POIs. Wang et al. [12] proposed a graph-enhanced spatial-temporal network that leverages the recurrent neural network (RNN) to learn user-specific temporal dependencies. Wang et al. [13] developed a time-aware position encoder to consider the temporal intervals among POIs separately.

### 2.2. Geographical Factor

The geographical factor is also one of the most important factors in capturing unique user preferences. Due to the expensive cost of check-in behavior, POIs which are far from the current location are not considered. Moreover, users tend to visit places near their lives. For instance, users may go to a nearby restaurant when they visit a mall. Hence, considering the geographical factor is beneficial to providing meaningful POIs for users.

Ye et al. [14] utilized the naïve Bayesian classifier to explore the geographical influence of POI recommendations based on the collaborative recommendation algorithm. Sun et al. [15] leveraged a geo-dilated RNN for short-term preference learning. Liu et al. [16] proposed a geographical probabilistic factor framework that leveraged matrix factorization techniques to integrate the geographical factor into the model to learn the user's presentations. Liu et al. [17] characterized the geographical clustering phenomenon more precisely based on the location visual content, which can improve the recommendation performance. Huo et al. [18] developed a geographical location privacy-preserving method based on Laplacian distributed noise to preserve the privacy of users. Su et al. [19] utilized the graph convolutional network (GCN) to integrate social relationships and geographical influence. Zhang et al. [20] proposed a personalized geographical influence method that jointly learns the geographical and diversity preferences of users. Liu et al. [21] developed a geographical-temporal awareness hierarchical attention network that utilizes the attention mechanism to capture the subtle POI-POI relationships from a multi-contextual perspective. For more works about POI recommendation, please refer to the recent surveys [22,23].

### 3. Preliminaries

In this section, we first introduce the definitions and notations in this paper. Then, we provide a description of Transformer architecture.

#### 3.1. Definitions

**Definition 1** (POI). A POI  $v$  contains several types of features, including geographical features, category features and attribute features.  $l_v$  denotes the geographical information of POI  $v$ , which is described by longitude and latitude. Category features and attribute features determine the property of POI  $v$ , for instance, “restaurant” is the category information of POI  $v$  and the decoration style is one of the attribute features of POI  $v$ . In this paper, we regard the category information and attribute information as the semantic features, and we utilize the widely used embedding technique Word2Vec [24] to generate the semantic features  $\mathbf{X} \in \mathbb{R}^{m \times d}$ , where  $d$  denotes the dimension of the feature vector and  $m$  denotes the number of POIs.

**Definition 2** (check-in record). A check-in record  $c = (u, v, t)$  contains the history of check-in behavior of a user  $u$ , where  $t$  denotes the check-in timestamps. In this paper, we leverage 24 h as the division of the timestamps.

**Definition 3** (top- $k$  POI recommendation). Given the history records  $C_u$  of user  $u$ , the goal of POI recommendation is to provide a recommended POI list  $\{v_0, \dots, v_{k-1}\}$  whose length is  $k$  according to the query  $(u, l, t)$ , where  $l$  denotes the current location, which is also described by longitude and latitude, and  $t$  denotes the current timestamp.

#### 3.2. Transformer

The Transformer’s architecture [25] follows the structure of an auto-encoder that consists of an encoder and a decoder. In this paper, we utilize the Transformer’s encoder to learn the representations of POIs. Next, we introduce the components of the encoder.

The Transformer’s encoder is composed of a multi-head attention (MHA) module and a feed-forward neural network (FFN) module. For brevity, we introduce single-head attention (SHA), which can easily extend to MHA. Suppose we have the input feature matrix  $\mathbf{H} \in \mathbb{R}^{n \times d}$ , the output of SHA is calculated as follows:

$$\mathbf{A} = \rho\left(\frac{(\mathbf{H}\mathbf{W}_Q)(\mathbf{H}\mathbf{W}_K)^T}{\sqrt{d_w}}\right), \quad (1)$$

$$\mathbf{H}_{out} = \mathbf{A}(\mathbf{H}\mathbf{W}_V), \quad (2)$$

where  $\mathbf{W}_Q$ ,  $\mathbf{W}_K$  and  $\mathbf{W}_V$  are the learnable parameter matrices and the dimensions of these matrices are the same  $d \times d_w$ .  $\rho(\cdot)$  denotes the softmax function [25] and  $\mathbf{A}$  represents the attention matrix, which preserves the interactions of all item pairs.

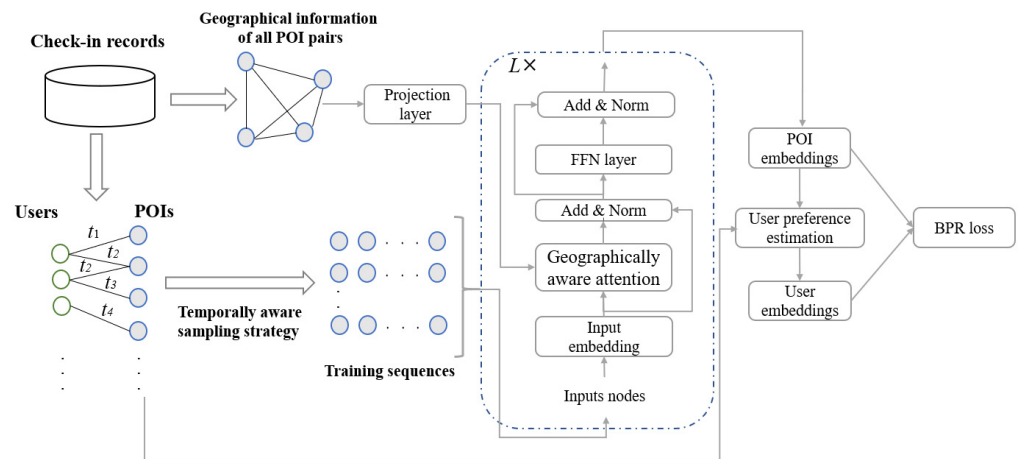
The FFN module contains two linear layers and a nonlinear activation function:

$$\mathbf{H}_{final} = \zeta_2(\sigma(\zeta_1(\mathbf{H}_{out}))), \quad (3)$$

where  $\zeta_1(\cdot)$  and  $\zeta_2(\cdot)$  denote the linear layer with different parameters, which is a basic module of a neural network.  $\sigma(\cdot)$  denotes the nonlinear activation function, and the residual connection technique is adopted to enhance the expressiveness of the FFN module’s output.

### 4. Methodology

In this section, we detail our proposed spatio-temporal aware transformer (STAT) method, which consists of two main designs: geographically aware attention mechanism and temporally aware sampling strategy. We introduce them in turn. The framework of STAT is shown in Figure 1.



**Figure 1.** The overall framework of STAT.

#### 4.1. Geographically Aware Attention Mechanism

As mentioned in Section 3.1, for each POI, we first constructed the semantic features, resulting in a feature matrix  $\mathbf{X}$ . Recent advanced graph deep learning-based methods transform check-in records into graph structural data, and further leverage popular graph deep learning techniques, such as graph embeddings and graph neural networks, to learn the representations of users and POIs. However, these techniques limit the model to capturing the long-range dependencies of POIs due to their fixed graph structure.

In light of this limitation, in this paper, we leveraged the self-attention mechanism of Transformer architecture to learn the interactions between POIs. Since the self-attention mechanism regards all input items as connected, the learned representations of POIs can obtain more meaningful global information.

However, the original self-attention mechanism was developed for learning the representations of words in a sentence, which does not carry extra information except semantic features. The POIs, discussed in Section 3.1, have complex information such as semantic information and geographical information, and the geographical information is important to learning the representations of POIs. Unfortunately, the calculation of the attention matrix is unable to capture this key information.

To enable the self-attention mechanism to preserve the geographical information of POIs, we propose a novel geographically aware attention mechanism. Intuitively, the relation of POIs is sensitive to the geographical distance of them. According to Tobler's first law of geography that states "Everything is related to everything else, but near things are more related than distant things" [26], we calculated the distances of all POI pairs as a bias to strengthen the attention matrix to capture the geographical influence:

$$\mathbf{A}_{i,j} = \frac{(\mathbf{X}_i \mathbf{W}_Q)(\mathbf{X}_j \mathbf{W}_K)^T}{\sqrt{d_w}} + \phi(\zeta(v_i, v_j)), \quad (4)$$

where  $\mathbf{X}_i$  and  $\mathbf{X}_j$  are the semantic feature vectors of POI  $v_i$  and POI  $v_j$ .  $\zeta(v_i, v_j)$  is a distance calculation function determined by the longitude and latitude of POIs.  $\phi(\cdot)$  denotes the learnable projection layer that transforms the distance into a scalar value. In this way, the geographical information can be carefully preserved in the attention matrix, which enables the self-attention mechanism to capture the influence of the geographical factor. Based on the geographically factor-aware attention matrix, we obtained the representations of POIs  $\mathbf{H}^V \in \mathbb{R}^{m \times d_w}$  according to Equations (2) and (3).

#### 4.2. Temporally Aware Sampling Strategy

Another limitation of generalizing Transformer to POI recommendation is the huge computational consumption of the self-attention mechanism, which is the quadratic computational and storage complexity with the number of POIs. In addition, in real-world applica-



tions, the number of POIs is usually very large, making it hard to train a Transformer-based method. On the other hand, not all POIs have close connections. As discussed in Section 2.1, POIs that have been visited at the same timestamp have more similar features. To preserve this similarity and reduce the training cost, in this paper, we propose a temporally aware sampling strategy, which samples the similar POIs according to check-in timestamps.

Specifically, for each POI  $v$ , we sampled a list  $N_{v,t} = \{v_0, \dots, v_{s-1}\}$  at the timestamp  $t$ , where  $s$  denotes the length of the list. Such a POI list can be regarded as a sentence in the field of natural language processing. And we sampled several lists for POI  $v$  according the check-in history record of  $v$ .

In this way, the training cost of the self-attention mechanism was reduced from  $O(m^2)$  to  $O(s^2)$ , which is affordable for most training environments. More importantly, the sampling process was conducted before the model training stage so that the mini-batch training technique could employed, which guaranteed the scalability of the proposed model.

Moreover, the temporally aware sampling strategy filtered out irrelevant POIs based on check-in history records, which also enhanced the effectiveness of the self-attention mechanism and captured the influence of the temporal factor.

#### 4.3. User Preference Estimation

Through the above components, we obtained the representation vectors of the POIs. Then, we leveraged them to estimate the representations of users given the fact that the visited POIs can reflect user preference. For instance, if a user loves food, there will be a large number of check-in records at restaurants. To calculate the representations of users, a naïve idea is to sum all the representation vectors of visited POIs to represent the preference of the target user  $u$ :

$$\mathbf{H}_u^U = \sum_{v \in C_u} \mathbf{H}_v^V, \quad (5)$$

where  $\mathbf{H}_u^U \in \mathbb{R}^{1 \times d_w}$  denotes the representation of user  $u$ . However, such a simple design makes it hard to capture dynamic user preferences since user preferences change over time. To address this limitation, in this paper, we proposed a user preference estimation method based on the check-in timestamp:

$$\mathbf{H}_u^U = \sum_{(v,t) \in C_u} \gamma(t - t_0) \mathbf{H}_v^V, \quad (6)$$

where  $t$  denotes the timestamp of the check-in record and  $t_0$  denotes the current timestamp.  $\gamma(\cdot)$  represents the sigmoid function. The motivation of this strategy is that the longer the time is, the smaller the impact of the check-in record on user preferences. In this way, we can carefully preserve the dynamics of user preferences.

#### 4.4. Parameter Learning

In this paper, we adopted a widely used optimization method, Bayesian personalized ranking [27], for learning the model's parameters. Specifically, we constructed the following objective function:

$$L_{STAT} = - \sum_{u=1}^m \sum_{(u,v_i) \in C_u} \ln \psi(\mathbf{H}_u^U \cdot \mathbf{H}_{v_i}^{V^T} - \mathbf{H}_u^U \cdot \mathbf{H}_{v_j}^{V^T}) + \mu \|\Theta\|^2, \quad (7)$$

where  $v_j$  denotes a random POI that user  $u$  has not visited.  $\Theta$  represents the model's parameters and  $\mu$  denotes the regularization coefficient.  $\psi(\cdot)$  denotes the sigmoid function. By minimizing Equation (7), we learnt the model's parameters based on the stochastic gradient descent method.

#### 4.5. Using STAT for Recommendation

For a POI recommendation service request  $S = (u, r, t)$ , where  $u$  denotes the target user,  $r$  denotes the current location of the user  $u$  and  $t$  denotes the timestamp of the request. We first estimated the user's preference at timestamp  $t$  via Equation (6). Then, we calculated the interesting scores between the target user  $u$  and POIs that had not appeared in  $u$ 's check-in records. Considering the cost of check-in behaviors, we filtered out the POIs more than 200 km away from the current location  $r$ , which is almost a two-hour drive. Finally, we ranked the POIs according to the interesting score and kept the top- $k$  items as recommended POIs.

### 5. Experiments

In this section, we first introduce the experimental settings of this paper, including dataset, baselines, evaluation metrics and parameter settings. Then, we introduce the designs of the experiments. Finally, we report the experimental results and provide the discussions.

#### 5.1. Dataset

In this paper, we adopted widely used real-world datasets [28], Foursquare [22] and Gowalla [22], for experiments.

The Foursquare dataset was collected from the famous location-based social network Foursquare, which contains user check-in records in New York and Tokyo from 2010 to 2014. Each check-in record contains rich data elements, including the information of users and POIs, the check-in timestamp and POI categories. Information on users' friendships is also involved in the dataset.

The Gowalla dataset was collected from the mobile social media platform Gowalla, which contains the check-in records from the mobile users from February 2009 to October 2010. In Gowalla, each record only contains the information of users, POIs and timestamps of check-ins.

The statistics of the datasets are reported in Table 1. In practice, we removed users and POIs that had less than 20 check-in records. We partitioned the datasets with the timeseries, and the former 80% of records were selected as the training set, following recent works [28,29]. The rest were regarded as the test set.

**Table 1.** Statistics of datasets.

Dataset	#Users	#POIs	#Records
Foursquare	24,941	28,593	1,196,248
Gowalla	18,737	32,510	1,278,274

#### 5.2. Baseline

In this paper, we chose four representative methods for performance comparison, GE, STA, GT-HAN, GPR and GNN-POI. The first two are embedding-based models, while the latter three are GNN-based methods.

GE [6]: GE is a graph embedding-based method that jointly learns the representations of POIs from multi-context features, including geographical features, temporal features and semantic features. Then, GE calculates the recommendation scores of POIs based on the above-learned features.

STA [30]: STA is a translation-based embedding method that leverages geographical and temporal information to learn the representations of users and POIs. Specifically, STA constructs the translation relationship of users, POIs and contextual information and further introduces the translation-based framework to learn the representations of users and POIs.

GT-HAN [21]: GT-HAN is a hybrid model based on attention networks, where it first utilizes a geographical-temporal attention network to learn the representations of POIs

from multi-contextual information and leverages a context-specific co-attention network to learn user preferences.

GPR [29]: GPR is a graph neural network-based method that extracts user preferences from check-in graphs constructed using geographical information. Specifically, GPR utilizes a graph auto-encoder to learn two types of geographical influences: ingoing influences and outgoing influences for learning complex geographical influences from users' check-in networks.

GNN-POI [28]: GNN-POI is hybrid model that leverages graph neural networks to learn node representations from a topological structure and utilizes bidirectional long short-term memory to model users' sequential check-in behavior, which comprehensively considers the influence of temporal and geographical factors.

### 5.3. Evaluation Metric

In this paper, we evaluated the performance of the recommendation task via three evaluation metrics followed [28,31]: precision [32] at  $k$ , recall [32] at  $k$  and normalized discounted cumulative gain (NDCG) at  $k$ , where  $k$  denotes the length of the recommendation list. We varied  $k$  from {5, 10, 15, 20} in experiments to observe the performance of models.

### 5.4. Performance Comparison

In this section, we evaluated all models on the Foursquare and Gowalla datasets via the mentioned evaluation metrics. Specifically, we ran each model ten times with different random seeds and reported the mean values of each evaluation metric. The results are shown in Figures 2–7.

From the figures, we can observe that the values are relatively low on each evaluation metric. This may be because of the high data sparsity of the real-world datasets, recent work [28,29] has also made similar observations. In addition, we can observe that our proposed STAT consistently outperforms all baselines on all datasets via the three evaluation metrics, which indicates the superiority of STAT. In addition, STAT beats GPR and GNN-POI, which implies that introducing the self-attention mechanism leads to better model performance than the message-passing mechanism. Moreover, we also observed that GNN-based methods outperform embedding-based methods, which declares that introducing semantic features can assist in learning more expressive representations of users and POIs.

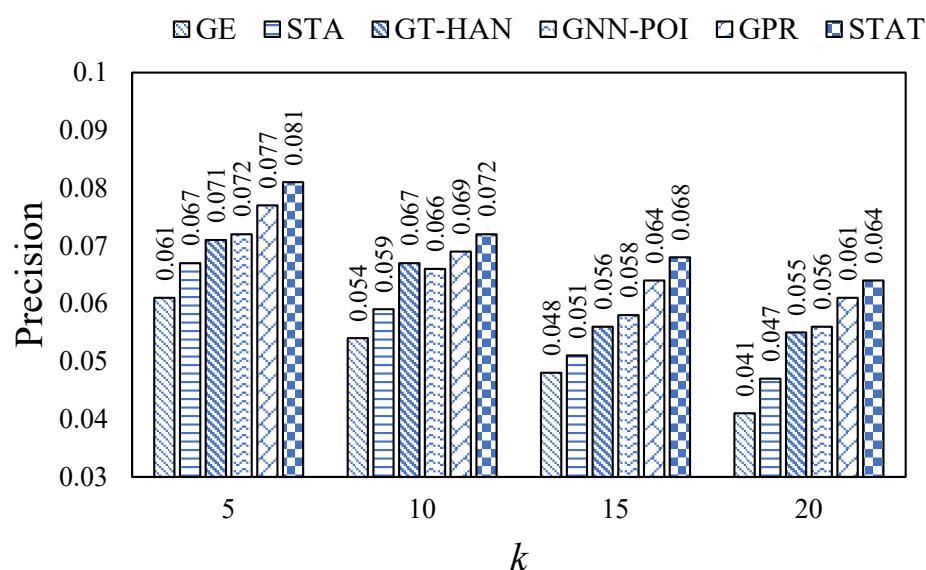


Figure 2. Performance of precision at  $k$  on Foursquare dataset.



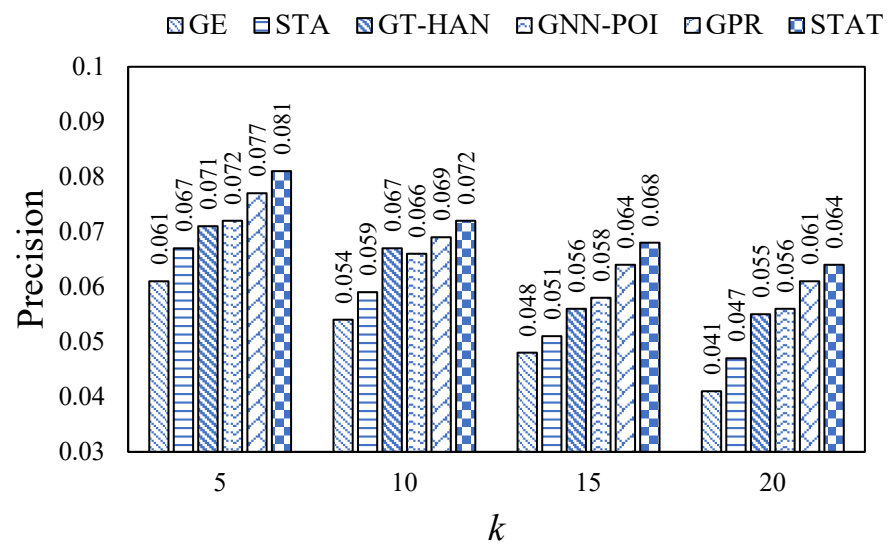


Figure 3. Performance of recall at  $k$  on Foursquare dataset.

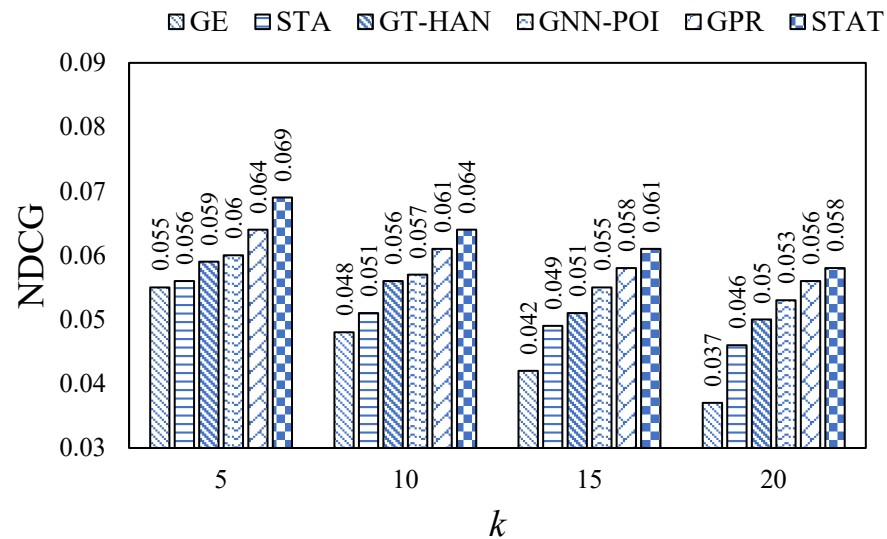


Figure 4. Performance of NDCG at  $k$  on Foursquare dataset.

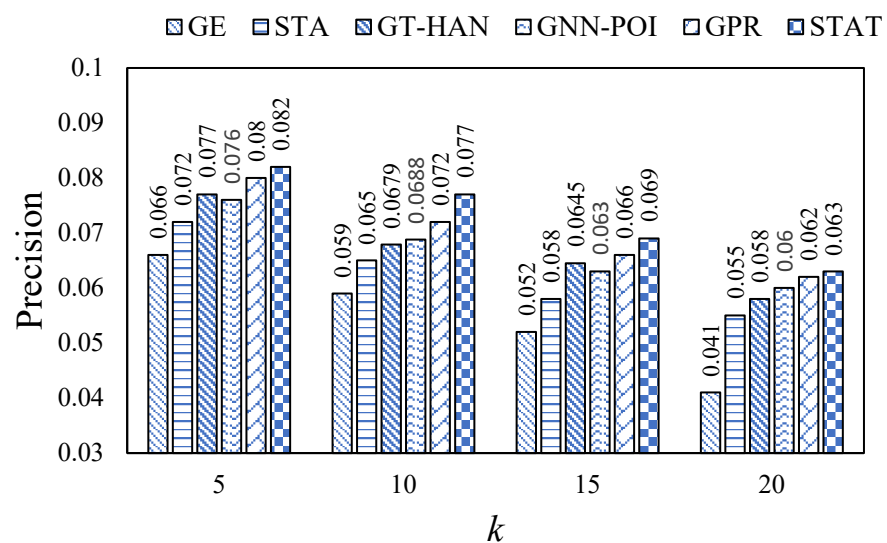


Figure 5. Performance of precision at  $k$  on Gowalla dataset.

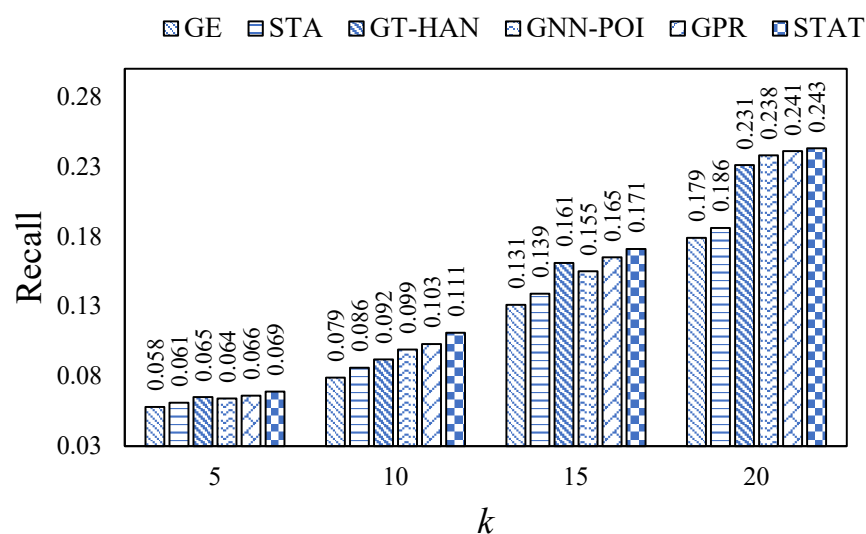


Figure 6. Performance of recall at  $k$  on Gowalla dataset.

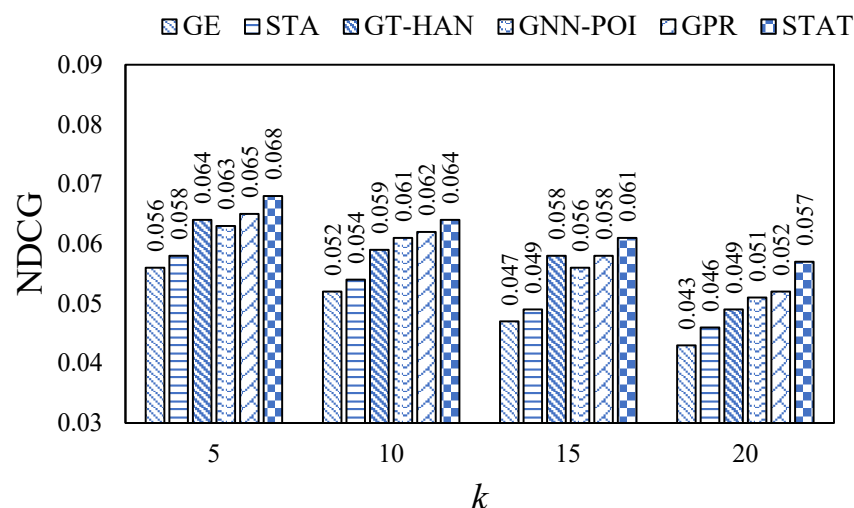


Figure 7. Performance of NDCG  $k$  on Gowalla dataset.

### 5.5. Study of the Parameter Sensitiveness

In this section, we study the parameter sensitiveness of STAT. There is a key parameter of STAT, the number of sampling nodes  $s$ . Intuitively, a large value of  $s$  brings more information since the sampling list contains more nodes for aggregating the information. To validate the impact of  $s$  on the model's performance, we conducted experiments using the Foursquare dataset by fixing  $k = 20$  and varying  $s$  from  $\{20, 40, 60, 80\}$ . The results are reported in Table 2. We can observe that the performance of the model increases first and then decreases with the increase of  $s$ , and STAT achieves the best performance when  $s = 40$ . This may be because when  $s$  is large, the sampling list may contain partial relevant nodes, which could introduce noisy information, further hindering the model's performance. Additionally, a small value of  $s$  achieves competitive performance. In this paper, we used the grid research method to determine the value of  $s$  in different datasets.

Table 2. The impact of  $s$  on Foursquare dataset.

	$s = 20$	$s = 40$	$s = 60$	$s = 80$
Precision at 20	0.061	0.064	0.056	0.051
Recall 20	0.234	0.249	0.211	0.196
NDCG 20	0.056	0.057	0.052	0.048

### 5.6. Ablation Study

In this section, we study the influence of different user preference estimation methods on model performance. As discussed in Section 4.3, STAT utilizes a time-aware method to calculate user preference via Equation (6), which can capture dynamic user preferences. Hence, we provided a variant of STAT named STAT-T. In STAT, we used Equation (5) to learn the user's preference, which means that we ignored the influence of the temporal factor on user preference. We ran the above two methods on the Foursquare dataset and the results are shown in Figures 8–10.

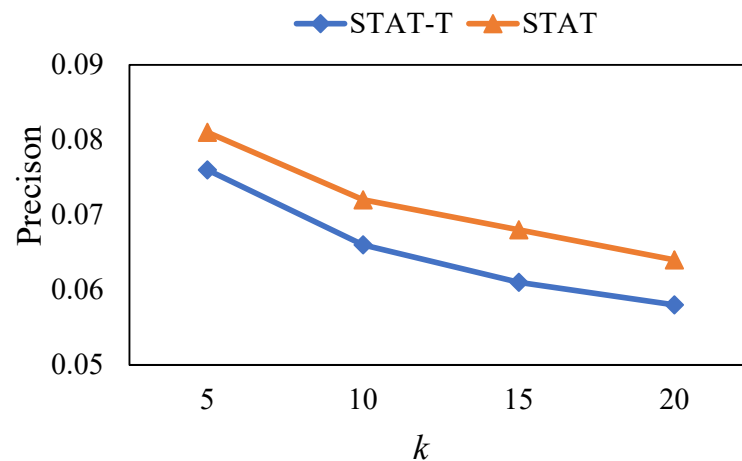


Figure 8. Precision at  $k$  of STAT and its variant on Foursquare dataset.

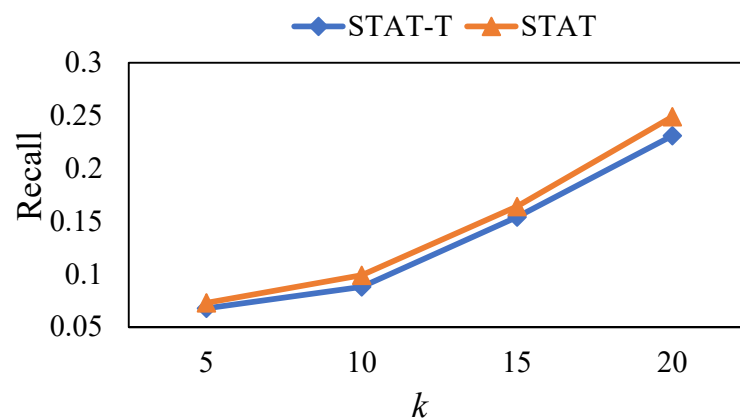


Figure 9. Recall  $k$  of STAT and its variant on Foursquare dataset.

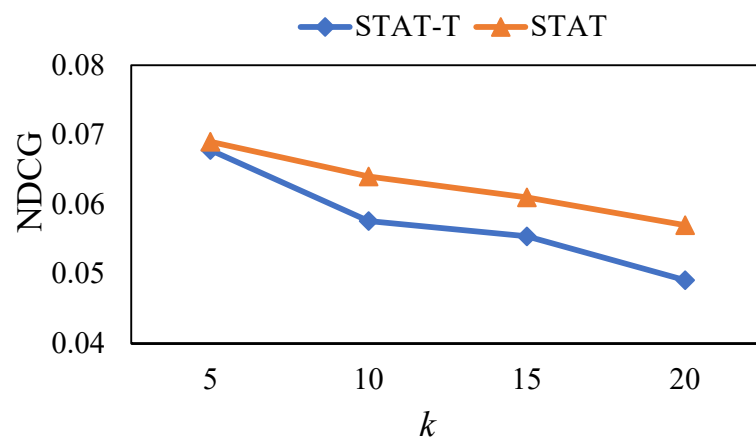


Figure 10. NDCG at  $k$  of STAT and its variant on Foursquare dataset.

Intuitively, user preferences change over time, and recent check-in behaviors better represent the user's preferences, while early check-in records generally have little influence on the user's preferences. From the results, we observe that STAT outperforms STAT-T. This phenomenon indicates that our developed time-aware user preference estimation method can better capture the user preference, demonstrating that the temporal factor is beneficial to accurately obtaining user preferences.

### 5.7. Efficiency Study

In this section, we conduct experiments to analyze the training cost of our proposed STAT. Since baselines are implemented using different programming languages, such as the official implementation of GE, which is based on C++. We only reported the running time and the memory cost of STAT, and our implementation is based on Python and the Pytorch framework. The experiments were conducted on a Linux server with one I9-10900K CPU and one RTX 2080TI GPU. The original cost of Transformer was too large to afford. Benefitting from our proposed temporally aware sampling strategy, we utilized the mini-batch training method for training the model on large-scale location-based social networks, guaranteeing the scalability of STAT. The experimental results are reported in Table 3.

**Table 3.** The training cost of STAT.

	Running Time (s)	Memory Cost (MB)
Foursquare	67.58	2349
Gowalla	46.18	2238

As mentioned in Section 4.2, the complexity of STAT is mainly related to the length of the sampling sequence. Hence, the memory cost was reduced to an affordable value. If we use the original Transformer model, we encounter the out-of-memory problem on these datasets due to the complexity square of the number of POIs.

## 6. Conclusions

In this paper, we propose STAT, a Transformer-based POI recommendation model, which takes the geographical factor and the temporal factor into account to learn the representations of POIs. Specifically, STAT develops a novel geographically aware attention mechanism that integrates the geographical influence into the self-attention mechanism to enhance the expressiveness of the attention matrix in the POI recommendation scenario. In addition, to generalize the Transformer architecture to a large-scale location-based social network, STAT proposes a temporally aware sampling strategy that samples several relevant nodes based on the check-in timestamp. In this way, the influence of the temporal factor was carefully preserved. Moreover, STAT develops a time-aware user preference estimation to capture dynamic user preferences. To validate the effectiveness of STAT, we conducted extensive experiments on real-world datasets via three widely used evaluation metrics. The experimental results indicate that STAT consistently achieved highest values on different evaluation metrics compared to baselines. For instance, STAT obtained 7.3% of recall on Foursquare when the length of recommendation list was 5, while the second-best performance was 6.8% from GPR. We also developed the efficiency study to analyze the training cost of STAT, and the results show that our proposed STAT has a good scalability for a large-scale network.

The advantage of STAT is to leverage the self-attention mechanism to learn the representations of POIs from complex contextual information. However, STAT only considers the temporal factor and geographical factor. Other important factors, such as social information of users and category information of POIs, are not involved. One potential future work would be to develop a reasonable framework to jointly model the above factors to better learn user preferences.

**Author Contributions:** Conceptualization, L.S.; Methodology, L.S. and J.L.; Software, J.L., P.Z. and H.H.; Validation, H.H.; Formal analysis, P.Z., D.E.B. and G.C.; Investigation, G.C.; Data curation, Z.L.; Writing—original draft, L.S., D.E.B. and Z.L.; Writing—review and editing, P.Z., G.C. and Z.L.; Project administration, J.L.; Funding acquisition, L.S., J.L. and G.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the Fundamental Research Funds for the Central Universities (No. CUC220C011, CUC220F002, 300102262513), Science and Technology Think Tank Young Talents Program of China Association for Science and Technology (No. 20220615ZZ07110274), Central Guidance on Local Science and the Technology Development Fund of Hebei Province (No. 226Z5404G), Shandong Provincial Natural Science Foundation (No. ZR2022LZH015, ZR2020MF006), Natural Science Foundation of Hebei Province of China (No. D2022508002), National Key Research and Development Program of China (No. 2022YFC3302103, 2020YFB2009304), National Natural Science Foundation of China (No. 72104016), R&D Program of Beijing Municipal Education Commission (No. SM202110005011), Guangxi Key Laboratory of Trusted Software.

**Data Availability Statement:** The Foursquare dataset can be downloaded from the link <https://sites.google.com/site/yangdingqi/home/foursquare-dataset> (accessed on 2 December 2022). The Gowalla dataset can be downloaded from the link <http://snap.stanford.edu/data/loc-gowalla.html> (accessed on 2 December 2022).

**Conflicts of Interest:** The authors of this publication declare that there are no conflict of interest.

## References

- Shi, L.; Du, J.; Cheng, G.; Liu, X.; Xiong, Z.; Luo, J. Cross-media search method based on complementary attention and generative adversarial network for social networks. *Int. J. Intell. Syst.* **2022**, *37*, 4393–4416. [CrossRef]
- Christoforidis, G.; Kefalas, P.; Papadopoulos, A.N.; Manolopoulos, Y. RELINE: Point-of-interest recommendations using multiple network embeddings. *Knowl. Inf. Syst.* **2021**, *63*, 791–817. [CrossRef]
- Shi, L.; Song, G.; Cheng, G.; Liu, X. A user-based aggregation topic model for understanding user's preference and intention in social network. *Neurocomputing* **2020**, *413*, 1–13. [CrossRef]
- Yuan, Q.; Cong, G.; Ma, Z.; Sun, A.; Magnenat-Thalmann, N. 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.
- Kefalas, P.; Manolopoulos, Y. A time-aware spatio-textual recommender system. *Expert Syst. Appl.* **2017**, *78*, 396–406. [CrossRef]
- Xie, M.; Yin, H.; Wang, H.; Xu, F.; Chen, W.; Wang, S. Learning graph-based poi embedding for location-based recommendation. In Proceedings of the 25th ACM International on Conference on Information and Knowledge Management, Indianapolis, IN, USA, 24–28 October 2016; pp. 15–24.
- Doan, K.D.; Yang, G.; Reddy, C.K. An attentive spatio-temporal neural model for successive point of interest recommendation. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*; Springer: Cham, Switzerland, 2019; pp. 346–358.
- Shu, T.; Shi, L.; Zhu, C.; Liu, X. A graph neural network framework based on preference-aware graph diffusion for recommendation. *Front. Psychiatry* **2022**, *13*, 1012980. [CrossRef] [PubMed]
- 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.
- Zhang, J.; Jiang, Y.; Wu, S.; Li, X.; Luo, H.; Yin, S. Prediction of remaining useful life based on bidirectional gated recurrent unit with temporal self-attention mechanism. *Reliab. Eng. Syst. Saf.* **2022**, *221*, 108297. [CrossRef]
- Dai, S.; Yu, Y.; Fan, H.; Dong, J. Spatio-Temporal Representation Learning with Social Tie for Personalized POI Recommendation. *Data Sci. Eng.* **2022**, *7*, 44–56. [CrossRef]
- Wang, Z.; Zhu, Y.; Zhang, Q.; Liu, H.; Wang, C.; Liu, T. Graph-Enhanced Spatial-Temporal Network for Next POI Recommendation. *ACM Trans. Knowl. Discov. Data* **2022**, *16*, 1–21. [CrossRef]
- Wang, E.; Jiang, Y.; Xu, Y.; Wang, L.; Yang, Y. Spatial-Temporal Interval Aware Sequential POI Recommendation. In Proceedings of the IEEE 38th International Conference on Data Engineering (ICDE), Kuala Lumpur, Malaysia, 9–12 May 2022; pp. 2086–2098.
- 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, 25–29 July 2011; pp. 325–334.
- Sun, K.; Qian, T.; Chen, T.; Liang, Y.; Nguyen, Q.V.H.; Yin, H. Where to Go Next: Modeling Long- and Short-Term User Preferences for Point-of-Interest Recommendation. *Proc. Conf. AAAI Artif. Intell.* **2020**, *34*, 214–221. [CrossRef]
- 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.

17. Liu, B.; Meng, Q.; Zhang, H.; Xu, K.; Cao, J. VGMF: Visual contents and geographical influence enhanced point-of-interest recommendation in location-based social network. *Trans. Emerg. Telecommun. Technol.* **2022**, *33*, e3889. [[CrossRef](#)]
18. Huo, Y.; Chen, B.; Tang, J.; Zeng, Y. Privacy-preserving point-of-interest recommendation based on geographical and social influence. *Inf. Sci.* **2020**, *543*, 202–218. [[CrossRef](#)]
19. Su, C.; Gong, B.; Xie, X. Personalized Point-of-Interest Recommendation Based on Social and Geographical Influence. In Proceedings of the 4th Artificial Intelligence and Cloud Computing Conference, Kyoto, Japan, 17–19 December 2021; pp. 130–137.
20. Zhang, Y.; Liu, G.; Liu, A.; Zhang, Y.; Li, Z.; Zhang, X.; Li, Q. Personalized Geographical Influence Modeling for POI Recommendation. *IEEE Intell. Syst.* **2020**, *35*, 18–27. [[CrossRef](#)]
21. Liu, T.; Liao, J.; Wu, Z.; Wang, Y.; Wang, J. Exploiting geographical-temporal awareness attention for next point-of-interest recommendation. *Neurocomputing* **2020**, *400*, 227–237. [[CrossRef](#)]
22. Islam, M.A.; Mohammad, M.; Das, S.; Ali, M.E. A survey on deep learning based Point-of-Interest (POI) recommendations. *Neurocomputing* **2022**, *472*, 306–325. [[CrossRef](#)]
23. Sánchez, P.; Bellogín, A. Point-of-Interest Recommender Systems based on Location-Based Social Networks: A Survey from an Experimental Perspective. *ACM Comput. Surv.* **2022**, *54*, 1–37. [[CrossRef](#)]
24. Le, Q.; Mikolov, T. Distributed representations of sentences and documents. In Proceedings of the International Conference on Machine Learning, PMLR, Beijing, China, 21–26 June 2014; pp. 1188–1196.
25. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; 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.
26. 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](#)]
27. Rendle, S.; Freudenthaler, C.; Gantner, Z.; Schmidt-Thieme, L. BPR: Bayesian personalized ranking from implicit feedback. *arXiv* **2012**, arXiv:1205.2618.
28. Zhang, J.; Liu, X.; Zhou, X.; Chu, X. Leveraging graph neural networks for point-of-interest recommendations. *Neurocomputing* **2021**, *462*, 1–13. [[CrossRef](#)]
29. Chang, B.; Jang, G.; Kim, S.; Kang, J. Learning graph-based geographical latent representation for point-of-interest recommendation. In Proceedings of the 29th ACM International Conference on Information & Knowledge Management, Ireland, 19–23 October 2020; pp. 135–144.
30. Qian, T.; Liu, B.; Nguyen, Q.V.H.; Yin, H. Spatiotemporal Representation Learning for Translation-Based POI Recommendation. *ACM Trans. Inf. Syst.* **2019**, *37*, 1–24. [[CrossRef](#)]
31. Shi, L.; Du, J.; Liang, M.; Kou, F. Dynamic topic modeling via self-aggregation for short text streams. *Peer-to-Peer Netw. Appl.* **2018**, *12*, 1403–1417. [[CrossRef](#)]
32. Memiş, S.; Enginoğlu, S.; Erkan, U. Fuzzy parameterized fuzzy soft k-nearest neighbor classifier. *Neurocomputing* **2022**, *500*, 351–378. [[CrossRef](#)]