



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

Lei Shi, Jia Luo, Peiying Zhang, Hongqi Han, Didier El Baz, Gang Cheng,
Zeyu Liang

► To cite this version:

Lei Shi, Jia Luo, Peiying Zhang, Hongqi Han, Didier El Baz, et al.. Understanding User Preferences in Location-Based Social Networks via a Novel Self-Attention Mechanism. Sustainability, 2022, 14 (24), pp.16414. 10.3390/su142416414 . hal-04258231

HAL Id: hal-04258231

<https://hal.science/hal-04258231v1>

Submitted on 27 Oct 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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

Lei Shi^{1, 2}, Jia Luo^{3, 4*}, Peiying Zhang⁵, Hongqi Han², Didier El baz⁴, Gang Cheng⁶, Zeyu Liang⁷

¹ State Key Laboratory of Media Convergence and Communication, Communication University of China, Beijing, China

² Institute of Scientific and Technical Information of China, Beijing, China

³ College of Economics and Management, Beijing University of Technology, Beijing, China

⁴ LAAS-CNRS, Université de Toulouse, CNRS, Toulouse, France

⁵ College of Computer Science and Technology, China University of Petroleum (East China), Qingdao, China

⁶ School of Computer Science, North China Institute of Science and Technology, Beijing, China

⁷ School of Data Science and Intelligent Media, Communication University of China, Beijing, China

* Correspondence: Jia Luo, jialuo@bjut.edu.cn

Abstract: The check-in behaviors of users are ubiquitous in location-based social network during their urban living. Understanding user preferences is critical for improving the recommendation service of social platform. And great quality of recommendation is also beneficial to the sustainable urban living since the user can easily find the Point-of-Interest (POI) to visit, which avoiding unnecessary consumption, such as long time for searching or driving. To capture the user preferences from the check-in behaviors, advanced methods transform the history records into the graph structure data and further leverage graph deep learning-based techniques to learn the user preferences. Despite effectiveness, existing graph deep learning-based methods are limited to capture the deep graph structural information due to the inherent limitations, such as the over-smoothing problem in graph neural networks, further leading to the suboptimal performance. To address the above issues, we propose a novel method built on the Transformer architecture named Spatio-Temporal Aware Transformer (STAT) via a novel graphical-aware attention mechanism. In addition, a new temporal-aware sampling strategy is developed to reduces the computational cost and enables 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;

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 the location-based social networks [1] (LBSNs). Such a place that a user is interested to visit is called 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 a large number on the Web in a sustainable manner. All kinds of urban life and daily clothes not only enrich people's life, but also bring about the problem of "choice paralysis". Point of interest recommendation is a personalized recommendation based on context information and location awareness. Point of interest recommendation relates users and points of interest, aiming to recommend some new interesting locations for users' lives. Point of interest recommendation based on location-based social networks plays an important role in providing 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, 3, 4]. Great POI recommendation services can improve the urban livings of users since the recommender system can help users easily find their POIs, avoiding the long-time search, which is also beneficial for the 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, like movie recommendation and music recommendation.

Such a challenging task has attracted great attention in the field of recommender systems. Various techniques [5, 6, 7, 8, 9] have been proposed to enhance the performance for this task, such as matrix factorization-based methods and collaborative filtering-based methods. Among a diversity of techniques, deep learning-based methods has shown remarkable performance and exhibited high flexibility, which attracts 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 constructed graphs 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.

While the goal of graph neural network-based models [8] is to leverage the advanced graph deep learning techniques to learn the representations of users and POIs from both topology features and semantic features. Similar to the former one, methods in this category also utilize the generated graphs to capture the influences of different contextual factors. Benefiting from the message-passing mechanism, graph neural network-based models can extract more deep information from the graph 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 the node representations. For instance, most graph embedding-based methods regard the immediate neighbors as the positive nodes, which indicates that the information of immediate neighbors is meaningful. While graph neural network-based methods depend on the message-passing mechanism, which has been proved that it will suffer from the over-smoothing problem when obtaining the remote nodes.

In light of the above limitations, in this paper, we propose a novel recommendation model built on the Transformer architecture named Spatio-Temporal Aware Transformer (STAT). By introducing the self-attention mechanism, STAT can carefully capture the long-range dependencies via the full attention matrix. To enhance the expressiveness of the original self-attention mechanism, we propose a novel geographical-aware attention mechanism that integrates the geographical information into the attention matrix, which can better preserve the interactions between POIs. Moreover, for generalizing Transformer to large-scale social networks, we develop a temporal-aware node sampling strategy that utilizes the temporal factor to sample relevant POIs for model training. To validate the effectiveness of 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 geographical-aware attention mechanism to learn the representations of POIs for the recommendation of urban life-related services.
- We develop a temporal-aware sampling strategy that samples relevant POIs according to the check-in timestamp, which can carefully preserve 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 to have a meal, not a bar, to have 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 capture 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] construct 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] propose a collaborative recommendation model, which incorporates temporal information for generating the recommended POIs. Kefalas et al. [5] consider the temporal dimension and measure the impact of time on various time intervals for learning user preferences. Xie et al. [6] propose a unified graph model to explore the semantic vectors from the temporal context. Different from the above studies, Doan et al. [7] utilize the Long Short Term Memory (LSTM) recurrent neural network to simultaneously captures both the sequential and temporal features in users' representations. Gao et al. [9] propose 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] develop a spatio-temporal neural network framework to utilize the check-in history and social ties of users for recommending personalized POIs. Wang et al. [12] propose a Graph-enhanced Spatial-Temporal network that leverages the recurrent neural network (RNN) to learn user-specific temporal dependencies. Wang et al. [13] develop 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 the check-in behavior, POIs which are far from the current location will not be 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 provide meaningful POIs for users.

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

3. Preliminaries

In this section, we first introduce the definitions and notations in this paper. Then we provide the description of the 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 $X \in \mathbb{R}^{m \times d}$ where d denotes the dimension of the feature vector and m denote the number of POIs.

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

Definition 3 (Top- k POI recommendation): Given the history records C_u of user u , the goal of the 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 architecture [25] follows the structure of Auto-encoder which consists of Encoder and Decoder. In this paper, we utilize the Encoder of Transformer to learn the representations of POIs. Next, we introduce the components of Encoder.

The Encoder of Transformer 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 $H \in \mathbb{R}^{n \times d}$, the output of SHA is calculated as follows:

$$A = \rho\left(\frac{(HW_Q)(HW_K)^T}{\sqrt{d_w}}\right), \quad (1)$$

$$H_{out} = A(HW_V), \quad (2)$$

where W_Q , W_K and 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 A represents the attention matrix which preserves the interactions of all item-pairs.

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

$$H_{final} = \zeta_2(\sigma(\zeta_1(H_{out}))), \quad (3)$$

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 output of FFN module.

4. Methodology

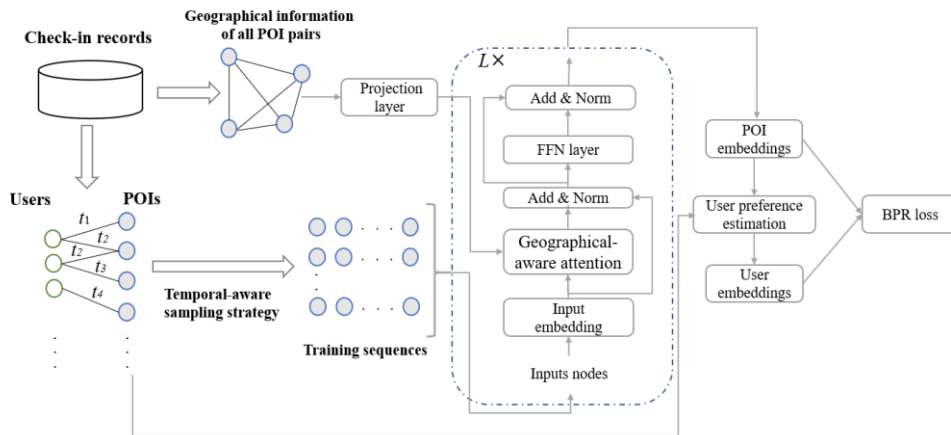


Figure 1. The overall framework of STAT.

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

4.1 Geographical-aware attention mechanism

As mentioned in Section 2.1, for each POI, we firstly construct the semantic features, resulting in a feature matrix X . Recent advanced graph deep learning-based methods transform the check-in records into the graph structural data, and further leverage the 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 capture the long-range dependencies of POIs due to the fixed graph structure.

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

However, the original self-attention mechanism is developed for learning the representations of words in a sentence, which does not carry extra information except semantic features. The POIs, discussed in Section 2.1, have complex information, such as semantic information and geographical information. And the geographical information is important to learn 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 geographical-aware attention mechanism. Intuitively, the relation of POIs is sensitive to the geographical distance of them. According to the Tobler’s first law of geography that “Everything is related to everything else, but near things are more related than distant things” [26], we calculate the distances of all POI pairs as a bias to strengthen the attention matrix to capture the geographical influence:

$$A_{i,j} = \frac{(X_i W_Q)(X_j W_K)^T}{\sqrt{d_w}} + \phi(\zeta(v_i, v_j)), \quad (4)$$

where X_i and 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 geographical factor-aware attention matrix, we can obtain the representations of POIs $H^V \in \mathbb{R}^{m \times d_w}$ according to Eq. (2) and Eq. (3).

4.2 Temporal-aware sampling strategy

Another limitation of generalize Transformer to the 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. And in the real-world applications, the number of POIs is usually very large, causing hard to train a Transformer-based method. On the other hand, not all POIs have close connections. As discussed in Section 1.1, POIs which 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 temporal-aware sampling strategy, which samples the similar POIs according to the check-in timestamps.

Specifically, for each POI v , we sample 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 sample several lists for POI v according the history check-in record of v .

In this way, the training cost of the self-attention mechanism can be reduced from $O(m^2)$ to $O(s^2)$, which is affordable for most training environments. More importantly, the sampling process can be

conducted before the model training stage, so that the mini-batch training technique can be employed, which guarantees the scalability of the proposed model.

Moreover, the temporal-aware sampling strategy can filter out irrelevant POIs based on the check-in history records, which also enhances the effectiveness of the self-attention mechanism and captures the influence of the temporal factor.

4.3 User preference estimation

Through the above components, we could obtain the representation vectors of POIs. Then we leverage them to estimate the representations of users. Given the fact that the visited POIs can reflect the user preference. For instance, if a user love food, there will be a large number of check-in records about 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 :

$$H_u^U = \sum_{v \in C_u} H_v^V, \quad (5)$$

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

$$H_u^U = \sum_{(v,t) \in C_u} \gamma(t-t_0) H_v^V, \quad (6)$$

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 dynamic user preference.

4.4 Parameter learning

In this paper, we adopt the widely-used optimization method, Bayesian Personalized Ranking [27], for learning the model parameters. Specifically, we construct the following objective function:

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

where v_j denote the random POI that user u have not visited. Θ represents the model parameters and μ denotes the regularization coefficient. $\psi(\cdot)$ denotes the sigmoid function. By minimizing Eq. (7), we can learn the model 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 estimate the user preference at timestamp t via Eq. (6). Then we calculate the interesting scores between the target user u and POIs that have not appeared in u 's check-in records. Considering the cost of check-in behaviors, we filter out the POIs whose are more than 200 kilometers away from the current location r which is almost a two-hour drive. Finally, we rank the POIs according to the interesting score and keep 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 adopt the widely-used real-world datasets [28], Foursquare [22] and Gowalla [22], for experiments.

The Foursquare dataset is collected from the famous location-based social network Foursquare which contains the 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 the category of POIs. The information of users' friendships is also involved in the dataset.

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

The statistics of datasets are reported in Table 1. In practice, we remove users and POIs which have less than 20 check-in records. We partition the datasets with the timeseries, and the former 80% of records are selected as the training set, followed recent works [28,30]. The rest are 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 choose 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 which jointly learn 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 [29]: STA is a translation-based embedding method which leverages the geographical information 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 introduce 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 geographical-temporal attention network to learn the representations of POIs from multi-contextual information and leverages a context-specific co-attention network to learn the user preferences.

GPR [30]: GPR is a graph neural network-based method that extract the user preference from the check-in graphs constructed by the geographical information. Specifically, GPR utilizes a graph auto-encoder to learn two types of geographical influences: ingoing influences and outgoing influences for learning the 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 topological structure and utilizes the Bi-directional Long Short-Term Memory to model users' sequential check-in behavior, which comprehensively considers the influence of temporal factors and geographical factors.

5.3 Evaluation metric

In this paper, we evaluate 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 vary k from {5, 10, 15, 20} in experiments to observe the performance of models.

5.4 Performance Comparison

In this section, we evaluate all models on Foursquare and Gowalla datasets via the mentioned evaluation metrics. Specifically, we run each model ten times with different random seeds and report the mean values of each evaluation metric. The results are shown in Figure 2 – Figure 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, 30] also have the similar observations. Besides, 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 can lead the better model performance than the message-passing mechanism. Moreover, we can also observe that GNN-based methods outperform embedding-based methods, which declares that introducing semantic features can learn more expressive representations of users and POIs.

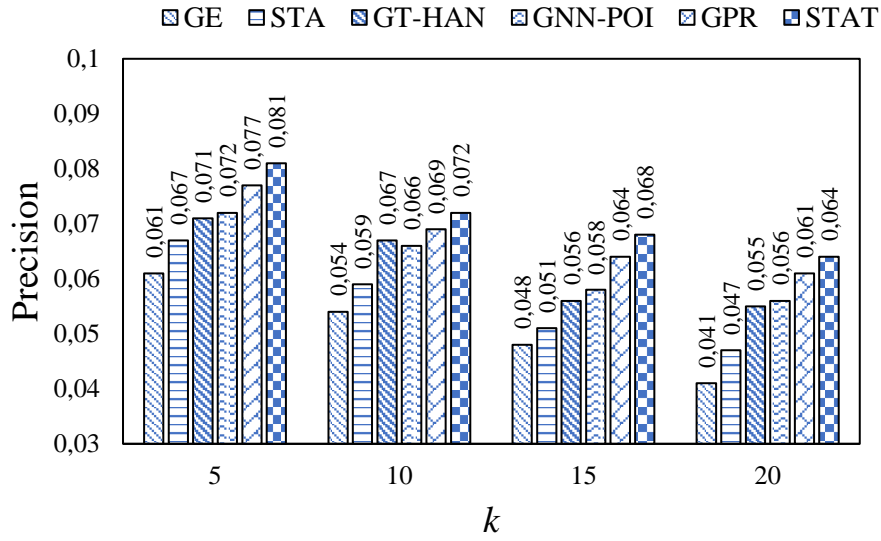


Figure 2. Performance of precision at k on Foursquare dataset.

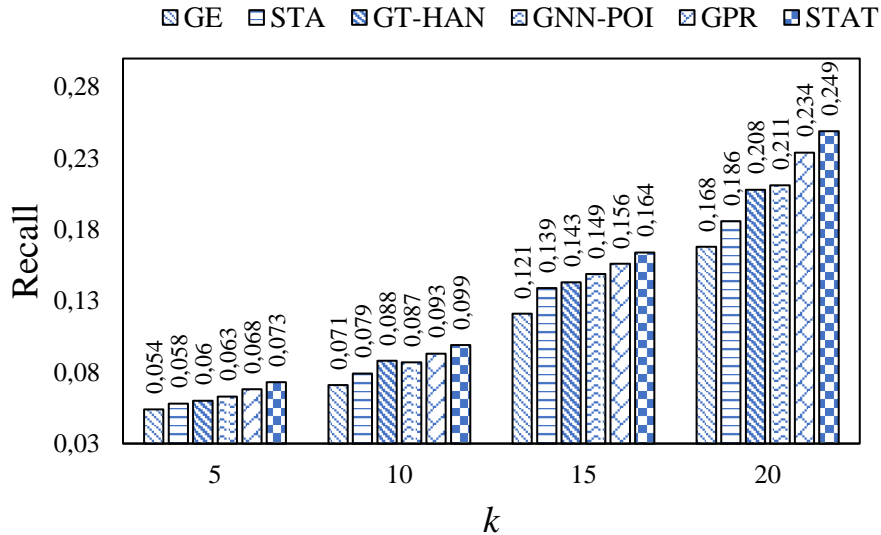


Figure 3. Performance of recall at k on Foursquare dataset.

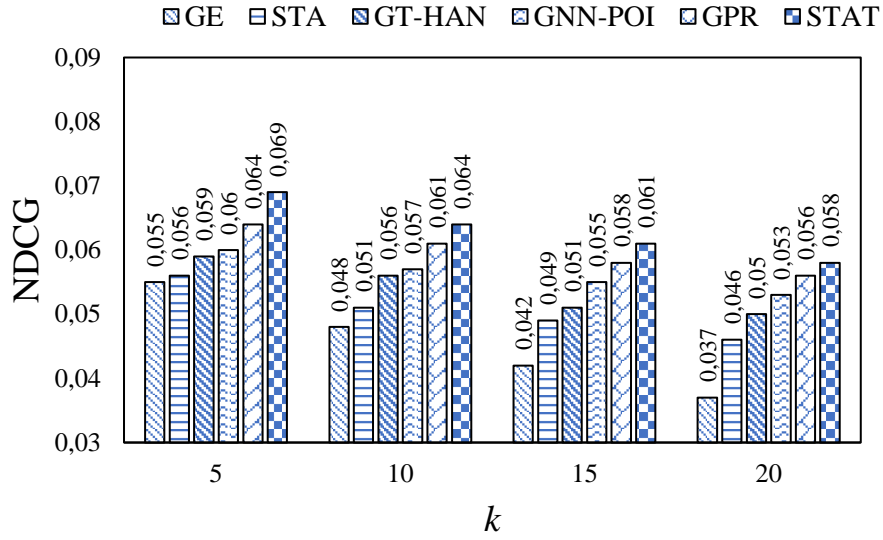


Figure 4. Performance of NDCG at k on Foursquare dataset.

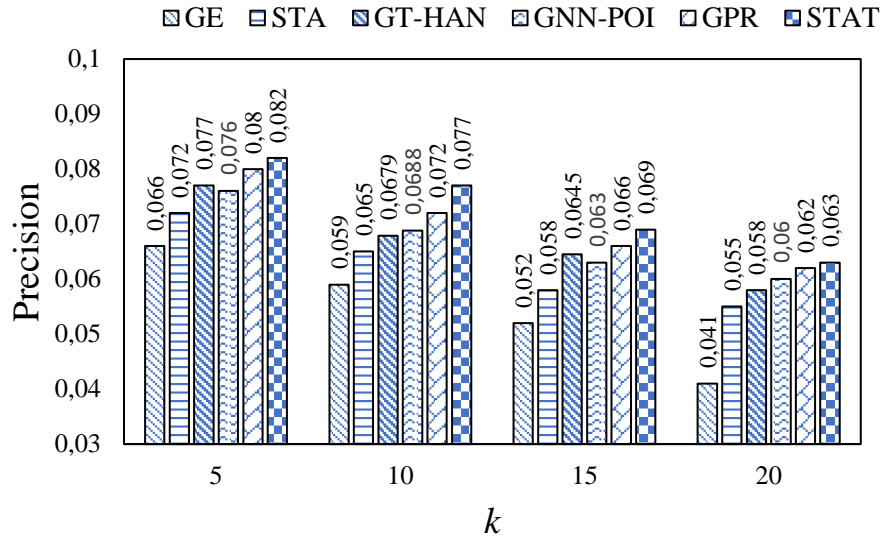


Figure 5. Performance of precision at k on Gowalla dataset.

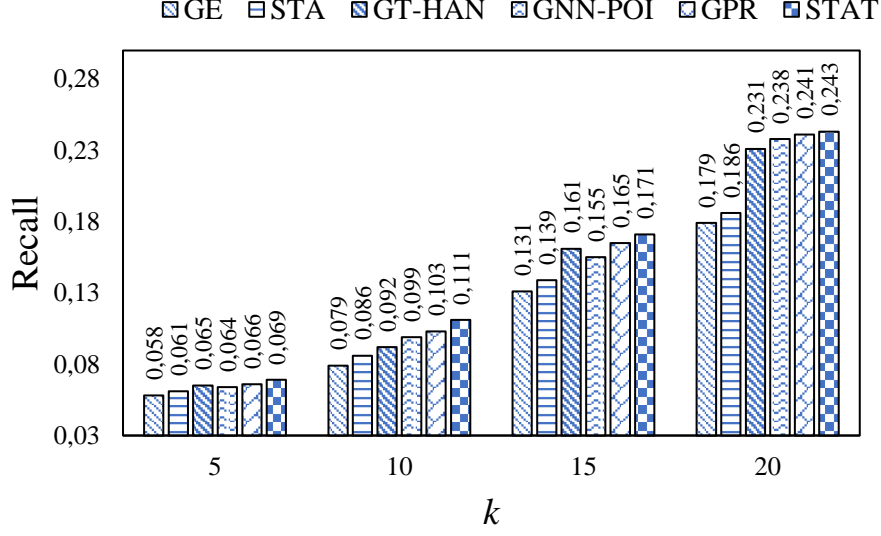


Figure 6. Performance of recall at k on Gowalla dataset.

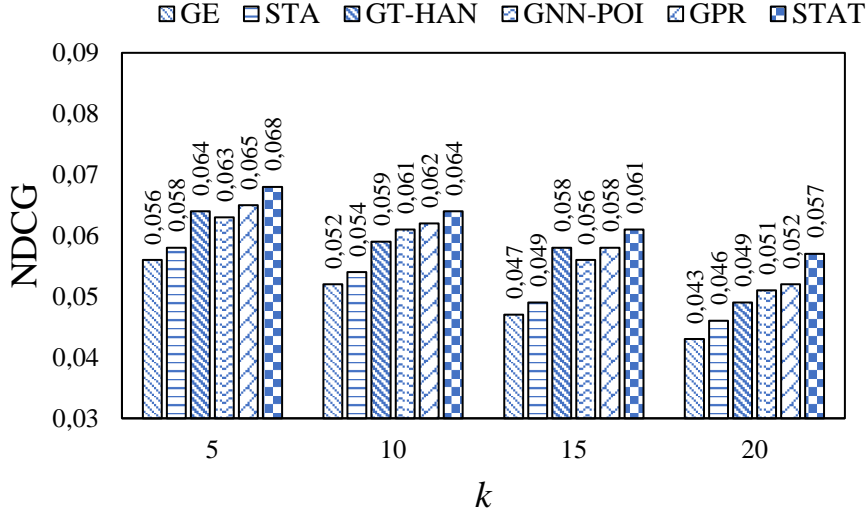


Figure 7. Performance of NDCG at k on Gowalla dataset.

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 can bring more information since the sampling list contains more nodes for aggregating the information. To validate the impact of s on the model performance, we conduct experiments on Foursquare dataset by fixing $k=20$ and varying s from $\{20, 40, 60, 80\}$. The results are reported in Table2. 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 because that when s is large, the sampling list may contain partial relevant nodes which could introduce the noisy information, further bothering the model performance. And a small value of s could achieve competitive performance. In this paper, we use the grid research method to determine the value of s on different datasets.

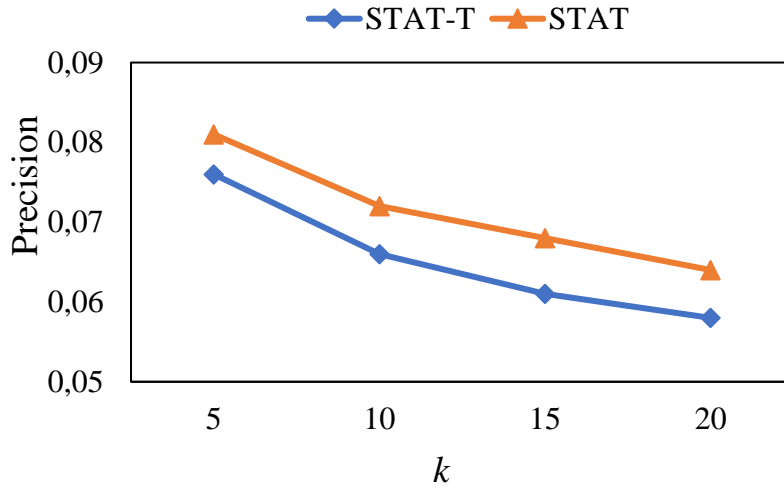
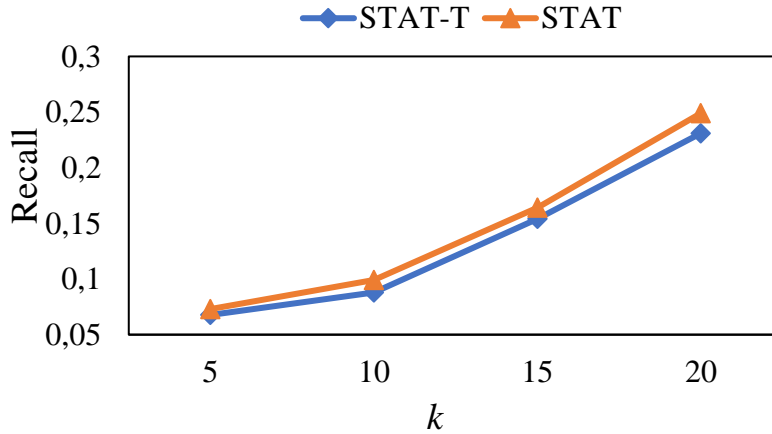
Table 2. The impact of s on Foursquare dataset

	$s=20$	$s=40$	$s=60$	$s=80$
Precision at 20	0.061	0.064	0.056	0.051
Recall at 20	0.234	0.249	0.211	0.196
NDCG at 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 3.3, STAT utilizes a time-aware method to calculate the user preference via Eq. (6), which can capture the dynamic user preferences. Hence, we provide a variant of STAT named STAT-T. In STAT, we use Eq. (5) to learn the user preference, which means that we ignore the influence of temporal factor on user preference. We run the above two methods on Foursquare dataset and the results are shown in Figure 8- Figure 10.

Intuitively, user preferences change over time. And the last check-in behaviors can better represent the user preferences. While early check-in records generally have little influence on the user preferences. From the results, we can 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 the temporal factor is beneficial for accurately obtaining user preferences.

**Figure 8.** Precision at k of STAT and its variant on Foursquare dataset.**Figure 9.** Recall at k of STAT and its variant on Foursquare dataset.

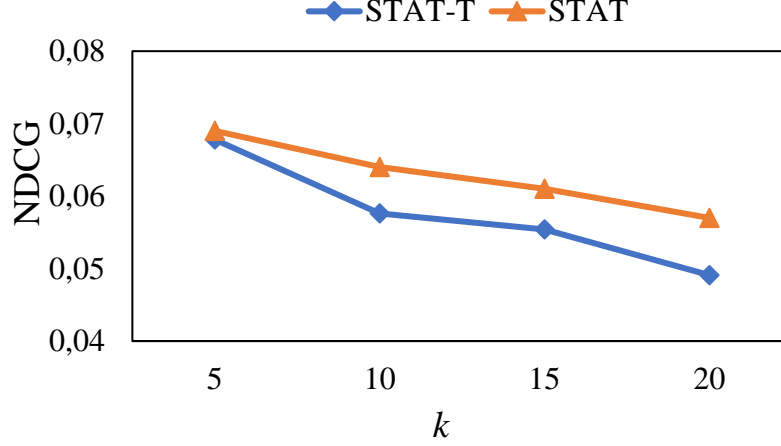


Figure 10. NDCG at k of STAT and its variant on Foursquare dataset.

5.7 Efficiency Study

In this section, we conduct experiments to analyze the training cost of our proposed STAT. Since baselines are implemented by different programming languages, such as the official implement of GE is based on C++. We only report the running time and the memory cost of STAT. And our implement is based on Python and Pytorch framework. The experiments are conducted on a Linux server with one I9-10900K CPU and one RTX 2080TI GPU. The original cost of Transformer is too large to afford. Benefited from our proposed temporal-aware sampling strategy, we can utilize 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	2,349
Gowalla	46.18	2,238

As mentioned in Section 3.2, the complexity of STAT is mainly related with the length of the sampling sequence. Hence, the memory cost can be reduced to an affordable value. And if we use the original Transformer model, we will meet the Out-of-Memory problem on these datasets due to the complexity square in the number of POIs.

6. Conclusion

In this paper, we propose STAT, a Transformer-based POI recommendation model, which takes the geographical factor and the temporal factor in account to learn the representations of POIs. Specifically, STAT develops a novel geographical-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 the large-scale location-based social network, STAT propose a temporal-aware sampling strategy that samples several relevant nodes based on the check-in timestamp. In this way, the influence of the temporal factor could be carefully preserved. Moreover, STAT develops a time-aware user preference estimation to capture the dynamic user preferences. To validate the effectiveness of STAT, we conduct extensive experiments on real-world datasets via three widely used evaluation metrics. The experimental results indicate that STAT consistently achieve highest values on different evaluation metrics compared to baselines. For instance, STAT obtains 7.3% of Recall on Foursquare when the length of recommendation list is 5, while the second performance is 6.8% from GPR. We also develop the efficiency study to analyze the training cost of STAT. And the results show that our proposed STAT have a good scalability for 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 consider the temporal factor and geographical factor. Other important factors, such as social information of users and category information of POIs, are not involved. One future work is to develop a reasonable framework to jointly model the above factors to better learn user preferences.

Funding: This work was supported by the Fundamental Research Funds for the Central Universities (No. CUC220C011, 300102262513), Youth Fund Project of Guangxi Natural Science Foundation (No. 2021JJB170032), 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), Natural Science Foundation of Hebei Province of China (No. D2022508002), National Key Research and Development Program of China (No.2022YFC3302103, 2020YFB2009304), 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>. The Gowalla dataset can be downloaded from the link <http://snap.stanford.edu/data/loc-gowalla.html>.

References

1. Shi L, Du J, Cheng G, et al. Cross-media search method based on complementary attention and generative adversarial network for social networks. *International Journal of Intelligent Systems*, 2021.
2. Christoforidis G, Kefalas P, Papadopoulos A N, et al. RELINE: point-of-interest recommendations using multiple network embeddings. *Knowledge and Information Systems*, 2021, 63(4): 791-817.
3. Shi L, Song G, Cheng G, et al. A user-based aggregation topic model for understanding user's preference and intention in social network. *Neurocomputing*, 2020, 413: 1-13.
4. Yuan Q, Cong G, Ma Z, et al. Time-aware point-of-interest recommendation. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*. 2013: 363-372.
5. Kefalas P, Manolopoulos Y. A time-aware spatio-textual recommender system. *Expert Systems with Applications*, 2017, 78: 396-406.
6. Xie M, Yin H, Wang H, et al. Learning graph-based poi embedding for location-based recommendation. In *Proceedings of the 25th ACM international on conference on information and knowledge management*. 2016: 15-24.
7. 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, 2019: 346-358.
8. Shu T, Shi L, Zhu C, et al. A graph neural network framework based on preference-aware graph diffusion for recommendation. *Frontiers in Psychiatry*, 2022, 13.
9. Gao H, Tang J, Hu X, et al. Exploring temporal effects for location recommendation on location-based social networks. In *Proceedings of the 7th ACM conference on Recommender systems*. 2013: 93-100.
10. Zhang J, Jiang Y, Wu S, et al. Prediction of remaining useful life based on bidirectional gated recurrent unit with temporal self-attention mechanism. *Reliability Engineering & System Safety*, 2022, 221: 108297.
11. Dai S, Yu Y, Fan H, et al. Spatio-Temporal Representation Learning with Social Tie for Personalized POI Recommendation. *Data Science and Engineering*, 2022, 7(1): 44-56.
12. Wang Z, Zhu Y, Zhang Q, et al. Graph-Enhanced Spatial-Temporal Network for Next POI Recommendation. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 2022, 16(6): 1-21.
13. Wang E, Jiang Y, Xu Y, et al. Spatial-Temporal Interval Aware Sequential POI Recommendation. In *Proceedings of the IEEE 38th International Conference on Data Engineering (ICDE)*. IEEE, 2022: 2086-2098.
14. Ye M, Yin P, Lee W C, et al. 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*. 2011: 325-334.
15. Sun K, Qian T, Chen T, et al. Where to go next: Modeling long-and short-term user preferences for point-of-interest recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence*. 2020, 34(01): 214-221.

16. Liu B, Fu Y, Yao Z, et al. Learning geographical preferences for point-of-interest recommendation. In Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining. 2013: 1043-1051.
17. Liu B, Meng Q, Zhang H, et al. VGMF: visual contents and geographical influence enhanced point-of-interest recommendation in location-based social network. Transactions on Emerging Telecommunications Technologies, 2022, 33(6): e3889.
18. Huo Y, Chen B, Tang J, et al. Privacy-preserving point-of-interest recommendation based on geographical and social influence. Information Sciences, 2021, 543: 202-218.
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. 2021: 130-137.
20. Zhang Y, Liu G, Liu A, et al. Personalized geographical influence modeling for POI recommendation. IEEE Intelligent Systems, 2020, 35(5): 18-27.
21. Liu T, Liao J, Wu Z, et al. Exploiting geographical-temporal awareness attention for next point-of-interest recommendation. Neurocomputing, 2020, 400: 227-237.
22. Islam M A, Mohammad M, Das S, et al. A survey on deep learning based Point-of-Interest (POI) recommendations. Neurocomputing, 2022, 472: 306-325.
23. Pablo Sánchez, Alejandro Bellogín. Point-of-Interest Recommender Systems based on Location-Based Social Networks: A Survey from an Experimental Perspective. ACM Computing Surveys, 2022, 54 (11s): 1-37.
24. Le Q, Mikolov T. Distributed representations of sentences and documents. In International conference on machine learning. PMLR, 2014: 1188-1196.
25. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. Advances in neural information processing systems, 2017, 30.
26. Liu B, Xiong H, Papadimitriou S, et al. A general geographical probabilistic factor model for point of interest recommendation. IEEE Transactions on Knowledge and Data Engineering, 2014, 27(5): 1167-1179.
27. Rendle S, Freudenthaler C, Gantner Z, et al. BPR: Bayesian personalized ranking from implicit feedback. arXiv preprint arXiv:1205.2618, 2012.
28. Zhang J, Liu X, Zhou X, et al. Leveraging graph neural networks for point-of-interest recommendations. Neurocomputing, 2021, 462: 1-13.
29. Qian T, Liu B, Nguyen Q V H, et al. Spatiotemporal representation learning for translation-based POI recommendation. ACM Transactions on Information Systems (TOIS), 2019, 37(2): 1-24.
30. Chang B, Jang G, Kim S, et al. Learning graph-based geographical latent representation for point-of-interest recommendation. In Proceedings of the 29th ACM International Conference on Information & Knowledge Management. 2020: 135-144.
31. Shi L, Du J, Liang M, et al. Dynamic topic modeling via self-aggregation for short text streams. Peer-to-peer Networking and applications, 2019, 12(5): 1403-1417.
32. Memiş S, Enginoğlu S, Erkan U. Fuzzy Parameterized Fuzzy Soft k-Nearest Neighbor Classifier. Neurocomputing, 2022, 500: 351-378.