*Electronic Research Archive*

ERA, 32(1): 486–504.
DOI: 10.3934/era.2024024
Received: 15 November 2023
Revised: 18 December 2023
Accepted: 22 December 2023
Published: 02 January 2024

http://www.aimspress.com/journal/ERA

*Research article*

# Learning user preferences from Multi-Contextual Sequence influences for next POI recommendation

**Jing Chen[1], Weiyu Ye[2,*] and Shaowei Kang[2]**

[1] Economic and Trade Department, Chongzuo Preschool Education College, Chongzuo 532200, China
[2] School of Electronic Engineering and Intelligence, Guangxi Vocational Institute of Technology, Chongzuo 532200, China

* **Correspondence:** Email: yeweiyuchj@sina.com.

**Abstract:** The recommendation of the next Point of Interest (POI) has attracted significant attention within the domain of POI recommendations in recent years. Existing methods for next POI recommendation are built on the original check-in sequences of users. Despite effectiveness, the original check-in sequences mix the influences of different contextual factors, which inevitably weakens the model ability of learning user preferences from the complex contextual information. To overcome this issue, we propose a novel Multi-Contextual Sequence-based Attention Network (MCSAN) for next POI recommendations. MCSAN first develops a new con-textual influence-based sampling strategy, which can transform the original check-in sequences into a series of contextual information-aware subsequences. Moreover, the constructed subsequences meticulously capture the impacts of various contextual information from the original check-in sequences. Then, MCSAN leverage the attention-based neural network to learn the representations of POIs from the generated subsequences. Finally, MCSAN develops a new feature fusion method that extracts user preferences from the learned POI presentations adaptively. Extensive experiments conducted on real-world datasets indicate the effectiveness of our proposed MCSAN for the next POI recommendation task, compared to recent representative methods.

**Keywords:** next POI recommendation; sequence; contextual information; attention

## 1. Introduction

The development of smartphones and global positioning systems has greatly improved the accuracy of location positioning. As a result, location-based social network (LBSN) [1–3] services like Gowalla, Yelp and Facebook, have gained immense popularity. LBSN offers a distinctive advantage over traditional social networks by allowing users to share their check-in information, such as visits to gyms and coffee shops, with their friends. Additionally, LBSN's Points of Interest (POI) recommendation [4,5] assists users in discovering relevant information from the extensive dataset of location-based social networks, enabling them to explore new places and enrich their daily experiences. Hence, the primary goal of the POI recommendation task is to suggest intriguing venues for users to explore next.

In this paper, we address the crucial task of next POI recommendations, which plays a vital role in the field of POI recommendation. The goal is to predict the user's next POI visit based on their recent check-in sequence, catering to their daily needs and preferences. The check-in sequence constructed from the user's recent history of check-in records. Extracting user check-in patterns from sequential information and multiple contextual factors, including geographical and temporal data, poses a major challenge in next POI recommendation.

In contrast to traditional POI recommendation task, the next POI recommendation scenario requires the recommendation model to generate a list of recommendations based on the recent user's check-in sequences. Hence, the next POI recommendation model is required to learn user preferences by the recent check-in sequences. Classic techniques for sequence analysis, such as Markov chain-based methods, recurrent neural network-based methods and graph neural network-based methods, are developed to extract the user preference from the check-in sequences.

Markov chain-based methods attempt to estimate the transition matrix that characterizes users' check-in patterns. However, the above methods encounter challenges due to the high data sparsity problem. The sparsity issue results in a significantly sparse transition matrix, subsequently impacting the overall performance of the model. Another category of existing methods for next POI recommendation is based on advanced deep learning techniques. These methods treat the check-in sequences of users as the input sequence from other domains, such as sentences in natural language processing, and further leverage the ability of recurrent neural networks to capture the inductive bias between the elements in the input sequence. Hence, these methods can effectively extract the user's check-in patterns from the POI check-in sequences. On the other hand, graph neural network-based methods utilize graph representation learning method for acquiring the representations of both POIs and users. Unlike previous methods extract the user preference from the check-in sequences, graph neural network-based methods begin by transforming the users' check-in sequences into graph structures. In this way, the relationship between POIs and users can be accurately preserved through the edges in the constructed graphs. Subsequently, graph neural network-based neural network blocks are developed to learn the representations of POIs and users from these generated graphs.

While recurrent neural network-based methods and graph neural network-based methods have proven effective, they rely solely on the original check-in sequence to learn user preferences. As a result, they overlook the impact of various contextual factors present in the check-in sequences, ultimately leading to suboptimal model performance. Intuitively, the original check-in sequence of a user may be too lengthy to accurately capture their unique preferences.

To address the aforementioned limitations, we present a novel method called the Multi-Contextual

Sequence-based Attention Network (MCSAN), which contains three key steps. In the first step, MCSAN develops a novel contextual influence-based sampling strategy, which can transform the original check-in sequences into a series of contextual information-aware subsequences. Therefore, our proposed method, MCSAN, is able to effectively preserve the influences of various contextual factors present in the lengthy check-in sequence. In the next step, MCSAN employs an attention-based neural network block to learn POI representations from different subsequences. Finally, MCSAN introduces a novel adaptive representation fusion method to dynamically learn the final user representations based on the extracted POI representations.

Compared to existing GNN- and recurrent neural network-based methods, MCSAN exhibits two major advantages: 1) MCSAN transforms the complex contextual information into separated contextual-based sequences, which are beneficial to learning representations from different context. 2) MCSAN develops a tailored Transformer-based neural network backbone with a learnable fusion layer, which enables the model to adaptively learn the final representations from the complex contextual information.

To evaluate the performance of our method for the next POI recommendation task, we conducted experiments using real-world datasets. Extensive experimental results have shown the effectiveness of MCSAN compared to other representative re-cent methods for next POI recommendation.

We summarize the major contributions of this paper as follows:

• We propose MCSAN, a novel method that can learn the representations of POIs and users from different contextual factors.

• We propose a new contextual influence-based sampling method that transform the original check-in sequences into a series of contextual information-aware subsequences, which exhibits significant differences with existing methods.

• We develop a novel adaptive representation fusion method, which allows for the adaptive learning of the ultimate representations based on the extracted POI representations for user.

• We conducted comprehensive experiments comparing MCSAN to recent representative methods for next POI recommendation to validate the effectiveness of our proposed method.

## 2. Related work

In the section, we provide a brief summary of recent studies of POI recommendation from two perspectives: General POI recommendation and next POI recommendation.

### *2.1. General POI recommendation*

The aim of POI recommendation is to generate a list of POIs that the user may find interesting based on their history of check-in records. Recent years, many techniques have been adopted and developed for this task, such as topic model-based methods [4–6], graph representation learning-based methods [7–10] and advanced neural net-work-based methods [11–13]. Ji et al. [4] propose a probabilistic generative model known as STARec. This model introduces the concept of distinguishing users' social preferences using the individual interests and combines them with individual activity preferences. By incorporating these factors, STARec aims to promote the accuracy for POI recommendations. Li et al. [14] propose the CTLM model, which addresses the issue of ill-matching. The model achieves this by separating city-specific features from common features shared using all cities. This approach enables users' interests in one city to be transferred to another, improving the recommendation accuracy across different cities. Guo et al. [15] propose Deep-PR, a deep distributed-learning-based method for POI recommendation. Deep-PR abstracts hidden feature components from both local and global subspaces, iteratively reoptimizes expressions of the feature space through propagation operations, and fine-tunes the feature spaces to achieve improved accuracy. Gan et al. [16] present a novel POI recommendation method that incorporates the spatio-temporal factor by constructing a unique hypersphere interest model for each user, which captures and depicts their preferences in a more accurate manner, enhancing the recommendation process. Hossain et al. [17] introduce Context-Aware Recency based Attention Network (CARAN), a recommendation method that incorporates weather information along with spatio-temporal context. CARAN focuses on recently visited places using attention mechanisms and enables interaction between non-adjacent check-ins through spatiotemporal matrices. Additionally, CARAN employs linear interpolation to ensure smooth representation of spatial distance in the recommendation process. Chen et al. [18] present a novel POI recommendation system that effectively captures and learns complex sequential transitions, which is achieved by incorporating time and distance irregularity, while dynamically weighting the decay values during the model learning process. Han et al. [19] propose AUC-MF, which focuses on two levels of geographic information: local similarity and global similarity. They further introduce innovative techniques to incorporate geographical information into AUC-MF, which includes sampling strategies to accelerate convergence and refining recommendations independently of model training.

### 2.2. *Next POI recommendation*

In contrast to general POI recommendation methods, the objective of next POI recommendation is to predict the next POI that users are likely to check in based on their previous check-in records. Hence, the sequential influence is a vital factor to understanding users' behaviors. Classic methods for modeling sequential factors, such as Markov chain [20–23] and recurrent neural network [24–27], have been introduced into the next POI recommendation task.

Wang et al. [28] utilizes graph embedding techniques to explicitly model complex geographical influences. Fang et al. [29] propose URPI-GRU, which consists of a short-term module and a long-term module. In the short-term module, the GRU model is utilized to learn the user's periodic and current preferences. In the long-term module, the user's long-term preference is extracted by considering K-nearest neighbor sequences. Zheng et al. [30] propose a memory augmented hierarchical attention network that integrates both short-term check-in sequences and long-term memories, which aims to capture complex spatiotemporal patterns of user movements by leveraging the combination of these two sources of information. Huang et al. [31] propose DAN-SNR, which addresses this gap by utilizing the self-attention mechanism to model both sequential and social influences within a unified framework. Lai et al. [32] propose the Multi-View Spatial-Temporal Enhanced Hypergraph Network, which simultaneously learns representations from both local and global perspectives. In the local view, they design a spatial-temporal enhanced graph neural network to capture and propagate spatial-temporal correlations based on user-POI interactions in an asymmetric manner. Lan et al. [33] propose a combination of the position-extended algorithm and the gated-deep network. They utilize a gated-deep network to capture long-term behavioral dependencies by generating auxiliary binary gates. Additionally, the position-extended algorithm is employed to enhance contextual interaction in recurrent neural networks.

Different from previous studies that consider only to extract the user preferences from the whole check-in sequence, our proposed MCSAN leverages a fine-grained manner by transforming the whole check-in sequence into contextual-aware subsequences. In this way, the proposed MCSAN can carefully preserve the influence of different context factors on the user preferences.

### 2.3. *Transformer related researches*

Transformer [34] has been the most successful deep learning [35] architecture in recent years. Due to its powerful modeling capacity, many researchers have applied this architecture in different domains, such as computer version [36,37], natural language processing [38,39] and graph mining [40,41]. For instance, Liu et al. [36] introduces a hierarchical design that divides the input image into non-overlapping patches at different scales. These patches are processed in a hierarchical manner, allowing the model to capture both local and global features effectively. Moreover, the proposed Swin Transformer demonstrates competitive performance on various computer vision tasks, including image classification and object detection. Due to the excellent performance of the transformer, in this paper, we introduce this architecture as the basic neural network backbone of our proposed method.

## 3. Background

In the section, we provide the preliminary knowledge of this paper, including key definitions and notations that are essential for understanding the subsequent content.

**Definition 1 (Users and POIs)**. The user and the POI are two vital elements of LBSNs. Let $U = \{u_1, \cdots, u_{n_u}\}$ denote the user set and $n_u$ is the total number of users. Let $P = \{p_1, \cdots, p_{n_m}\}$ denote the POI set and $n_m$ is the total number of POIs. Each POI $p$ is associated with additional information, such as geographical information $l_p$ and category information $c_p$. In this paper, we partition the entire geographical space into multiple blocks based on urban administrative divisions, yielding a set of distinct locations blocks $L = \{l_1, \cdots, l_{n_o}\}$, where $n_o$ denotes the number of location blocks. We also utilize $C = \{c_1, \cdots, c_{n_c}\}$ to represent the set of categories and $n_c$ is the number of categories.

**Definition 2 (A check-in record)**. A check-in record represents a check-in behavior of the target user, which is described by the triplet $(u, p, t)$. $u$ and $p$ represent the target user and POI, respectively, and$t$ denotes the check-in time. In this paper, we utilize $T = \{t_1, \cdots, t_{n_t}\}$ to represent the set of check-in timestamps.

**Definition 3 (A check-in sequence)**. A check-in sequence $S_u = \{(u_1, p_1, t_1), \cdots, (u_{q_u}, p_{q_u}, t_{q_u})\}$ is used to describe a series of check-in behaviors of the user $u$, where $q_u$ represents the sequence length of $S_u$.

**Definition 4 (Next POI recommendation)**. Given the user's previous check-in sequence, the objective of the next POI recommendation is to generate a list of POIs $\{p_1, \cdots, p_k\}$ that the user $u$ may have an interest in visiting at the next timestamp. $k$ represents the length of the recommendation list.

In the rest of the paper, we adopt uppercase to represent the sets used in the paper, and lowercase is adopted to represent the elements of the set. For matrices and vectors, we adopt the bold uppercase and lowercase to represent them, respectively. We summarize the key notations with their descriptions in Table 1.

**Table 1.** Summary of the key notations.

| Notation | Description |
|---|---|
| $U$, $u$ | The set of users, a user |
| $P$, $p$ | The set of POIs, a POI |
| $L$, $l$ | The set of location blocks, a location |
| $C$, $c$ | The set of categories, a category |
| $T$, $t$ | The set of location timestamps, a timestamp |
| $S_u$ | The check-in sequence of user $u$ |
| $s_u^{c,p}$ | The semantic-aware sequence of user $u$ |
| $s_u^{t,p}$ | The temporal-aware sequence of user $u$ |
| $s_u^{g,p}$ | The geographical-aware sequence of user $u$ |
| $\boldsymbol{H}_p^{c,u}$ | The representations of POIs from $s_u^{c,p}$ |
| $\boldsymbol{H}_p^{t,u}$ | The representations of POIs from $s_u^{t,p}$ |
| $\boldsymbol{H}_p^{g,u}$ | The representations of POIs from $s_u^{g,p}$ |
| $\boldsymbol{H}_u^{U}$ | The ultimate representations of user $u$ |
| $\boldsymbol{H}_p^{P}$ | The final representations of POI $p$ |
| $\delta(\cdot)$ | The linear layer |
| $\rho(\cdot)$ | The softmax function |
| $\alpha^{u,c}$ | The fusion weight for $\boldsymbol{H}_p^{c,u}$ |
| $\alpha^{u,t}$ | The fusion weight for $\boldsymbol{H}_p^{t,u}$ |
| $\alpha^{u,g}$ | The fusion weight for $\boldsymbol{H}_p^{g,u}$ |

## 4. Multi-Contextual Sequence-based Attention Network

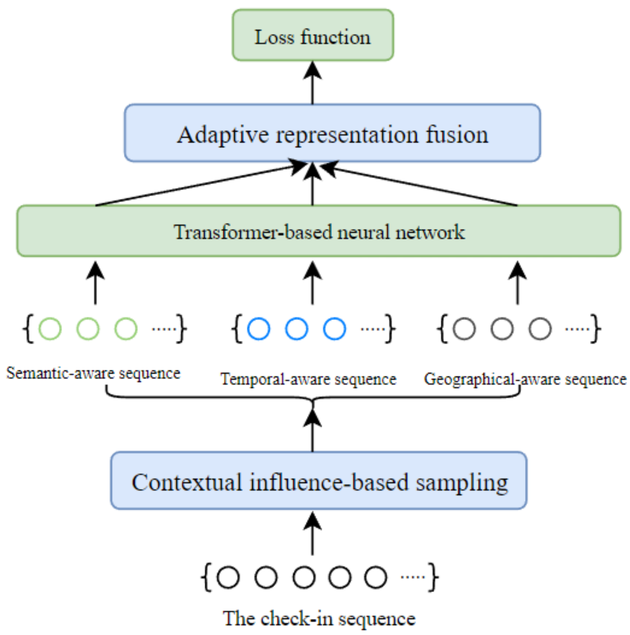


**Figure 1.** The overall framework of MCSAN.

In this section, we introduce our proposed Multi-Contextual Sequence-based Attention Network (MCSAN). Specifically, MCSAN contains three modules, contextual influence-based sampling module, Transformer-based neural network module and adaptive representation fusion module. Figure 1 shows the framework of MCSAN.

*4.1. Contextual influence-based sampling*

As mentioned before, the check-in sequence $S_u$ contains the recent check-in behaviors of the target user $u$, which is crucial for understanding the user preferences. Recent works pay attention to leveraging the check-in sequence to extract the user representations for the recommendation. However, they tend to leverage the full sequence to learn user preferences, which inevitably ignores the contextual information involved in the previous check-in records. Moreover, the full check-in sequence is too long to precisely learn the relations between visited POIs, leading to learning the suboptimal user representations.

To gain a comprehensive understanding of user preferences by the check-in sequence, we introduce a novel strategy known as contextual influence-based sampling (CIS). The goal of CIS is to sample several key POIs from the full check-in sequence according to different contextual information, resulting in different subsequences. These subsequences convey different contextual influences, which reflect user preferences in different aspects of contextual factor, such as temporal factor and geographical factor. Compared to the full sequence, a series of subsequences is more beneficial for learning user preferences.

Hence, we propose CIS to transform the full sequence into contextual influence-aware subsequences. In this paper, we consider three contextual factors, temporal factor, geographical factor and semantic factor. CIS samples different nodes to construct the subsequence for the corresponding contextual factor. Since CIS utilizes the same strategy to generate the subsequence for different contextual information, we take the semantic factor-aware subsequence generation for example to detail our proposed CIS.

In this paper, we leverage the category information to capture the semantic information of POIs. Given the check-in sequence $S_u$ of the user $u$, we first transform it into the semantic aware subsequence $s_u^c$:

$$s_u^c = \phi(s_u) \tag{1}$$

where $\phi(\cdot)$ is the transformation function. $s_u^c = \{c_1^u, \cdots, c_{|s_u^c|}^u\}$ denotes the transformed semantic aware subsequence and $c_i^u \in s_u^c$ denotes the category information of the visited POI. The operation of $\phi(\cdot)$ is to merge the adjacent POIs that belong to the same category in the check-in sequence and replace them with the category information, resulting the semantic aware check-in subsequence. Hence, we can also obtain the category-POI matrix $\boldsymbol{M}_u^c \in \mathbb{R}^{|s_u^c| \times |s_u|}$ to describe the relations between categories and merged POIs according to the check-in records of user $u$. In practice, we adopt the contextual-aware sampling strategy according to the check-in timestamps, location or category information. We select POIs belong to similar check-in timestamps, locations or categories to construct the semantic-aware subsequences. Then, we utilize the weighted sampling method to construct the semantic aware check-in sequences of POIs. Since each element in the sequence could contains several

POIs, we need to define a sampling strategy to construct the POI sequence. In this paper, we adopt the weighted sampling described as follows:

$$Pro_p = \frac{d_p}{\sum_{p_i \in s_u} d_{p_i}} \tag{2}$$

where $d_p$ denotes the frequency of POI $p$ appearing in the check-in sequence $s_u$. Intuitively, higher frequency represents higher interest. Hence, such POIs should be more suitable to be sampled for learning user preferences. After the sampling process, we can obtain the set of subsequences involved corresponding POIs $s_u^{c,p}$. Compared to the original check-in sequence $s_u$, $s_u^{c,p}$ is more precisely to represent the user preferences from the perspective of the semantic contextual information. Followed the same strategy, we can obtain the temporal subsequences $s_u^{t,p}$ and the geographical subsequences $s_u^{g,p}$.

*4.2. Transformer-based neural network*

The constructed subsequences $s_u^{c,p}$, $s_u^{t,p}$ and $s_u^{g,p}$ describe the check-in patterns of the user $u$ from the perspective of different contextual information. Next, we need to extract the presentations of POIs from the generated subsequences. MCSAN develops an attention-based neural network block to learn the hidden representations of POIs. Specifically, the self-attention mechanism of the Transformer architecture [34] is adopted for the feature learning process. To facilitate the explanation of the attention-based neural network block, we illustrate the processing of semantic-aware subsequences as an example.

For the input feature matrix $\boldsymbol{H}^{c,u} \in \mathbb{R}^{|s_u^{c,p}| \times d}$, where $|s_u^{c,p}|$ is the length of the input sequence and $d$ denotes the dimension of the input feature, we first leverage the projection layer to transform the original input features into hidden features:

$$\boldsymbol{H}^{c,0} = \delta(\boldsymbol{H}^{c,u}) \tag{3}$$

where $\delta(\cdot)$ denotes the linear layer. Next, we employ the self-attention mechanism to learn the inductive biases between POIs in the sequence:

$$\boldsymbol{H}^{c,i+1} = \rho\left(\boldsymbol{Q}^i \boldsymbol{K}^{iT}\right) \cdot \boldsymbol{V}^i \tag{4}$$

where $\rho(\cdot)$ denotes the softmax function. $\boldsymbol{Q}^i$, $\boldsymbol{K}^i$ and $\boldsymbol{V}^i$ are transformed feature matrices which are calculated as follows:

$$\boldsymbol{Q}^i = \boldsymbol{H}^{c,i} \cdot \boldsymbol{W}^{q,i}, \boldsymbol{K}^i = \boldsymbol{H}^{c,i} \cdot \boldsymbol{W}^{k,i}, \boldsymbol{V}^i = \boldsymbol{H}^{c,i} \cdot \boldsymbol{W}^{v,i} \tag{5}$$

where $\boldsymbol{W}^{q,i}$, $\boldsymbol{W}^{k,i}$ and $\boldsymbol{W}^{v,i}$ are learnable parameter matrices of the $i$-th neural network layer. While we also introduce the feedforward neural network-based block [34] to enhance the representation learning of the self-attention layer.

Through several Transformer layers, we can obtain the POI representations $\boldsymbol{H}_p^{c,u}$ extracted from the semantic-aware sequence. Following the same neural network architecture, we can obtain $\boldsymbol{H}_p^{t,u}$ and $\boldsymbol{H}_p^{g,u}$ from the temporal-aware sequence and geographical-aware sequence, respectively.

*4.3. Adaptive representation fusion module*

Based on the Transformer layer-based neural network block, we extract the POI representations from different contextual information. Finally, we have to estimate the user preferences according the POI representations. The naïve strategy is to directly sum or concat the representations of POIs. However, this simple strategy ignores the unique user preference. The influences of different contextual factors vary for different users. This phenomenon motivates us to develop an adaptive representation fusion module to estimate the user representations.

Specifically, we utilize the following the strategy to adaptively calculate the fusion weights:

$$\alpha^{u,c} = \frac{exp\left(\boldsymbol{H}_p^{c,u}\boldsymbol{H}_p^{c,u^{\top}}\right)}{exp\left(\boldsymbol{H}_p^{c,u}\boldsymbol{H}_p^{c,uT}\right)+exp\left(\boldsymbol{H}_p^{t,u}\boldsymbol{H}_p^{t,uT}\right)+exp\left(\boldsymbol{H}_p^{g,u}\boldsymbol{H}_p^{g,u^{\top}}\right)} \tag{6}$$

where $\alpha^{u,c}$ is the fusion weight for $\boldsymbol{H}_p^{c,u}$. Following the same strategy, we can obtain the fusion weights $\alpha^{u,t}$ and $\alpha^{u,g}$ for $\boldsymbol{H}_p^{t,u}$ and $\boldsymbol{H}_p^{g,u}$. Finally, the preference of user $u$ is estimated as follows:

$$\mathbf{H}_u^U = (\alpha^{u,c} \cdot \frac{1}{|s_u^{c,p}|}\sum_{p_i \in s_u^{c,p}} \mathbf{H}_{p_i}^{c,u}) + (\alpha^{u,t} \cdot \frac{1}{|s_u^{t,p}|}\sum_{p_j \in s_u^{t,p}} \mathbf{H}_{p_j}^{t,u}) + (\alpha^{u,g} \cdot \frac{1}{|s_u^{g,p}|}\sum_{p_k \in s_u^{g,p}} \mathbf{H}_{p_k}^{g,u}) \tag{7}$$

Additionally, for the representations of POI $p$, we fuse the representations from different contextual factors to obtain the final representation $\boldsymbol{H}_p^P$.

The computational complexity of MCSAN is mainly from the calculation of the Transformer layers. Suppose each contextual sequence has the same length $n_s$, and the complexity of MCSAN could be roughly represented in $O((n_s)^2 ld)$ where $l$ represents the layers of Transformer backbone and $d$ denotes the dimension of hidden representations.

*4.4. Parameter estimation*

Following prior studies [32,42], we utilize the cross-entropy loss function to train the model parameters. Specifically, we compute the recommendation scores between user $u$ and the unvisited POIs using the following approach:

$$\boldsymbol{Y}_u^{'} = \tau\left(\boldsymbol{H}_u^U \cdot \boldsymbol{H}_{un}^{P^{\top}}\right) \tag{8}$$

where $H_{un}^P$ denotes the feature matrix of the unvisited POIs and $\tau(\cdot)$ denotes the softmax function. Moreover, $\boldsymbol{Y}_u^{'}$ denotes the matrix of recommendation scores for each unvisited POI.

The objective function of MCSAN is as follows:

$$Loss_{MCSAN} = -\sum_{u \in U} \boldsymbol{Y}_u log \boldsymbol{Y}'_u + \eta \left|\left|\boldsymbol{\theta}\right|\right|^2 \tag{9}$$

where $\boldsymbol{Y}_u$ denote the POIs that the user $u$ has visited, $\boldsymbol{\theta}$ represents the parameters and $\eta$ represents the regularization weight. By minimizing the objective function, we can adopt the stochastic gradient descent method to learn the model parameters.

### *4.5. Implementation of the proposed method*

Here, we provide the implementation of MCSAN. As mentioned before, for the input check-in sequence of a user, MCSAN first transforms the sequence into contextual-aware subsequences via the proposed transformation strategy. Then, MCSAN leverages the standard Transformer layer to learn the representations from the constructed sequences. Finally, MCSAN adopts a learnable fusion layer to adaptive extract the final user representations from the various contextual-aware sequences. The overall learning algorithm for MCSAN is outlined in Algorithm 1.

---

Algorithm 1: learning algorithm of MCSAN

---

Input: The check-in sequence *S*, the regularization weight η, the max training step $T_{max}$.
Output: Model parameters $\boldsymbol{\theta}$.

1. For each user, generate the contextual factor-aware subsequences $s_u^{c,p}$, $s_u^{t,p}$ and $s_u^{g,p}$ according to Eqs (1) and (2);
2. **for** step in $T_{max}$ **do:**
3. Calculating the hidden representations of POIs according to Eq (4);
4. Estimating the user preferences according to Eq (7)
5. Calculating the training loss according to Eq (9);
6. Updating the parameters $\boldsymbol{\theta}$;
7. **end for**
8. return $\boldsymbol{\theta}$.

---

## 5. Experiments

In this section, we conduct comprehensive experiments on real-world datasets to evaluate the effectiveness of our proposed MCSAN. We begin by introducing the fundamental experiment settings, which include the datasets used, baselines considered, evaluation metrics employed and parameter configurations. Next, we present the experimental results comparing our model with other existing models on the selected datasets. Last, we perform ablation studies to show the effectiveness of our proposed method.

### *5.1. Datasets*

Following recent work [43], we adopt the check-in records of users of three cities collected from the famous LBSN Foursquare, called Phoenix, New York and Singapore. The Foursquare dataset typically refers to the data available through Foursquare's API, which includes information about venues, user check-ins, user-generated tips and reviews and other location-related details. Each dataset contains check-in records from users in their respective cities. The dataset statistics are summarized in

Table 2. For the task setting, we adopt 8:1:1 split for all datasets according to the check-in timestamp for each user. While the earliest 80% is the training set, the latter 10% is the validation set and the rest is the test set.

**Table 2.** Statistics of datasets.

| Dataset | #Users | #POIs | #Records |
|---|---|---|---|
| Phoenix | 2946 | 7247 | 47,980 |
| New York | 16,387 | 56,252 | 511,431 |
| Singapore | 8648 | 33,712 | 355,337 |

*5.2. Baselines*

We choose the baselines from three aspects: Markov chain-based methods, recur-rent neural network-based methods and graph neural network-based methods.

For Markov chain-based methods, we choose TAD-FPMC [20] which is a typical method built on the Markov chain to compute the transition matrix to forecast the next check-in POI for users; For recurrent neural network-based methods, we select NeuNext [24] and LSPL [25], which leverage the recurrent neural network or its variants to learn the user preferences; For graph neural network-based method, we select MSTHN [32] which integrates the graph neural networks to extract the representations of POIs and users. We introduce the detailed information of these methods as follows:

TAD-FPMC: Decomposes the user-POI interaction tensor, which considers the timing of user interactions with POIs into lower-dimensional representations to extract latent user preferences.

NeuNext: Introduces the RNN-based neural network backbone to learn the user preferences from the constructed check-in sequences.

LSPL: Involves capturing historical behavior to understand long-term preferences and incorporating contextual information to model short-term preferences.

MSTHN: Introduces hypergraph networks to model relationships between different entities by leveraging the function of hypergraphs, which offers a more expressive way to capture complex relationships than traditional graphs.

*5.3. Evaluation metrics*

In this paper, we utilize two widely adopted metrics, Recall and Normalized Discounted Cumulative Gain (NDCG), to evaluate the performance of all models for the task of next POI recommendation. Each evaluation metric is calculated as follows:

$$Recall@k = \frac{|R_k \cap R_{test}|}{|R_{test}|} \tag{10}$$

$$NDCG@k = \frac{1}{|R_{test}|} \sum_{(u,p)\in R_{test}} \frac{\log(opt+1)}{\log(rank(u,p)+1)} \tag{11}$$

where $R_{test}$ denotes the test set and $R_k$ denotes the POI recommendation list with length *k*. *opt* denotes the optimal rank of the predicted POI which is set 1 and $rank(u,p)$ denote the rank of the POI $p$ in the recommendation list of user $u$.

*5.4. Parameter settings*

For the chosen baseline models, we adopt their open-resource implementations and follow their recommended parameter settings. We also conduct the parameter tuning operation for all models on the datasets. Furthermore, for our proposed method, we try the hidden dimension of neural network in $\{64, 128, 256\}$, learning rate in $\{0.01, 0.005, 0.001\}$ and the regularization coefficient in $\{0.0001, 0.00001\}$. In practice, we use the grid-search method to determine the optimal values of hyperparameters.

*5.5. Performance comparison*

We execute each method 10 times using independent random seeds and present the average values of each evaluation indicator to ensure a more robust comparison. The results are illustrated in Figures 2–7.

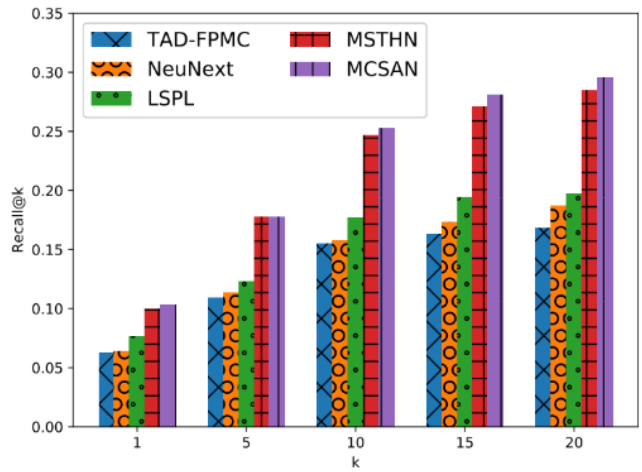


**Figure 2.** The overall framework of MCSAN.

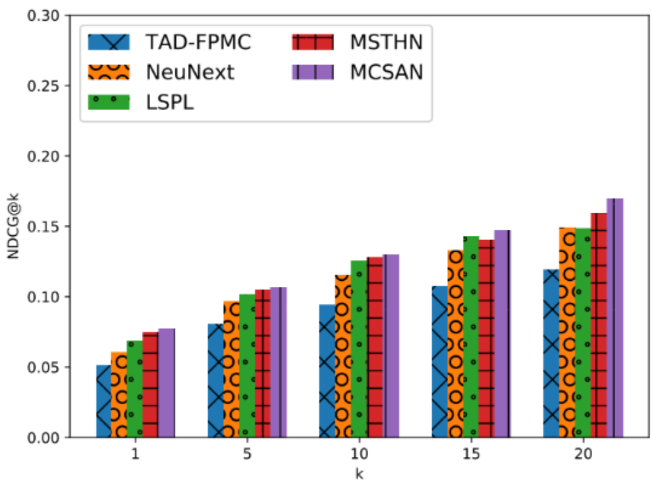


**Figure 3.** The performance on Phoenix in terms of NDCG.

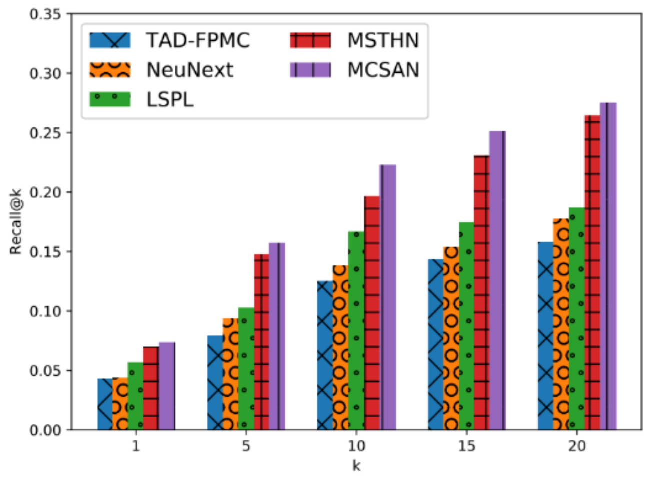


**Figure 4.** The performance on New York in terms of Recall.

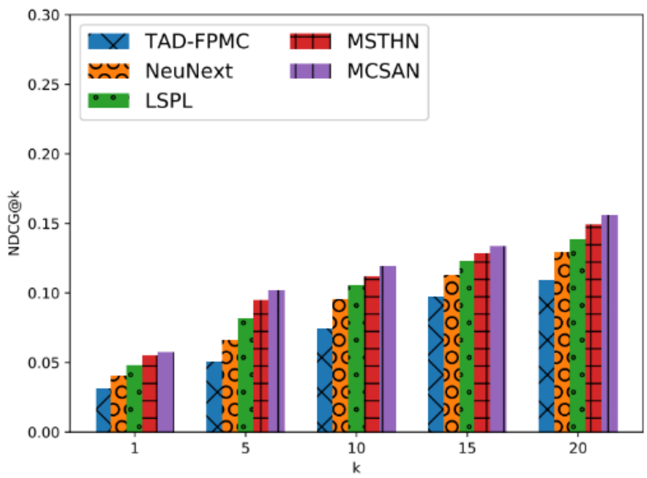


**Figure 5.** The performance on New York in terms of NDCG.

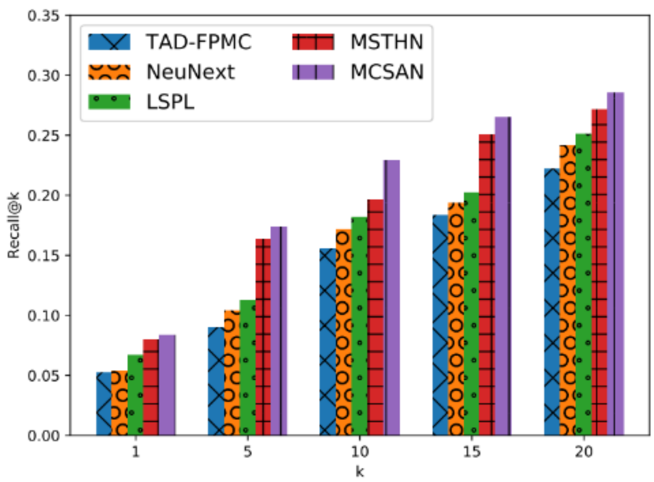


**Figure 6.** The performance on Singapore in terms of Recall.

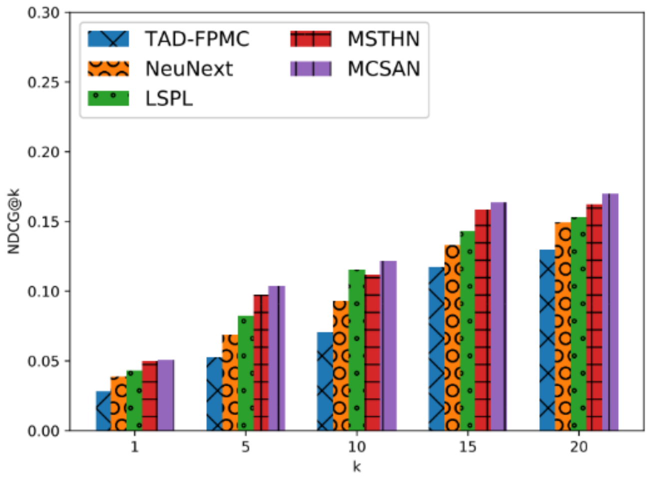


**Figure 7.** The performance on Singapore in terms of NDCG.

By the experimental results across all datasets, we can obtain the following observations: Overall, our proposed MCSAN consistently outperforms other methods on all datasets, particularly in comparison to recurrent neural network-based methods that consider the entire check-in sequences of users as input. This suggests the effectiveness of our proposed designs, namely contextual influence-based sampling and adaptive feature fusion. We can also observe that MCSAN achieve the lowest performance on the New York dataset, compared to other datasets. This is because New York dataset exhibits worst data sparsity problem which largely influences the model performance. In general, graph neural network-based methods tend to exhibit better performance compared to recurrent neural network-based approaches. This could be attributed to the fact that transforming the original check-in sequences into graph structural data may facilitate the learning of user preferences, as opposed to treating the check-ins as a single long sequence.

### *5.6. Ablation study*

To further validate the effectiveness of our proposed adaptive feature fusion module, we conducted additional experiments. Specifically, we introduced two variants of MCSAN, namely MCSAN-S and MCSAN-C. MCSAN-S equally aggregates the representations of POIs from various contextual information to derive the final node representation. On the other hand, MCSAN-C utilizes vector concatenation to fuse the POI representations and calculate the final user preferences. We kept the recommendation list length fixed at 10 and compared these variants with MCSAN across all datasets. The results are presented in Tables 3–5.

**Table 3.** The performance of MCSAN and its variants on Phoenix.

| Model | Recall | NDCG |
|---|---|---|
| MCSAN-S | 0.2467 | 0.1271 |
| MCSAN-C | 0.2489 | 0.1278 |
| MCSAN | 0.2525 | 0.1297 |

**Table 4.** The performance of MCSAN and its variants on New York.

| Model | Recall | NDCG |
|---|---|---|
| MCSAN-S | 0.2186 | 0.1169 |
| MCSAN-C | 0.2201 | 0.1176 |
| MCSAN | 0.2225 | 0.1187 |

**Table 5.** The performance of MCSAN and its variants on Singapore.

| Model | Recall | NDCG |
|---|---|---|
| MCSAN-S | 0.2210 | 0.1182 |
| MCSAN-C | 0.2234 | 0.1193 |
| MCSAN | 0.2292 | 0.1217 |

Based on the aforementioned results, it is evident that our proposed adaptive feature fusion module effectively improves the model performance across all real-world datasets, showing the effectiveness of our design. Furthermore, it can be observed that MCSAN-C surpasses MCSAN-S on all datasets. This observation suggests that the vector concatenation operation is better suited for learning user preferences than the vector summation operation.

## 6. Conclusions

In this paper, we propose a novel method called MCSAN for the task of next POI recommendation. Unlike existing methods that rely on the complete check-in sequences of users to learn their preferences, MCSAN introduces a new module called CIS. This module transforms the full check-in sequence into contextual information-aware subsequences, allowing MCSAN to effectively capture the influences of different contextual factors present in the original check-in sequence. MCSAN then adopts an attention-based neural network block to learn representations of POIs from these constructed subsequences. Hence, MCSAN can effectively capture the contextual information and its impact on user preferences. Furthermore, MCSAN incorporates an adaptive feature fusion module that dynamically adjusts the aggregation weights of different contextual factors based on the check-in patterns of individual users. This approach preserves the unique user preferences by giving more weight to the relevant contextual factors for each user. We conducted extensive experiments on three real-world datasets to evaluate the performance of MCSAN. The experimental results demonstrate the effectiveness of our proposed method in next POI recommendation, as compared to recent state-of-the-art approaches. As discussed before, MCSAN can adaptively learn the user preferences from different contextual factors. However, our proposed MCSAN highly dependents on the sequence construction. Hence, the construction strategy and the sampling method may have large influence on the model performance. In the future work, we will develop more suitable sequence construction and sampling strategies for the POI recommendation task.

## Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

## Acknowledgments

This work is supported by the 2022 Guangxi University Young and Middle School Teachers' Basic Scientific Research Ability Improvement Project (No. 2022KY1886).

## Conflict of interest

The authors declare that there are no conflicts of interest.

## References

1. Y. Hwangbo, K. J. Lee, B. Jeong, K. Y. Park, Recommendation system with minimized transaction data, *Data Sci. Manage.*, **4** (2021), 40–45. https://doi.org/10.1016/j.dsm.2022.01.001
2. L. Shi, J. Luo, C. Zhu, F. Kou, G. Cheng, X. Liu, A survey on cross-media search based on user intention understanding in social networks, *Inf. Fusion*, **91** (2023), 566–581. https://doi.org/10.1016/j.inffus.2022.11.017
3. L. Shi, J. Du, G. Cheng, L. Xia, Z. Xiong, J. Luo, Cross-media search method based on complementary attention and generative adversarial network for social networks, *Int. J. Intell. Syst.*, **37** (2022), 4393–4416. https://doi.org/10.1002/int.22723
4. W. Ji, X. Meng, Y. Zhang, STARec: Adaptive learning with spatiotemporal and activity influence for POI recommendation, *ACM Trans. Inf. Syst.*, **40** (2021), 1–40. https://doi.org/10.1145/3485631
5. W. Ji, X. Meng, Y. Zhang, SPATM: A social period-aware topic model for personalized venue recommendation, *IEEE Trans. Knowl. Data Eng.*, **34** (2020), 3997–4010. https://doi.org/10.1109/TKDE.2020.3029070
6. L. Shi, G. Song, G. Cheng, X. Liu, A user-based aggregation topic model for understanding user's preference and intention in social network, *Neurocomputing*, **413** (2020), 1–13. https://doi.org/10.1016/j.neucom.2020.06.099
7. M. Xie, H. Yin, H. Wang, F. Xu, W. Chen, S. Wang, Learning graph-based poi embedding for location-based recommendation, in *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, ACM, (2016), 15–24. https://doi.org/10.1145/2983323.2983711
8. X. Hu, J. Xu, W. Wang, Z. Li, A. Liu, A graph embedding based model for fine-grained POI recommendation, *Neurocomputing*, **428** (2021), 376–384. https://doi.org/10.1016/j.neucom.2020.01.118
9. T. Yang, H. Gao, C. Yang, C. Shi, Q. Xie, X. Wang, et al., Memory-enhanced period-aware graph neural network for general POI recommendation, in *Proceedings of the International Conference on Database Systems for Advanced Applications*, Springer, (2023), 462–472. https://doi.org/10.1007/978-3-031-30672-3_31
10. B. Hu, Y. Ye, Y. Zhong, J. Pan, M. Hu, Transmkr: Translation-based knowledge graph enhanced multi-task point-of-interest recommendation, *Neurocomputing*, **474** (2022), 107–114. https://doi.org/10.1016/j.neucom.2021.11.049

11. S. Xing, F. Liu, Q. Wang, X. Zhao, T. Li, Content-aware point-of-interest recommendation based on convolutional neural network, *Appl. Intell.*, **49** (2019), 858–871. https://doi.org/10.1007/s10489-018-1276-1
12. G. Li, Q. Chen, B. Zheng, H. Yin, Q. V. H. Nguyen, X. Zhou, Group-based recurrent neural networks for POI recommendation, *ACM/IMS Trans. Data Sci.*, **1** (2020), 1–18. https://doi.org/10.1145/3343037
13. C. Liu, J. Liu, J. Wang, S. Xu, H. Han, Y. Chen, An attention-based spatiotemporal gated recurrent unit network for point-of-interest recommendation, *ISPRS Int. J. Geo-Inf.*, **8** (2019), 355. https://doi.org/10.3390/ijgi8080355
14. D. Li, Z. Gong, D. Zhang, A common topic transfer learning model for crossing city POI recommendations, *IEEE Trans. Cybern.*, **49** (2018), 4282–4295. https://doi.org/10.1109/TCYB.2018.2861897
15. Z. Guo, K. Yu, N. Kumar, W. Wei, S. Mumtaz, M. Guizani, Deep-Distributed-Learning-Based POI recommendation under mobile-edge networks, *IEEE Internet Things J.*, **10** (2022), 303–317. https://doi.org/10.1109/JIOT.2022.3202628
16. M. Gan, Y. Ma, Mapping user interest into hyper-spherical space: A novel POI recommendation method, *Inf. Process. Manage.*, **60** (2023), 103169. https://doi.org/10.1016/j.ipm.2022.103169
17. M. B. Hossain, M. S. Arefin, I. H. Sarker, M. Kowsher, P. K. Dhar, T. Koshiba, CARAN: A Context-Aware Recency based Attention Network for point-of-interest recommendation, *IEEE Access*, **10** (2022), 36299–36310. https://doi.org/10.1109/ACCESS.2022.3161941
18. Y. C. Chen, T. Thaipisutikul, T. K. Shih, A learning-based POI recommendation with spatiotemporal context awareness, *IEEE Trans. Cybern.*, **5** (2020), 2453–2466. https://doi.org/10.1109/TCYB.2020.3000733
19. P. Han, S. Shang, A. Sun, P. Zhao, K. Zheng, X. Zhang, Point-of-interest recommendation with global and local context, *IEEE Trans. Knowl. Data Eng.*, **34** (2021), 5484–5495. https://doi.org/10.1109/TKDE.2021.3059744
20. X. Li, M. Jiang, H. Hong, L. Liao, A time-aware personalized point-of-interest recommendation via high-order tensor factorization, *ACM Trans. Inf. Syst.*, **35** (2017), 1–23. https://doi.org/10.1145/3057283
21. J. He, X. Li, L. Liao, Category-aware next point-of-interest recommendation via listwise bayesian personalized ranking, in *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, AAAI Press, (2017), 1837–1843.
22. J. He, X. Li, L. Liao, Next point-of-interest recommendation via a category-aware Listwise Bayesian Personalized Ranking, *J. Comput. Sci.*, **28** (2018), 206–216. https://doi.org/10.1016/j.jocs.2017.09.014
23. X. Li, D. Han, J. He, L. Liao, M. Wang, Next and next new POI recommendation via latent behavior pattern inference, *ACM Trans. Inf. Syst.*, **37** (2019), 1–28. https://doi.org/10.1145/3354187
24. P. Zhao, A. Luo, Y. Liu, J. Xu, Z. Li, F. Zhuang, et al., Where to go next: A spatio-temporal gated network for next poi recommendation, *IEEE Trans. Knowl. Data Eng.*, **34** (2020), 2512–2524. https://doi.org/10.1109/TKDE.2020.3007194
25. Y. Wu, K. Li, G. Zhao, X. Qian, Long-and short-term preference learning for next POI recommendation, in *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, ACM, (2019), 2301–2304. https://doi.org/10.1145/3357384.3358171

26. J. Ni, P. Zhao, J. Xu, J. Fang, Z. Li, X. Xian, et al., Spatio-temporal self-attention network for next POI recommendation, in *Web and Big Data: 4th International Joint Conference*, Springer, (2020), 409–423. https://doi.org/10.1007/978-3-030-60259-8_30
27. K. Zhao, Y. Zhang, H. Yin, J. Wang, K. Zheng, X. Zhou, et al., Discovering subsequence patterns for next POI recommendation, in *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*, (2020), 3216–3222. https://doi.org/10.24963/ijcai.2020/445
28. Z. Wang, Y. Zhu, Q. Zhang, H. Liu, C. Wang, T. Liu, Graph-enhanced spatial-temporal network for next POI recommendation, *ACM Trans. Knowl. Discovery Data*, **16** (2022), 1–21. https://doi.org/10.1145/3513092
29. J. Fang, X. Meng, URPI-GRU: An approach of next POI recommendation based on user relationship and preference information, *Knowledge-Based Syst.*, **256** (2022), 109848. https://doi.org/10.1016/j.knosys.2022.109848
30. C. Zheng, D. Tao, J. Wang, L. Cui, W. Ruan, S. Yu, Memory augmented hierarchical attention network for next point-of-interest recommendation, *IEEE Trans. Comput. Social Syst.*, **8** (2020), 489–499. https://doi.org/10.1109/TCSS.2020.3036661
31. L. Huang, Y. Ma, Y. Liu, K. He, DAN-SNR: A deep attentive network for social-aware next point-of-interest recommendation, *ACM Trans. Internet Technol.*, **21** (2020), 1–27. https://doi.org/10.1145/3430504
32. Y. Lai, Y. Su, L. Wei, G. Chen, T. Wang, D. Zha, Multi-view Spatial-Temporal Enhanced Hypergraph Network for next POI recommendation, in *Database Systems for Advanced Applications*, Springer, (2023), 237–252. https://doi.org/10.1007/978-3-031-30672-3_16
33. P. Lan, Y. Zhang, H. Xiang, Y. Wang, W. Zhou, Spatio-temporal position-extended and gated-deep network for next POI recommendation, in *Database Systems for Advanced Applications*, Springer, (2023), 505–520. https://doi.org/10.1007/978-3-031-30672-3_34
34. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, et al., Attention is all you need, in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Curran Associates Inc., (2017), 6000–6010.
35. T. Wu, R. Zhu, S. Wan, Semantic map guided identity transfer GAN for person re-identification, *ACM Trans. Multimedia Comput. Commun. Appl.*, **2023** (2023). https://doi.org/10.1145/3631355
36. Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, et al., Swin Transformer: Hierarchical vision transformer using shifted windows, in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, IEEE, (2021), 9992–10002. https://doi.org/10.1109/ICCV48922.2021.00986
37. Z. Liu, J. Ning, Y. Cao, Y. Wei, Z. Zhang, S. Lin, et al., Video Swin Transformer, in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (2022), 3192–3201. https://doi.org/10.1109/CVPR52688.2022.00320
38. J. Devlin, M. W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Association for Computational Linguistics, (2019), 4171–4186. https://doi.org/10.18653/v1/N19-1423
39. H. Sun, S. Geng, J. Zhong, H. Hu, K. He, Graph Hawkes Transformer for extrapolated reasoning on temporal knowledge graphs, in *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, (2022), 7481–7493. https://doi.org/10.18653/v1/2022.emnlp-main.507

40. C. Ying, T. Cai, S. Luo, S. Zheng, G. Ke, D. He, et al., Do transformers really perform badly for graph representation, in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., **34** (2021), 28877–28888.
41. J. Chen, K. Gao, G. Li, K. He, NAGphormer: A tokenized graph transformer for node classification in large graphs, in *Proceedings of the Eleventh International Conference on Learning Representations*, 2022.
42. Q. Li, X. Xu, X. Liu, Q. Chen, An attention-based spatiotemporal GGNN for next POI recommendation, *IEEE Access*, **10** (2022), 26471–26480. https://doi.org/10.1109/ACCESS.2022.3156618
43. L. Zhang, Z. Sun, Z. Wu, J. Zhang, Y. S. Ong, X. Qu, Next point-of-interest recommendation with inferring multi-step future preferences, in *Proceedings of the 31st International Joint Conference on Artificial Intelligence*, International Joint Conferences on Artificial Intelligence Organization, (2022), 3751–3757. https://doi.org/10.24963/ijcai.2022/521