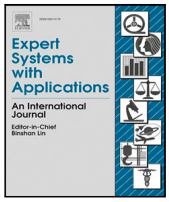
Expert Systems With Applications 256 (2024) 124884

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa



Collaborative trajectory representation for enhanced next POI recommendation

Jiankai Zuo, Yaying Zhang *

Key Laboratory of Embedded System and Service Computing of Ministry of Education, Tongji University, Shanghai, China

ARTICLE INFO

Dataset link: https://github.com/JKZuo/CTRNext

Keywords:
Next POI recommendation
Trajectory similarity
Attention mechanism
Representation learning

ABSTRACT

Point-of-Interest (POI) recommendation stands as the cornerstone within a variety of location-based applications and services that intend to anticipate upcoming movements that users may be interested in. The current state-of-the-art methods have effectively explored spatio-temporal contextual features and users' long-term and short-term preference patterns. Nevertheless, most existing work lacks the ability to effectively capture group movement patterns from trajectory collaboration. Additionally, they pay close attention to the accuracy of personalized recommendations, neglecting recommendation diversity, which refers to offering broader location options that extend beyond a user's typical preferences, thereby avoiding excessive homogeneity or repetition. To address these gaps, this study proposes a Collaborative Trajectory Representation model (CTRNext), which enhances the diversity of recommendations while maintaining the precision of personalized preferences. To be specific, we first design two trajectory embedding layers to extract joint semantic interactions and the explicit spatiotemporal context-aware representation. Then, a trajectory semantic similarity calculation module that captures collaborative signals from potentially similar-minded users and eliminates barriers caused by trajectory length is proposed. Next, the implicit correlation and further updated representation between different check-in records are achieved through a multi-head self-attention aggregation module. Finally, we put forward a dual-driven user preference matching module to generate the preference-based next POI recommendation while enhancing diversity. Our approach demonstrates its remarkable recommendation accuracy through extensive experimentation on four real-world datasets, surpassing the performance of state-of-the-art methodologies.

1. Introduction

The point-of-interest (POI) recommendation (Islam et al., 2022; Sánchez & Bellogín, 2022), commonly known as next location prediction or mobility prediction, holds substantial significance in the realm of location-based social networks (LBSNs). This predictive modeling, aimed at foreseeing destinations that might be of interest to users, has garnered noteworthy interest due to its potential relevance in catering to user preferences within LBSNs. Traditional POI recommendation models, based primarily on non-deep-learning methods such as Markov chains (Rendle et al., 2010) and matrix factorization (Cheng et al., 2013; Feng et al., 2015), excel in learning users' short-term preferences but are constrained by data sparsity. With the advancement of deep learning, RNN-based models employing gating mechanisms (Liu et al., 2016; Zhao et al., 2022, 2019) or leveraging attention mechanisms (Fang & Meng, 2022; Liu et al., 2016; Long et al., 2022; Luo et al., 2021; Wu et al., 2022; Yang et al., 2022) have been adept at capturing users' long-term preferences. Additionally, GNN-based models (Fu et al., 2024; Rao et al., 2022; Wang et al., 2023, 2022; Xia et al., 2023; Yan et al., 2023; Yin et al., 2023) alleviate data sparsity by constructing various graph structures (user social graphs, POI transition graphs, user-POI hypergraphs, etc.). However, there are still two challenges: (1) lack of ability to effectively capture collaborative information from the trajectories of similar-minded users and (2) difficulty to strike a balance between personalized and diverse recommendations.

Although existing studies have taken factors such as spatial, social relations, and temporal elements influencing the effectiveness of POI recommendations into account, their focus tends to lean toward fitting individual users' historical check-in behaviors. These approaches often fall short in learning the movement patterns of groups, particularly the collaborative information among similar-minded users, thereby neglecting valuable insights into their collective mobility behaviors. In fact, predicting the next visit of a specific user not only relies on that user's individual historical check-in POI (i.e., POI-POI Interaction) but

* Corresponding author.
E-mail addresses: tj_zjk@tongji.edu.cn (J. Zuo), yayingzhang@tongji.edu.cn (Y. Zhang).

https://doi.org/10.1016/j.eswa.2024.124884
Received 21 March 2024; Received in revised form 30 June 2024; Accepted 23 July 2024
Available online 27 July 2024

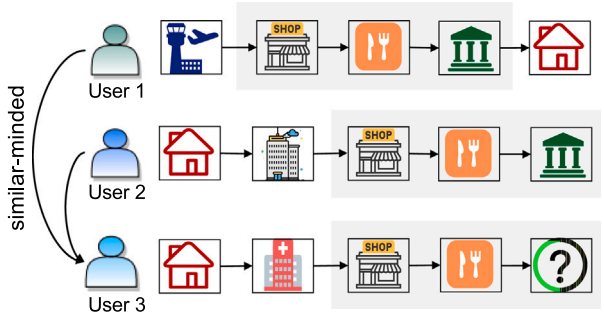


Fig. 1. An illustration of the similar-minded users. Three users share similar habitual preferences, from the shop to the restaurant.

also correlates with the trajectories of other users who share similar habitual preferences (i.e., User–User Interaction), as shown in Fig. 1. Users 1 and 2 checked in to a museum after visiting the shop and restaurant. If it is determined that users 1, 2, and 3 are similar-minded, and after user 3 has visited shops and restaurants, a museum may be recommended. This collaborative information from similar-minded users' trajectories contributes significantly to forecasting the next movement, especially for inactive or cold-start users. Currently, there are primarily two categories of approaches to identify potentially similar-minded users: the explicit social relationship (Fang & Meng, 2022; Perifanis et al., 2023; Seyedhoseinzadeh et al., 2022) and implicit trajectory similarity (Chen et al., 2020; Furtado et al., 2016; Li et al., 2021; Liu et al., 2020; Xiao et al., 2020). The former methods heavily rely on prior knowledge of user social connections (e.g., real-life friendships between users). They are highly dependent on the availability and quality of auxiliary data. The latter approaches face challenges related to varying lengths of user trajectories (such as trajectory matching, alignment, completion, and padding issues), which are mainly caused by differences in individual user activities. Therefore, a key challenge lies in ensuring that the recommendation model can capture collaborative information without being influenced by differences in trajectory length and data sparsity.

In addition, balancing personalization (Feng et al., 2015; Rendle et al., 2010; Wang et al., 2019; Wu et al., 2022; Zhao et al., 2020) and diversity (Chen et al., 2021, 2023; Meng & Fang, 2020; Werneck et al., 2021; Zang et al., 2021) in POI recommendations is another major challenge. Specifically, diversity means providing users with a wider range of location options beyond their habitual preferences. For instance, if a user has mainly visited coffee shops in the past, recommendation diversity enables that he would also receive recommendations for restaurants, bookstores, and other coffee-themed places, based on the preferences of other users who share similar trajectory patterns. On the other hand, personalization entails tailoring the recommended results to individual users based on their preferences and past behavior, aiming to meet their unique interests. Emphasizing personalization excessively would constrain recommendations, potentially limiting the diversity of suggested POIs. Only treating user behaviors as fixed and unchanging static preferences will ignore the periodic changes in points-of-interest and the dynamic dependence of preferences between different stages. This could result in a narrowed scope of recommendations that align closely with users' known specific preferences but fail to introduce or explore novel and diverse POIs, impeding the model's ability to cater to evolving user interests.

To address the aforementioned challenges, we propose a Collaborative Trajectory Representation model, named CTRNext, to enhance the next POI recommendation. Specifically, we propose a trajectory semantic similarity calculation module to capture collaborative signals among potentially similar-minded users and eliminate barriers posed by trajectory lengths. This module represents different trajectories as length-agnostic representation matrices, which incorporate semantic relevance and spatiotemporal context features to better describe user behavioral patterns and preferences. To strike a balance between recommendation diversity and personalization, we propose a dual-driven user preference matching module which consists of a cosine-based preference similarity matching layer and an attention-based candidate location matching layer. The former aims to enhance the diversity of recommendations by integrating collaborative features from potentially similar-minded users. The latter aggregates explicit spatiotemporal context-aware interactions to generate personalized recommendations.

The main contributions are summarized as follows.

- We capture collaborative signals from similar-minded users' trajectories and overcome length-related issues by converting trajectories into length-agnostic matrices. These representation matrices encapsulate semantic relevance and spatiotemporal context, effectively characterizing user regular behaviors and specific preferences.
- We develop a dual-driven user preference matching module to harmonize personalized recommendations with diversity by learning collaborative features based on potentially similar-minded users and introducing an attention-based candidate location matching layer.
- Extensive experiments are conducted on four real-world LBSN datasets. The test results demonstrate that the CTRNext model prominently outperforms state-of-the-art baseline methods in both complete and incomplete (i.e., missing or sparse) check-in trajectory data. Moreover, case studies highlight the model's capacity for the diversity of recommendations.

The remainder of this paper is organized as follows: Section 2 investigates existing research on the next POI recommendations. Section 3 presents preliminary descriptions and the problem statement. Section 4 elaborates on the CTRNext model proposed in this paper. Section 5 discusses the experiments and result analysis. Finally, Section 6 makes the conclusion and gives future directions.

2. Related work

We mainly brief the related work from two aspects: POI recommendations and trajectory similarity calculation.

2.1. Next POI recommendation

The point-of-interest (POI) recommendation primarily relies on the historical interaction sequences between users and points of interest, aiming to explore users' subsequent decision-making processes. For sequential dependency modeling, researchers have introduced different recommendation methodologies to unearth the evolving preferences within user mobility patterns. For example, Cheng et al. (2013) extended the factorizing personalized Markov chains (FPMC) (Rendle et al., 2010) to the task of the next point-of-interest recommendation by incorporating personalized embeddings and regional information. Feng et al. (2015) devised a personalized ranking metric embedding approach (PRME), which integrates sequential information with geographical influences. Liu et al. (2016) introduced a spatiotemporal recurrent neural network (ST-RNN), employing specific transition matrices for varying time intervals and geographical distances.

Attention-based models also have been proposed to capture more comprehensive dependencies. DeepRP model (Long et al., 2022)1u adopted an RNN with backtracking attention to discern users' regular preferences at different temporal granularities by retrieving the context of a trajectory. BayMAN (Xia et al., 2023) studied robust POI recommendations by using the Bayesian-enhanced graph and proposed multi-perspective attention from time-, distance-, and semantic-aware. STAN (Luo et al., 2021) considered both non-adjacent and

non-successive visits by combining self-attention with a spatiotemporal correlation matrix. GETNext (Yang et al., 2022) introduced a transition attention component and Transformer layers to improve recommendation accuracy for inactive users with short trajectories.

In recent years, graph neural networks (GNNs) have been employed to address the cold-start problem inherent in POI recommendation tasks by leveraging their feature aggregation ability among neighboring nodes. For example, Wang et al. (2023) came up with an enhanced encoder–decoder network (EEDN) by incorporating hyper-graph convolutions and matrix factorization. Rao et al. (2022) defined a similarity function and a user-POI knowledge graph to reflect the transition patterns among POIs. To further capture the latent relations between POIs, Yin et al. (2023) proposed a sequence-based and dynamic neighbor graph to search for similar neighborhood POIs. However, they all imply low-order POI-POI relationships and fail to obtain higher-order features from trajectory collaboration to gain insight into global group movement patterns.

2.2. Trajectory similarity calculation

Researchers have improved the performance of POI recommendations by exploring trajectory and user similarity. For instance, Liu et al. (2020) proposed a multi-level attention model (At2vec), which addresses the problem of uneven sampling rate and individual activity differences in trajectory similarity calculation. Li et al. (2021) captured the context-aware similarity between users through the proposed trajectory grouping strategy. Xiao et al. (2020) constructed a trajectory similarity matrix to measure the regular travel behavior of different individuals and introduced transfer learning to alleviate data sparsity. Wang et al. (2019) explored trust-enhanced user similarity by considering the geographical and temporal influences of trajectories and user preferences. Chen et al. (2020) proposed a trajectory pair filtering technique to obtain the parallel semantic trajectory join features. Bok et al. (2019) employed a pattern-matching method to determine the POI trajectory similarity between different users. Yu et al. (2022) used three features (similarity, popularity, and location) of POIs to output the initial interest points and then integrated users with similar trajectories to enhance recommendation. However, they are unable to encapsulate richer semantic representations for downstream tasks of POI recommendations. In this work, we propose a novel trajectory semantic similarity calculation module to capture the latent relation among POIs as well as the interaction between users and POIs.

3. Preliminaries

In the next POI recommendation system, we first define three basic sets. Let $\mathbb{U} = \{u_i\}_{i=1}^{U} = \{u_1, u_2, \ldots, u_U\}$ indicate a set of users. $\mathbb{L} = \{l_i\}_{i=1}^{L} = \{l_1, l_2, \ldots, l_L\}$ be a set of locations (POIs). $\mathbb{T} = \{t_i\}_{i=1}^{T} = \{t_1, t_2, \ldots, t_T\}$ means a set of timestamps. For convenience, we have summarized the notations of our paper in Table 1.

Definition 1 (*Check-in Record*)**.** A check-in event is denoted by a tuple $r = \langle u, l, t \rangle$, signifying that a user u visited a location l at the timestamp t. Among them, each location $l \in \mathbb{L}$ is composed of a quadruplet $l = \langle id, lon, lat, cat \rangle$, incorporating identification, longitude, latitude, and location category (shopping malls, hospitals, and stations, etc.).

Definition 2 (*Check-in Trajectory Sequence*)**.** For any given user $u \in \mathbb{U}$, we define his movement trajectory as a chronological sequence $\mathcal{T}_u = \{r_1, r_2, \ldots, r_n\}$ containing n check-in records. It is noted that the length n of such trajectories is user-specific.

Definition 3 (*Next POI Recommendation*)**.** Formally, given the historical trajectory dataset $\{\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_U\}$ of all users, alongside a current trajectory $\mathcal{T}_u^c = \{r_1, r_2, \ldots, r_c\}$ for a target user $u \in \mathbb{U}$, and an impending check-in time t_{c+1}, the problem of the next POI location recommendation is to predict probability scores $list(u) = \{\hat{y}_{l_i}^u | i = 1, 2, \ldots, L\}$ for all POI candidate locations. Then the Top-k POIs based on the probability ranking are recommended to user u who might visit at the next movement l_{c+1}.

Table 1
Main notations used in our paper.

Notations	Descriptions
$\mathbb{U}, \mathbb{L}, \mathbb{T}$	User set, location set, timestamp set
u, l, t	User $u \in \mathbb{U}$, POI $l \in \mathbb{L}$, timestamp $t \in \mathbb{T}$
r	Check-in event
$\mathcal{T}_u$	Check-in trajectory sequence of user u
e_u, e_l, e_t	User embedding, location embedding, time embedding
e_s, e_c	Semantic embedding, time-conscious category embedding
d_u, d_l, d_t	Embedding sizes of user, location, time
d	Uniform embedding dimension
$\Delta t, \Delta d$	Time interval, distance interval
$e_{\Delta t}, e_{\Delta d}$	Time and distance interval embeddings
$\Delta T, \Delta D$	Temporal and spatial interval matrices
$M_{\Delta t}, M_{\Delta d}$	Explicit time and distance interval matrices
$\mathbf{M_J}, \mathbf{M_E}$	Joint feature matrix, explicit feature matrix
$\mathbf{M_H}, \mathbf{M_L}$	Hidden feature matrix, latent feature matrix
$\mathbf{M_F}$	User-location check-in frequency matrix
$\mathbf{M_R}$	Length-agnostic representation matrix
$\mathbf{M_A}$	Self-attention aggregation matrix
λ, W_*	Time-decay rate, trainable weight
n_h, q	Number of heads, number of attention channels

4. Methodology

The overall framework of our proposed CTRNext model is illustrated in Fig. 2, comprising primarily of two embedding layers that learn the dense representations from user historical check-in trajectories, and three learning modules that refine and update these representations to generate recommendations tailored to user preferences. These three learning modules are summarized as follows:

(i) Trajectory Semantic Similarity Calculation Module (TSSC): This pivotal module capitalizes on a joint feature matrix $\mathbf{M_J}$, derived from the Trajectory Joint Feature Embedding Layer, in concert with a user-location check-in frequency matrix $\mathbf{M_F}$. Its highlight lies in computing a length-agnostic representation matrix $\mathbf{M_R}$ that encapsulates composite semantic information of trajectories, overcoming the constraints imposed by the variable trajectory length of different users.

(ii) Multihead Self-Attention Aggregation Module (MSAA): Building on the rich explicit feature matrix $\mathbf{M_E}$ extracted from the Explicit SpatioTemporal Interval Embedding Layer, this module synergizes with a latent feature matrix $\mathbf{M_L}$. Through a series of intricate self-attention calculations, it dynamically updates representations, culminating in a self-attention aggregation matrix $\mathbf{M_A}$ that embodies the implicit interaction of spatiotemporal factors.

(iii) Dual-driven User Preference Matching Module (DUPM): As the decoder of our framework, DUPM seamlessly amalgamates the representation matrix $\mathbf{M_R}$ with the aggregation matrix $\mathbf{M_A}$. The result is a dual-informed engine that generates a diverse and personalized list of POI recommendations, echoing the unique preferences of each user.

Hereafter, every component in the proposed CTRNext model will be described in detail.

4.1. Trajectory joint feature embedding layer

To transform raw check-in trajectories into rich learnable vector representations, we propose a trajectory joint feature embedding layer. This layer projects a raw trajectory $\mathcal{T} = \{r_1, r_2, \ldots, r_n\}$ into a joint feature matrix $\mathbf{M_J} = (v_1, v_2, \ldots, v_n) \in \mathbb{R}^{n \times d}$. Specifically, for the user $\mathbb{U}$ and location $\mathbb{L}$ sets, we trained two embedding functions to distinguish the general behavior for each specific user $u \in \mathbb{U}$, as well as discerning

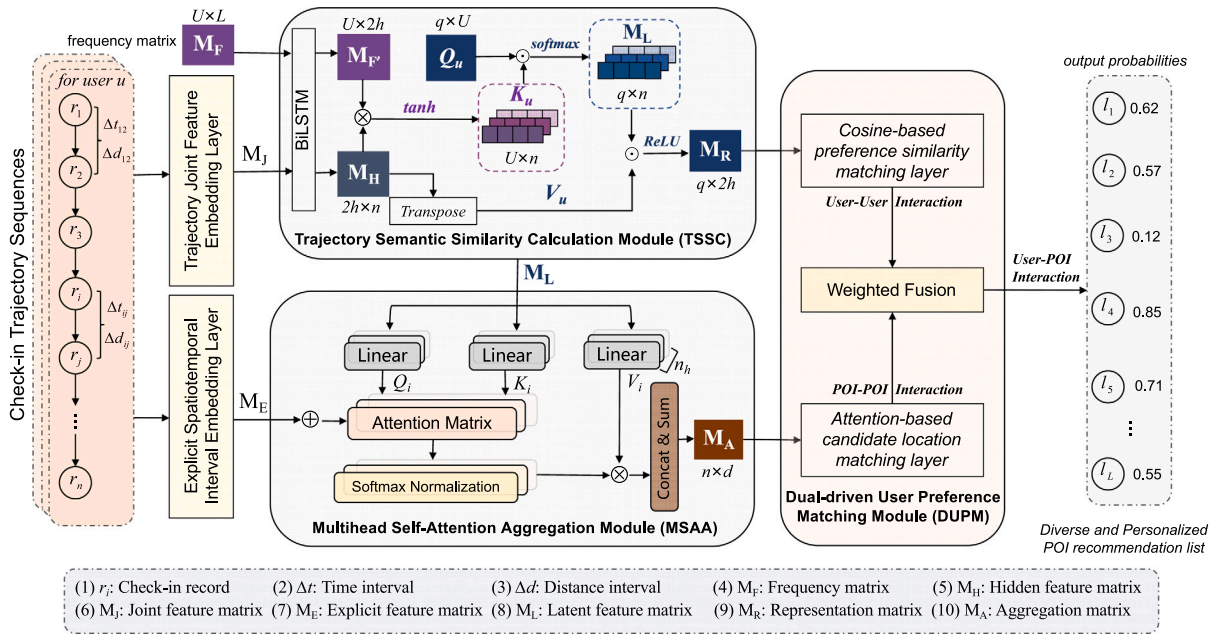


Fig. 2. The framework of proposed Collaborative Trajectory Representation (CTRNext) model.

the spatial location distribution for any given POI $l \in \mathbb{L}$, formalized in Eqs. (1) and (2), respectively.

$$e_u = \mathbf{F_{embed}}(u) \in \mathbb{R}^{d_u}, \tag{1}$$

$$e_l = \mathbf{F_{embed}}(l) \in \mathbb{R}^{d_l}, \tag{2}$$

Considering the temporal dynamics associated with POI categories – for instance, transportation stations experience a surge in passenger flow during rush hours – a uniform embedding approach is insufficient to represent the time-varying POI categories. To this end, we devise a time-conscious category embedding strategy. Concretely, based on the activities of individual users in different categories of POIs in actual scenarios, we first summarize the d_s attributes of POIs as (*public, consumption, entertainment, learning, sightseeing, sports, health, transportation, office, etc*). A point-wise feed-forward neural network (FNN) is then employed to semantically score these attributes, thus yielding a vector representation $e_s = \mathbf{FNN}(\text{cat}) \in \mathbb{R}^{d_s}$ that resonates with the POI categories' semantic features. For instance, the learned semantic representation of the station category is $e_s(station) = (0.8, 0.3, 0.1, 0.2, 0.6, 0.5, 0.1, 0.9, 0.2, \dots)$.

Additionally, leveraging the Time2Vec (Kazemi et al., 2019) encoding scheme to encapsulate the visiting periodicity across categories, we integrate $e_t = \sin(\omega t + \varphi) \in \mathbb{R}^{d_t}$ into our model. The resultant time-conscious category embedding e_c is fused as Eq. (3) to capture the temporal pattern of user activities.

$$e_c = \mathbf{Fusion}\left(e_s, e_t\right) \in \mathbb{R}^{d_c}, \tag{3}$$

where $\mathbf{Fusion}(\cdot)$ is a $ReLU$-activated feed-forward neural network devoid of bias terms.

To preserve the generality of representation without forfeiting specificity, we standardize the embedding dimensions d_u, d_l, and d_c to a uniform dimension d. The culminating representation of each check-in event r_i is a weighted aggregation of the three embeddings $v_i = \mathbf{WeightAgg}(e_u, e_l, e_c) \in \mathbb{R}^d$, thereby reflecting a composite semantic interaction inherent in the check-in trajectories.

4.2. Explicit SpatioTemporal interval embedding layer

When analyzing user behavior patterns from the perspective of temporal dimension, only considering the check-in timestamp is not comprehensive enough. For example, for two users with similar interests but divergent meal timings, the description of the daily routines of the two users is shown in Fig. 3(a). Only timestamp embedding could skew the model with time biases in the learning process, misrepresenting the similarity in their behavioral patterns. This discrepancy can be mitigated by introducing time intervals that capture the approximate temporal proximity of activities. In addition, the spatial interval of other similar-minded users has analogous situations.

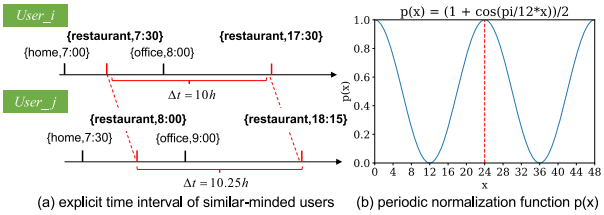


Fig. 3. Explicit time interval and the periodic function.

To further mine the spatiotemporal context inherent in trajectory data, we construct matrices $\Delta T \in \mathbb{R}^{n\times n}$ and $\Delta D \in \mathbb{R}^{n\times n}$ to encode temporal and spatial intervals, respectively. The matrix entries $\Delta t_{i,j} = |t_j - t_i|$ and $\Delta d_{i,j} = |d_j - d_i|$ signify the explicit temporal and spatial intervals between records r_i and r_j. Similar to the trajectory joint feature embedding layer, we also need to obtain dense vector representations of these two explicit interval matrices.

Acknowledging the cyclic nature of human activities, we observe a 24-h periodicity with a high intraday correlation that attenuates over longer intervals. Daytime and nocturnal activities exhibit stark dissimilarities, suggesting a transformation threshold at the 12-h mark. To encapsulate this phenomenon, we devise a periodic normalization function $p(x)$ grounded in trigonometric principles, as Eq. (4) and more intuitively see Fig. 3(b).

$$p(\Delta t) = \frac{1 + \cos\left(\frac{\pi}{12}(\Delta t)\right)}{2}, \tag{4}$$

In light of the observed correlation in check-in records decreasing with the rise of time intervals, we multiply a time exponential decay rate to represent this temporal evolution. Finally, a trainable time-aware parameter matrix W^T is applied to obtain the embedding vectors

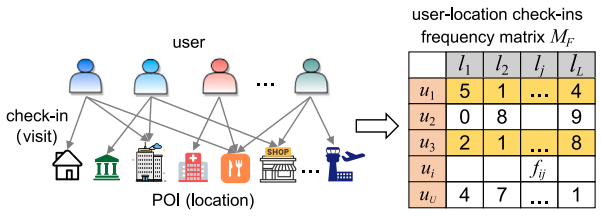


Fig. 4. User-location check-in frequency matrix.

$e_{\Delta t}$ for each Δt, as Eq. (5).

$$e_{\Delta t} = W^T(p(\Delta t) \times e^{-\lambda \Delta t}) \in \mathbb{R}^{d_\delta}, \tag{5}$$

where λ denotes a time decay rate, reflecting the weight of the impact on time interval.

For explicit distance interval, a trainable distance-aware parameter matrix W^D is employed to gain the embedding vectors $e_{\Delta d}$ for each Δd, as Eq. (6).

$$e_{\Delta d} = W^D(\Delta d) \in \mathbb{R}^{d_\delta}, \tag{6}$$

Therefore, dense vector representations of two explicit interval matrices can be expressed as $M_{\Delta t} \in \mathbb{R}^{n \times n \times d_\delta}$ and $M_{\Delta d} \in \mathbb{R}^{n \times n \times d_\delta}$. We apply a weighted aggregation on the last dimension of two representation matrices to obtain an explicit feature matrix $\mathbf{M_E} = \mathbf{WeightAgg}(M_{\Delta t}, M_{\Delta d}) \in \mathbb{R}^{n \times n}$ that reflects the joint spatiotemporal context-aware interactions of check-in trajectories.

4.3. Trajectory semantic similarity calculation module

The intricacies of human mobility patterns are often encoded in the spatiotemporal trajectories generated by user check-ins. Traditional models tend to overlook the nuanced semantics of these trajectories, leading to a deficient representation that is unable to effectively capture the complex interplay of location, time, and user behavior. Recognizing this gap, we introduce the Trajectory Semantic Similarity Calculation Module (TSSC). It is mainly designed based on bidirectional LSTM and attention mechanism. The objective is to obtain a low-dimensional representation matrix $\mathbf{M_R}$ that is agnostic to the sequence length n. Given a trajectory $\mathcal{T}$ comprising n check-in records, we can express it as $\mathbf{M_J} = (v_1, v_2, \ldots, v_n) \in \mathbb{R}^{n \times d}$. Each v_i is a d-dimensional vector encapsulating semantic features, derived from the joint feature embedding layer. We process $\mathbf{M_J^\top} \in \mathbb{R}^{d \times n}$ through a BiLSTM to learn the dependencies between adjacent check-in points within a trajectory. The resulting forward and backward hidden state vectors from the BiLSTM, $\overrightarrow{H}_i \in \mathbb{R}^h$ and $\overleftarrow{H}_i \in \mathbb{R}^h$ respectively, where h is the dimension size, provide a comprehensive representation of the trajectory by capturing both past and future context, as Eqs. (7) and (8).

$$\overrightarrow{H}_i = \mathbf{ForwardLSTM}\left(\overrightarrow{H_{i-1}}, v_i\right), \tag{7}$$

$$\overleftarrow{H}_i = \mathbf{BackwardLSTM}\left(\overleftarrow{H_{i-1}}, v_i\right), \tag{8}$$

These hidden state matrices $\overrightarrow{H}$ and $\overleftarrow{H}$ are all in $\mathbb{R}^{h \times n}$, which are combined to form a hidden feature matrix $\mathbf{M_H} = [\overrightarrow{H} \parallel \overleftarrow{H}] \in \mathbb{R}^{2h \times n}$. This matrix reflects the temporal correlations of each check-in point within a trajectory.

For POI recommendations across U users and L locations (POIs), we could construct a check-in frequency matrix $\mathbf{M_F} \in \mathbb{R}^{U \times L}$, where each element f_{ij} signifies the frequency of visits by user i to location j, illustrated as Fig. 4. However, this generated user-POI interaction matrix is typically sparse because the number of POIs visited by a certain user is limited. Since the sparsity of user-POI interactions has been a persistent problem, it is necessary to compress $\mathbf{M_F}$ while retaining the sequential dependency patterns among different locations. By leveraging a BiLSTM to $\mathbf{M_F}$, we obtain an enriched frequency matrix $\mathbf{M'_F} \in \mathbb{R}^{U \times 2h}$.

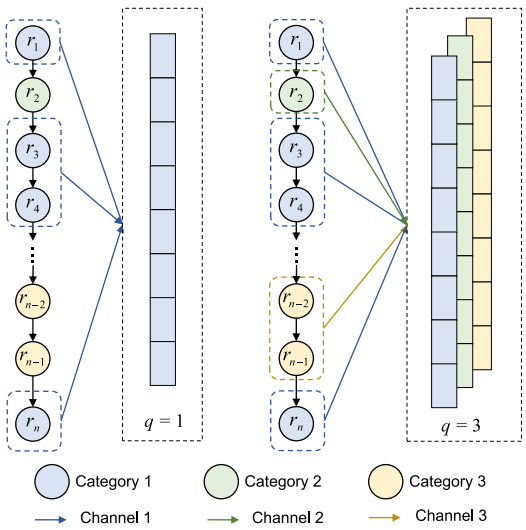


Fig. 5. An illustration of the attention channel.

Utilizing both the enriched frequency matrix $\mathbf{M'_F}$ and hidden feature matrix $\mathbf{M_H}$, and by introducing a self-attention mechanism, we can encode any trajectory with variable length n into a length-agnostic representation matrix $\mathbf{M_R}$. This is achieved by learning a latent feature matrix $\mathbf{M_L} \in \mathbb{R}^{q \times n}$ that captures visiting frequencies and temporal correlations through q attention channels, as Eq. (9). Specifically, $\mathbf{M_H}$ is a triple composite semantic matrix with temporal correlation that fuses user identity e_u, POI location e_l, and time-conscious category embedding information e_c. We combine $\mathbf{M_H}$ with the enriched frequency matrix $\mathbf{M'_F}$ to obtain $\mathbf{K_u} = tanh(\mathbf{M'_F} \times \mathbf{M_H}) \in \mathbb{R}^{U \times n}$ that considers the personalized preferences of each individual user.

$$\mathbf{M_L} = softmax(\mathbf{Q_u} \odot tanh(\mathbf{M'_F} \times \mathbf{M_H})), \tag{9}$$

where $\mathbf{Q_u}$ is a trainable parameter that can be interpreted as the Query matrix of $\mathbf{M_H}$. The symbol $\odot$ signifies the dot-product attention and $softmax(\cdot)$ ensures that the sum of all attention weights between $\mathbf{Q_u}$ and $\mathbf{K_u}$ is equal to one.

In an effort to capture broader semantic contexts, we employ a multi-channel attention mechanism, allowing us to aggregate multiple contexts and focus on the most relevant movement patterns and activity behaviors. Here, for $\mathbf{Q_u} \in \mathbb{R}^{q \times U}$ matrix, q represents the number of attention channels. We take the visible explicit POI category information as an example, described in Fig. 5. We found that when $q = 1$, the obtained $\mathbf{M_L} \in \mathbb{R}^{1 \times n}$ can only reflect local semantics in a trajectory, such as successive check-in records with the same categories or adjacent visited points in spatial. Hence, the single channel fails to capture context-aware correlations between non-successive check-in records with distinct categories. However, $\mathbf{M_L} \in \mathbb{R}^{q \times n}$ can capture the latent semantic similarity between trajectory segments from q different scales.

Finally, a $ReLU$-transformed dot-product attention between $\mathbf{M_L} \in \mathbb{R}^{q \times n}$ and the transposed $\mathbf{M_H^\top} \in \mathbb{R}^{n \times 2h}$ is implemented, as Eq. (10). We can obtain a length-agnostic representation matrix $\mathbf{M_R} \in \mathbb{R}^{q \times 2h}$ that encapsulates rich semantic information and depicts latent relations between POIs within a complete check-in trajectory.

$$\mathbf{M_R} = ReLU(\mathbf{M_L} \odot \mathbf{M_H^\top}). \tag{10}$$

It is worth noting that the size of the trajectory representation matrix we ultimately learned is independent of the trajectory length n. This provides us with a convenient means to measure the similarity of user activity preferences, as it is insensitive to the variations in trajectory lengths.

4.4. Multihead self-attention aggregation module

So far, we have obtained a latent feature matrix, denoted as $\mathbf{M_L}$, which combines both personalized frequency and sequence dependencies from the historical trajectory. However, utilizing it directly might yield sub-optimal results, as it lacks the aggregation of implicit dependencies and encoding of extensive spatio-temporal information. To aggregate the implicit correlation between different check-in records in historical trajectories from multiple perspectives and further update the representation of the latent feature matrix, we propose a Multihead Self-Attention Aggregation module (MSAA).

The MSAA module first computes multi-head matrix representations of the latent feature matrix $\mathbf{M}_\mathbf{L}^\top \in \mathbb{R}^{n\times q}$. This involves generating the matrices Q, K, and V, corresponding to Query, Key, and Value components of each head in the attention mechanism, as Eq. (11).

$$head_i = \mathbf{Attention}\left(\mathbf{M}_\mathbf{L}^\top W_i^Q, \mathbf{M}_\mathbf{L}^\top W_i^K, \mathbf{M}_\mathbf{L}^\top W_i^V, \mathbf{M_E}\right), \tag{11}$$

where W_i^Q, W_i^K, and W_i^V (all in $\mathbb{R}^{q\times d/n_h}$) are the Query, Key, and Value weight parameter matrices associated with the ith head, respectively, and n_h is the number of heads.

Distinct from conventional self-attention methodologies, MSAA incorporates an explicit feature matrix $\mathbf{M_E}$, enriching the model with additional contextual spatio-temporal information. This matrix will contribute to aggregate specific patterns, such as temporal and spatial intervals. Specifically, calculate the attention distribution of each head, as Eq. (12). To prevent the dot product of Q and $K^\top$ to be too large, and to control the scale of $\mathbf{M}_\mathbf{L}^\top W^Q (\mathbf{M}_\mathbf{L}^\top W^K)^\top \in \mathbb{R}^{n\times n}$ and the explicit feature matrix $M_E \in \mathbb{R}^{n\times n}$ within a reasonable range, we divide them by $\sqrt{d/n_h}$ to improve the stability. Then, we apply the $softmax$ operation to the scaled dot product and convert it into an attention-weight distribution. This distribution $\alpha = softmax(QK^\top + \mathbf{M_E}) \in \mathbb{R}^{n\times n}$ is used for the weighted sum Value matrix $V \in \mathbb{R}^{n\times d/n_h}$. It represents the attention probability of each underlying candidate POI for all visited locations. During the training process, by adjusting the matrix $\mathbf{M_E}$, the proposed model can be guided to incorporate explicit spatio-temporal relationships into the interaction of self-attention aggregation.

$$\mathbf{Attention}(Q, K, V, \mathbf{M_E}) = softmax\left(\frac{QK^\top + \mathbf{M_E}}{\sqrt{d/n_h}}\right)V, \tag{12}$$

Furthermore, each $head_i \in \mathbb{R}^{n\times d/n_h}$ is linearly amalgamated through an output weight matrix $\mathbf{W^O} \in \mathbb{R}^{d\times d}$ to obtain the attention head features $\mathbf{H_A} \in \mathbb{R}^{n\times d}$, as Eq. (13).

$$\mathbf{H_A} = \mathbf{Concat}\left(head_1, head_2, \ldots, head_{n_h}\right)\mathbf{W^O}, \tag{13}$$

where $\mathbf{Concat}(\cdot)$ function is used to concatenate the output matrix of the multi-head attention according to each column.

Finally, we perform a dropout and layer normalization operation on $\mathbf{H_A}$ to receive the final self-attention aggregation matrix, $\mathbf{M_A} = Dropout(LayerNorm(\mathbf{H_A})) \in \mathbb{R}^{n\times d}$, which can reflect implicit relationships between visited historical locations and potential candidate POIs.

4.5. Dual-driven user preference matching module

This module aims to learn the probabilities for visiting the next POI based on the trajectory semantic features of other similar-minded users and the user's own. Subsequently, it performs a weighted fusion of both to output the ultimate probability scores.

Cosine-based Preference Similarity Matching Layer. Let $\mathbf{M_R^i}$ and $\mathbf{M_R^j}$ denote the trajectory representation matrix of two users u_i and u_j, both with dimensions $q \times 2h$. The total sum of cosine similarities between the corresponding row vectors of the two users' trajectory representation matrices can be expressed as Eq. (14).

$$Sim\left(\mathbf{M_R^i}, \mathbf{M_R^j}\right) = \sum_{k=1}^{q} \frac{\mathbf{M_R^i}[k,:] \cdot \mathbf{M_R^j}[k,:]^\top}{\left\|\mathbf{M_R^i}[k,:]\right\| \cdot \left\|\mathbf{M_R^j}[k,:]\right\|}, \tag{14}$$

where $\mathbf{M_R}[k,:]$ denotes the kth row vector in the matrix $\mathbf{M_R}$ and $\|\cdot\|$ represents the Euclidean norm. The summation computes the total cosine similarity between corresponding row vectors of the two users' trajectory matrices. A larger value for Sim indicates a higher similarity between the trajectories.

For a target user u, predict their rating probability $\hat{p}_{l_i}^u$ for a POI l_i by searching for the k most similar-minded users based on cosine similarity, as Eq. (15).

$$\hat{p}_{l_i}^u = \frac{\sum_{j\in N(u)} Sim(u,j) \cdot p_{l_i}^j}{\sum_{j\in N(u)} |Sim(u,j)|}, \tag{15}$$

where $\hat{p}_{l_i}^u$ denotes the predicted rating probability of user u for POI l_i based on other similar-minded users, $N(u)$ represents the set of k most similar-minded users to the target user u, $Sim(u,j)$ signifies the cosine similarity between users u and j based on their trajectory representation matrices, and $p_{l_i}^j$ indicates the rating probability of user j for POI l_i.

Attention-based Candidate Location Matching Layer. For a target user u, given the location embedding matrix $\mathbf{E_L} = \{e_l^1, e_l^2, \ldots, e_l^L\} \in \mathbb{R}^{L\times d}$ and learned self-attention aggregation matrix $\mathbf{M_A} \in \mathbb{R}^{n\times d}$ from their historical trajectories, we calculate the probability score of each POI (location), as Eq. (16).

$$\hat{S}_u = \mathbf{WeightAgg}(softmax(\mathbf{E_L} \odot \mathbf{M_A}^\top)) \in \mathbb{R}^L, \tag{16}$$

Here, we apply a weighted aggregation on the last dimension of $softmax(\mathbf{E_L} \odot \mathbf{M_A}^\top) \in \mathbb{R}^{L\times n}$ and transform its dimension into $\mathbb{R}^L$ that reflects the impact of n historical check-in on L candidate locations. Each value $\hat{s}_{l_i}^u \in \hat{S}_u$ denotes the predicted score of user u for POI l_i based on the user's own trajectory features.

Considering that $\hat{s}_{l_i}^u$ is predominantly influenced by users' historical visit behaviors and patterns (i.e., preference-based), while $\hat{p}_{l_i}^u$ emphasizes more on the collective characteristics among similar-minded users, to achieve more accurate personalized POI recommendations accompanied by diversity, we performed a weighted fusion to strike a balance between the two aspects, as Eq. (17).

$$\hat{y}_{l_i}^u = \mathbf{Fusion}(\hat{p}_{l_i}^u, \hat{s}_{l_i}^u) = w_p^u \cdot \hat{p}_{l_i}^u + w_s^u \cdot \hat{s}_{l_i}^u, \tag{17}$$

where $\mathbf{Fusion}(\cdot)$ is a feed forward neural network activated by $softmax$, and $\hat{y}_{l_i}^u$ is the predicted preference probability of user u for POI l_i. The learned weights w_p^u and w_s^u derived from the fusion function vary for different users u to ensure adaptive personalized recommendations. Additionally, for inactive users, greater w_p^u weight is placed to alleviate the cold start issue and enhance recommendation diversity.

We utilize cross-entropy loss, as Eq. (18), to optimize all trainable parameters $\theta = \{W^*, b^*, d_*\}$, incorporating ℓ_2 regularization to mitigate overfitting.

$$\mathcal{L} = -\sum_{i=1}^{L} y_{l_i}^u \log\left(\hat{y}_{l_i}^u\right) + \left(1 - y_{l_i}^u\right)\log\left(1 - \hat{y}_{l_i}^u\right) + \lambda_\theta \|\theta\|_2^2, \tag{18}$$

where $y_{l_i}^u$ is the ground truth in which $y_{l_i}^u = 1$ when the user's actual check-in is POI l_i, otherwise, $y_{l_i}^u = 0$ and λ_θ is the corresponding weight of the ℓ_2 regularization.

5. Experiment and result analysis

5.1. Experiment settings

5.1.1. Datasets

We conduct our experiments on four widely used cities for the POI location recommendation task, namely New York City (NYC), Tokyo (TKY), Singapore (SIN), California and Nevada (CA), all collected from the two most popular location-based social networks (LBSNs) service providers, i.e., Foursquare (Yang et al., 2019, 2014) and Gowalla (Cho et al., 2011; Yuan et al., 2013). The basic statistical information of the four datasets is shown in Table 2. For different datasets, the mean

Table 2
Basic dataset statistics (# denotes the number of).

	NYC	TKY	SIN	CA
Period	Apr. 2012~Feb. 2013	Apr. 2012~Feb. 2013	Aug. 2010~Jul. 2011	Feb. 2009~Oct. 2010
#Users	1083	2293	2321	4318
#POIs (Locations)	5135	59,739	5596	9923
#Check-ins	147,938	565,100	194,108	250,780
#POI-categories	318	384	252	301
Mean of trajectory length	136.88	246.45	83.65	36.38
Std of trajectory length	161.27	220.74	105.09	52.30

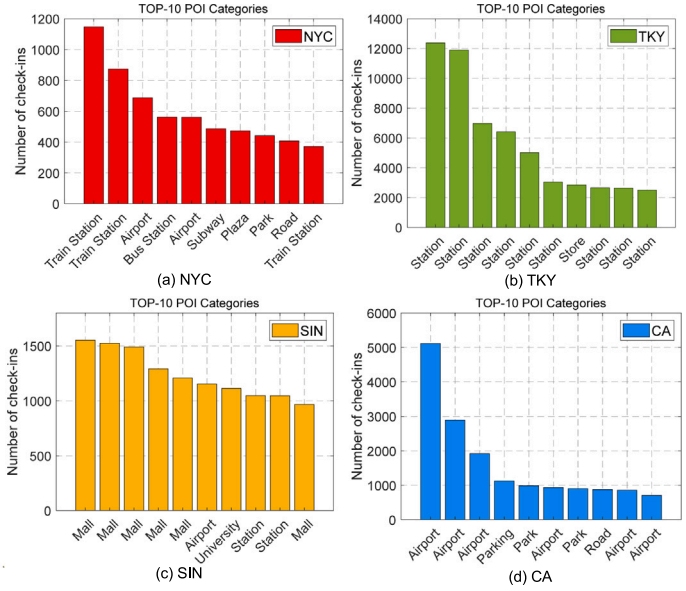


Fig. 6. The most frequently visited POI categories.

and standard deviation of trajectory length for each user are various, reflecting the check-in activity level of users in each city. Each record consists of the user ID, POI ID, POI's GPS-specific coordinates, POI category, and check-in time. For some locations without POI categories, we have supplemented them using the OpenStreetMap-based Nominatim tool (Clemens, 2015). The top ten most check-in location categories among the four cities are shown in Fig. 6. It can be seen that the four datasets are heterogeneous, with NYC and TKY having the most check-in POIs mainly being ground transportation stations, CA mostly being airports, and SIN being consumer shopping malls, reflecting significant differences in user preferences among four different cities.

5.1.2. Evaluation metrics

The final output of our model are probability scores of all POI candidate locations. Based on these predicted scores on different POIs, we recommend the Top-k POIs to potential interested users. We introduce three commonly utilized metrics to evaluate the recommendation performance of models. The larger values of the three metrics represent better performance results.

The first one is Accuracy@k (Acc@k) which measures the count of accurately predicted POIs within the Top-k recommendations, as Eq. (19), where k represents the number of items (POIs) recommended in the list. In this paper, we select $k = \{1, 5, 10, 20\}$ for testing.

$$Acc@k = \frac{|Hit|}{|Q_n|}, \tag{19}$$

where $|Hit|$ is the total count of the ground-truth POI that appears in the top-k recommendation list, and $|Q_n|$ indicates the number of testing queries.

The second indicator is the normalized discounted cumulative gain (NDCG@k), which evaluates the quality of recommendations within the top-k positions, as Eq. (20).

$$NDCG@k = \frac{1}{|Q_n|}\sum_{i=1}^{|Q_n|}\frac{DCG_i@k}{IDCG_i@k}, \tag{20}$$

where $DCG_i@k = \sum_{i=1}^{k}\frac{2^{rel_i}-1}{\log_2(i+1)}$ is the discounted cumulative gain, and rel_i represents the relativity of ith recommended POI and user ground-truth check-in POI. $IDCG_i@k$ is the ideal discounted cumulative gain representing the maximum possible gain that could be achieved with a perfect ranking.

The third indicator is the mean reciprocal rank (MRR), which measures the average position of the correct recommended POI, as Eq. (21).

$$MRR = \frac{1}{|Q_n|}\sum_{i=1}^{|Q_n|}\frac{1}{rank_i}, \tag{21}$$

where $rank_i$ is the rank of the correct recommended POI in testing queries Q_n. If the recommended POI is not in the top-k positions, then $rank_i = \infty$.

5.1.3. Hyper-parameters selection

During the training phase, we select the Adam optimizer to update all parameters of the model with $learning\_rate = 2e^{-3}$, $weight\_decay = 3e^{-4}$, and $epoch = 100$. Three embedding sizes of user, location, and time-conscious category all uniformly are $d = d_u = d_l = d_c = 64$ for NYC, SIN, and CA, $d = 96$ for TKY. The number of heads n_h in the multi-head self-attention aggregation module (MSAA) is set to $\{1, 2, 4, 2\}$ for NYC, TKY, SIN, and CA, respectively. All experiments were conducted in a Pytorch environment under a Linux server with one NVIDIA Tesla V100-SXM2 GPU card with 32 GB memory. For all baseline methods, we tested them by applying their default hyper-parameter settings.

5.1.4. Baseline methods

To verify the performance of the proposed CTRNext model, the following eight representative baseline methods are used in the empirical study.

- FPMC (Rendle et al., 2010) integrates two foundational paradigms, Matrix Factorization (MF) and Markov Chain (MC), to perceive user preferences and characterize transitions between visited locations.
- PRME (Feng et al., 2015) designs a novel pair-wise embedding strategy called personalized ranking metric embedding to explore POIs continuing transition.
- ST-RNN (Liu et al., 2016) constructs temporal and spatial specific transition matrices based on traditional RNNs to learn local contextual information of POI sequences.
- STAN (Luo et al., 2021) utilizes two types of self-attention structures and a multi-modality embedding to aggregate point-to-point interactions which is capable of forecasting the next POI to users among non-adjacent and non-successive check-in records.
- NeuNext (Zhao et al., 2022) is an improved version of the spatiotemporal gated network (STGN) (Zhao et al., 2019) which extends new gates on the basis of LSTM to be competent in learning dependency relationships between successive visits.

Table 3
Recommendation performance comparison with baselines on four heterogeneous datasets.

Dataset	Model	Acc@1	Acc@5	Acc@10	Acc@20	NDCG@5	NDCG@10	NDCG@20	MRR
NYC	FPMC	0.1003	0.2126	0.2970	0.3323	0.1652	0.1972	0.2152	0.1701
	PRME	0.1159	0.2236	0.3105	0.3643	0.1655	0.1982	0.2162	0.1712
	ST-RNN	0.1483	0.2923	0.3622	0.4502	0.2415	0.2771	0.3017	0.2193
	NeuNext	0.1801	0.3469	0.4281	0.5259	0.2842	0.3346	0.3680	0.2801
	PLSPL	0.1917	0.3678	0.4523	0.5370	0.2884	0.3358	0.3682	0.2806
	STAN	0.2292	0.4697	0.5849	0.6459	0.3397	0.3571	0.3822	0.3293
	SLS-REC	0.2344	0.4712	0.5908	0.6675	0.3402	0.3658	0.3892	0.3423
	GETNext	0.2416	0.4990	0.6151	0.6885	0.3595	0.3850	0.4082	0.3641
	CTRNext	**0.2598**	**0.5285**	**0.6438**	**0.7201**	**0.3880**	**0.4052**	**0.4308**	**0.3902**
	Improvement	+7.53%	+5.91%	+4.67%	+4.59%	+7.93%	+5.25%	+5.54%	+7.17%
TKY	FPMC	0.0934	0.2079	0.2800	0.3492	0.1397	0.1785	0.1862	0.1401
	PRME	0.1068	0.2289	0.3094	0.3689	0.2005	0.2507	0.2750	0.1899
	ST-RNN	0.1591	0.3090	0.3608	0.4815	0.2372	0.2783	0.3069	0.2302
	NeuNext	0.1711	0.3402	0.4015	0.4799	0.2690	0.3008	0.3299	0.2564
	PLSPL	0.1893	0.3644	0.4500	0.4911	0.2721	0.3099	0.3450	0.2611
	STAN	0.2095	0.3874	0.4593	0.5203	0.3001	0.3253	0.3647	0.2983
	SLS-REC	0.2117	0.4255	0.5022	0.5578	0.3217	0.3675	0.3854	0.3162
	GETNext	0.2296	0.4428	0.5299	0.5849	0.3395	0.3865	0.4109	0.3299
	CTRNext	**0.2783**	**0.5042**	**0.5801**	**0.6457**	**0.3810**	**0.4305**	**0.4651**	**0.3891**
	Improvement	+21.21%	+13.87%	+9.47%	+10.39%	+12.22%	+11.38%	+13.19%	+17.94%
SIN	FPMC	0.0532	0.1149	0.1561	0.1711	0.0683	0.0944	0.1172	0.0759
	PRME	0.0744	0.1352	0.1741	0.1907	0.1108	0.1355	0.1432	0.1117
	ST-RNN	0.1101	0.1500	0.2010	0.2419	0.1195	0.1396	0.1485	0.1201
	NeuNext	0.1360	0.1762	0.2374	0.2706	0.1306	0.1508	0.1698	0.1410
	PLSPL	0.1490	0.1877	0.2536	0.2958	0.1399	0.1599	0.1735	0.1452
	STAN	0.2098	0.2605	0.3465	0.3771	0.1892	0.2001	0.2345	0.1900
	SLS-REC	0.2307	0.2805	0.3549	0.4078	0.2001	0.2149	0.2398	0.2058
	GETNext	0.2334	0.2860	0.3674	0.4195	0.2003	0.2155	0.2407	0.2060
	CTRNext	**0.2531**	**0.3010**	**0.3942**	**0.4560**	**0.2159**	**0.2296**	**0.2538**	**0.2201**
	Improvement	+8.44%	+5.24%	+7.29%	+8.70%	+7.79%	+6.54%	+5.44%	+6.84%
CA	FPMC	0.0383	0.0702	0.1159	0.1682	0.0923	0.1201	0.1358	0.0911
	PRME	0.0521	0.1034	0.1425	0.1954	0.1074	0.1259	0.1366	0.1002
	ST-RNN	0.0799	0.1423	0.1940	0.2477	0.1433	0.1596	0.1649	0.1429
	NeuNext	0.0961	0.2097	0.2613	0.3245	0.1912	0.2423	0.2707	0.1712
	PLSPL	0.1072	0.2278	0.2995	0.3401	0.1986	0.2495	0.2801	0.1847
	STAN	0.1104	0.2348	0.3018	0.3502	0.2049	0.2507	0.2819	0.1869
	SLS-REC	0.1362	0.2833	0.3357	0.4055	0.2336	0.2692	0.2951	0.2034
	GETNext	0.1357	0.2852	0.3590	0.4241	0.2327	0.2685	0.3009	0.2103
	CTRNext	**0.1591**	**0.3051**	**0.3805**	**0.4533**	**0.2491**	**0.2841**	**0.3182**	**0.2243**
	Improvement	+16.81%	+6.98%	+5.99%	+6.89%	+6.64%	+5.53%	+5.75%	+6.66%

* The improvement rate refers to the performance enhancement of the proposed model (**bold**) over the suboptimal baseline (underline).

- PLSPL (Wu et al., 2022) takes advantage of the attention and LSTM to capture users' long- and short-term preference patterns, respectively, and integrates the two through one user-based combination layer.
- GETNext (Yang et al., 2022) leverages the spectral GCNs to capture enhanced trajectory flows between different POIs, and a Transformer framework to decode global collaborative signals for the next POI recommendation.
- SLS-REC (Fu et al., 2024) designs a spatio-temporal Hawkes attention hypergraph network and a dynamic propagation-based GNN to capture users' short-term and long-term preferences, respectively.

5.2. Performance comparison with baselines

Table 3 displays the comparative recommendation performance results between the proposed model and baselines across four heterogeneous datasets. The proposed CTRNext model achieved notable improvements on three metrics—$Acc@k$, $NDCG@k$, and MRR—by +5%∼+21%, +5%∼+12%, and +6%∼+18%, respectively. In particular, when compared to the second-best GETNext approach, CTRNext exhibits a significant increase in Top-1 accuracy by +7.53% on NYC, +21.21% on TKY, +8.44% on SIN, and +16.81% on CA. It reflects the model's precision in predicting the most relevant POI for users. This improvement is primarily attributed to our model's comprehensive learning of semantic interactions within similar-minded user check-in behaviors and the capture of higher-order collaborative features from their trajectories.

In terms of baseline methods, self-attention models such as PLSPL, STAN, and GETNext outperform sequence-based approaches. Specifically, ST-RNN directly leverages the time and spatial intervals between consecutive visits as explicit information to enhance model performance, showing the importance of explicit contextual information. However, it requires a substantial volume of user historical point-of-interest sequence data for training, potentially facing a user cold-start problem. NeuNext mitigates the user cold start by extending novel gating mechanisms, yet it falls short in effectively modeling check-ins between non-adjacent locations. To aggregate non-adjacent visits, STAN explicitly utilizes the relative spatiotemporal interval of all check-in records along its trajectory by taking advantage of self-attention layers. Its performance has been significantly improved, but the computational cost is relatively high. SLS-REC uses two kinds of graph structures for self-supervised contrastive learning to characterize the long-term and short-term preferences of users. GETNext also performs commendably on the NDCG metric, indicating its capability to POI recommendations. This could be attributed to GETNext's more advanced transformer mechanisms for understanding the complex patterns in user behavior.

In terms of heterogeneous datasets, the performance improvement rate of our proposed CTRNext model is not consistent across the four datasets. The most significant improvement rate is on TKY, with all indicators greater than +10% except for $Acc@10 = +9.47\%$. In contrast, the other three datasets have approaching improvement rates, ranging from an average of +5% to +8%. This variability may be influenced by a

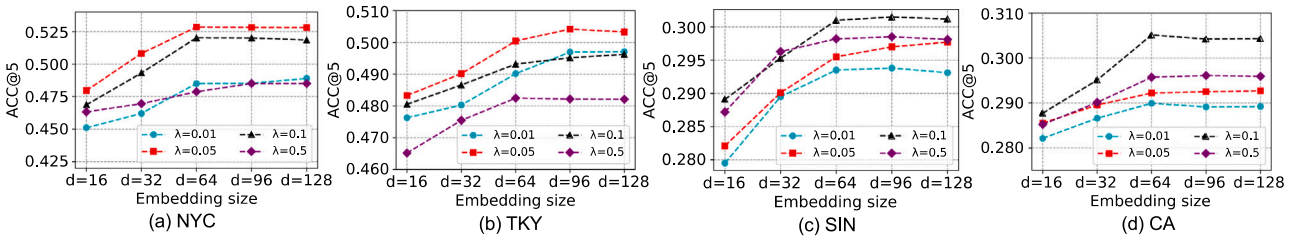


Fig. 7. Performance of CTRNext under different values of the embedding size d and time decay rate λ.

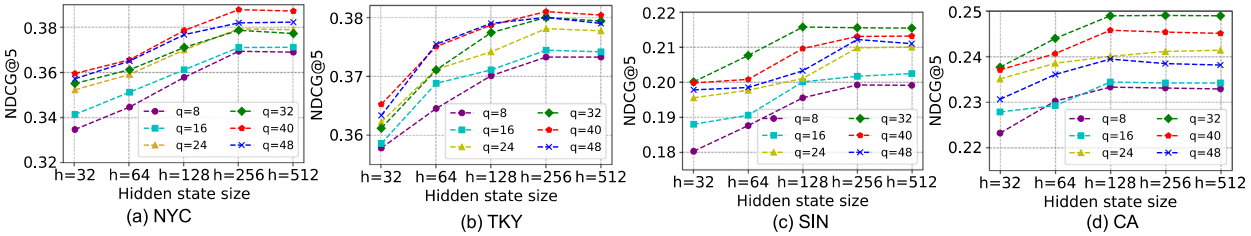


Fig. 8. Performance of CTRNext under different values of the hidden state size h and the number of attention channels q.

combination of three factors: the number of users, the amount of check-ins, and the length of trajectories. From Table 2, it becomes evident that these factors are all the largest on the TKY dataset compared to the others. As the number of users increases, the range of collaborative signals that CTRNext can perceive from potentially similar-minded users will also expand. Although SIN and TKY have a close number of users, the mean trajectory length in the SIN dataset is only 1/3 of TKY's. The number of check-ins and trajectory length reflect the level of user activity from global and local (individual) perspectives. As the value of both increases, it provides more historical movement behavior for CTRNext. Thus, the superior performance of CTRNext on TKY can be attributed to a larger user base and more extensive trajectory data, offering deeper insights into user behavior patterns.

5.3. Hyper-parameter sensitivity

We investigated the impact of four crucial hyper-parameters (i.e., embedding size d, time decay rate λ, hidden state size h, and attention channel quantity q) on the recommendation accuracy ($Acc@k$) and ranking quality ($NDCG@k$) of the CTRNext model. Figs. 7 and 8 depicts the final results ($k = 5$), from which we draw the following conclusions:

(i) Four hyper-parameters need to be fine-tuned based on the heterogeneity across the four different datasets. Variations in individual-users' visited POI categories and activity levels lead to differences in user behavioral preferences, trajectory lengths, and the number of inactive users.

(ii) The embedding size d is set to $\{16, 32, 64, 96, 128\}$ and time decay rate λ is set to $\{0.01, 0.05, 0.1, 0.5\}$. Increasing the embedding size d enables CTRNext to capture richer semantic interaction features within trajectories, while the time decay rate λ reflects the degree of dependency on user periodic preferences. The CTRNext model achieves its highest recommendation accuracy when ($d = 64, \lambda = 0.05$) on NYC dataset, ($d = 96, \lambda = 0.05$) on TKY, and ($d = 64, \lambda = 0.1$) on both SIN and CA datasets.

(iii) The hidden state size h is set to $\{32, 64, 128, 256, 512\}$ and attention channel quantity q is from 8 to 48 growing with a step size of 8. The optimal ranking performance is achieved when ($h = 256, q = 40$) on both NYC and TKY datasets, and ($h = 128, q = 32$) on both SIN and CA datasets. When both parameters are increased to their optimal values, there is a gradual improvement in $NDCG$ until they reach their optimal values. Further increases beyond these optimal only lead to marginal improvements or even a decline.

(iv) The hyper-parameters h and q determine the dimensions of a representation matrix $\mathbf{M_R}$. When h exceeds 256, it will cause redundancy in the temporal features of check-in records. When q is less than 32, some complex trajectories (e.g., check-in records with numerous POI categories, or significant differences in distances between consecutive POI visits by the same user) become difficult to attention aggregation, resulting in representation deficiency.

5.4. Ablation study

In this section, an ablation study is undertaken to assess the individual contributions of each component within the proposed model toward recommendation performance. Specifically, six variant models are conducted.

- **Full Model**: the proposed CTRNext model.
- **w/o Freq**: remove frequency interaction matrix $\mathbf{M_F}$ and replace it with a new matrix (dimension is the same as $\mathbf{M_F}$) learned from a fully connected layer.
- **w/o TimeCat**: remove the time-conscious category embedding e_c and replace it with only category embedding without involving temporal dynamics.
- **w/o Explicit**: remove the explicit spatiotemporal interval embedding $\mathbf{M_E}$ during multi-head self-attention aggregation in Eq. (12).
- **w/o MSAA**: replace the multi-head self-attention aggregation module with an ordinary attention layer that neglects the aggregation of implicit correlation in trajectories from multiple perspectives.
- **w/o TSSC**: remove the trajectory semantic similarity calculation module and the $\mathbf{M_L}$ originally inputted to the MSAA is replaced with the $\mathbf{M_J}$.

Fig. 9 depicts the outcomes derived from the ablation study. Through sequentially eliminating the six key components, we have the following observations: among these components, "TSSC" and "MSAA" have the most significant impact on model performance (both are also the inputs of the dual-driven user preference matching

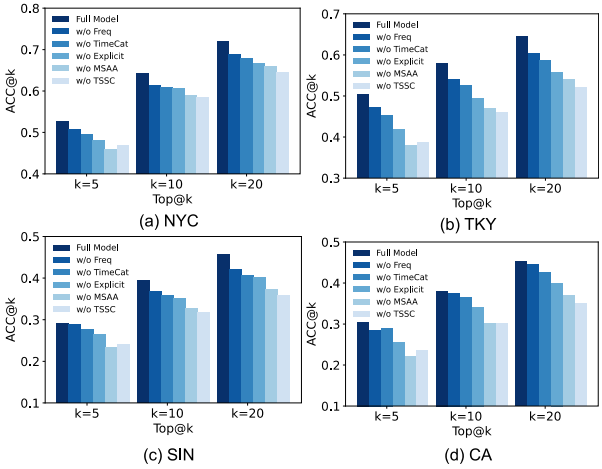


Fig. 9. Ablation study results on four datasets.

module), followed by "Explicit", then "TimeCat", and finally "Freq". When the number of recommended POIs is small ($k = 5$), the decrease on the Acc@k metric is more significant with w/o MSAA compared to w/o TSSC. However, as the number of recommendations increases ($k = 10, 20$), the TSCC plays a more important role than MSAA. This finding underscores the importance of leveraging collaborative information from similar user trajectories to better comprehend user preferences and to learn their intent when visiting the next POI. Because compared to only learning the historical trajectory of the target user, the trajectory semantic similarity calculation (i.e., user–user interaction) enhances the diversity of recommendations. Exclusively relying on an individual user's trajectory representation demonstrates inadequate in generating diverse recommendations, potentially leading to recommended POI categories being the same or homogeneous. In addition, we have observed that "Freq" is more important than "TimeCat" on the CA dataset compared to other datasets for Acc@5. This could be attributed to the larger regions within CA, where some POIs exhibit a more sparse distribution, resulting in certain time intervals having limited temporal correlations among POI categories.

5.5. Research on model robustness

To test the robustness of models under data interference, we randomly deleted a certain proportion $p\% = \{0\%, 10\%, 20\%, 30\%\}$ of historical trajectory check-in data. The comparison results with the two most competitive models (STAN and GETNext) on four datasets are shown in Fig. 10. Overall, all three models exhibit some performance degradation due to data sparsity and incompleteness. STAN's performance on the Acc metric decreased by 7.96%~30.18%, GETNext decreased by 5.96%~28.95%, and Ours decreased by 5.13%~22.91% across four datasets. For the MRR metric, the performance degradation of the STAN ranges from 5.03% to 25.02%, GETNext from 5.00% to 20.98%, and Ours from 4.50% to 18.06%. The decrease in recommendation accuracy Acc of the model is significantly faster than that of ranking quality MRR. However, the proposed model did not cause severe linear decay at higher missing proportions. Especially on the TKY dataset, our model has only decreased by 5.94%, 11.58%, and 14.68% on the Acc metric, respectively, with the missing proportion $p\%$ equals to 10%, 20%, and 30%. This is mainly because our model has the ability to acquire the POI transition probability and capture collaborative information between similar-minded users, rather than relying solely on the historical check-in trajectory of a single user. Due to the implementation of point-to-point attention interaction between different POI locations in the self-attention aggregation layer of the STAN model, performance degradation is significant when some locations are deleted (for example, Acc of STAN changes from 0.2292 to 0.1601 on the NYC when $p\%$ from 0% to 30%). The global collective trajectory flow used in the GETNext model can enhance the POI recommendation for inactive users and alleviate the cold start problem of short trajectories. At the same time, the captured generic movement pattern of GETNext also overcomes the data sparsity problem caused by data deletion to a certain extent (when the missing proportion $p\%$ is below 20%).

Based on this study, it is evident that our proposed model demonstrates superior accuracy and robustness in recommending POI compared to baseline methods, particularly in scenarios where user check-in data is inherently incomplete and sparse.

5.6. Case study and interpretability analysis

In order to understand the ability of the proposed model to provide insight into trajectory features among similar-minded users and enhance the recommendation diversity, we conducted a case study on the Tokyo dataset and analyzed the interpretability of CTRNext. Specifically, we selected two similar-minded users' check-in trajectories (UserID = 5,12), which are visualized in Fig. 11(a). User 12 is more active compared to User 5, with a wider distribution of POI locations visited. To illustrate some similar behaviors and preferences between two users, we have marked the locations of several special POIs, where POIs 1~6 all represent train stations, and 0 is an electronics store. POI 2 is the location where two users check in most frequently, which User 5 visited 39 times and User 12 visited 32 times. Other train stations, such as 1,3,4,5, are also regularly visited (i.e., with similar explicit time intervals) by two users. Fig. 11(b) shows the attention scores learned by the proposed model. The element $corr(i, j)$ of the correlation weight matrix indicates the degree of impact of User 12 visiting POI i on User 5 visiting POI j. We can see that the upper left corner of the matrix (POIs 0,1,2) has a strong positive correlation. The correlation $corr(1, 2) = 0.87$ between POIs 1 and 2 is easy to understand as they are both train stations and geographically close. Users tend to check in at the surrounding points of interest, and there is a spatial clustering phenomenon.

The positive correlation $corr(0, 1) = 0.63$ and $corr(0, 2) = 0.85$ between POI 0 and POI 1, POI 2 is not easily intuitively understood (i.e., it is implicit). By querying the check-in trajectories of two users, we found that the actual situation is user 12 visited POI 0 a total of 10 times, all within a period of time after visiting POI 2. User 5 has only visited POI 0 once, so it can be determined that POI 0 (an electronics store) is a cold start item that user 5 has never visited before. This case study demonstrates that the proposed model not only captures correlations of historical trajectories for user 5 himself but also captures other users' collaborative features and spatio-temporal contextual interactions in the trajectories (user 12) with similar habits and latent preferences. The proposed CTRNext model successfully recommends the new item POI 0 which might be of interest to User 5, which enhances the diversity of recommendations. In addition to nearby locations, some POIs that are not close in distance also have correlations. For example, the strongly correlated $corr(1, 6) = 0.65$ illustrates that the model's ability to perceive the semantic similarity of trajectories for similar-minded users does not rely solely on the explicit spatial distance.

6. Conclusion

Current POI recommendations, known as preference-based or catered recommendations, rely heavily on users' historical preferences and behaviors. In this paper, we propose a CTRNext method that aims to encourage users to broaden their scope of interests by capturing collaborative signals from potentially similar-minded users, which enhances diversity and provides a more enriched and personalized POI recommendation experience. In case of incomplete and sparse user check-in records, it can be inferred from the model robustness experiment that our proposed CTRNext model demonstrates superior

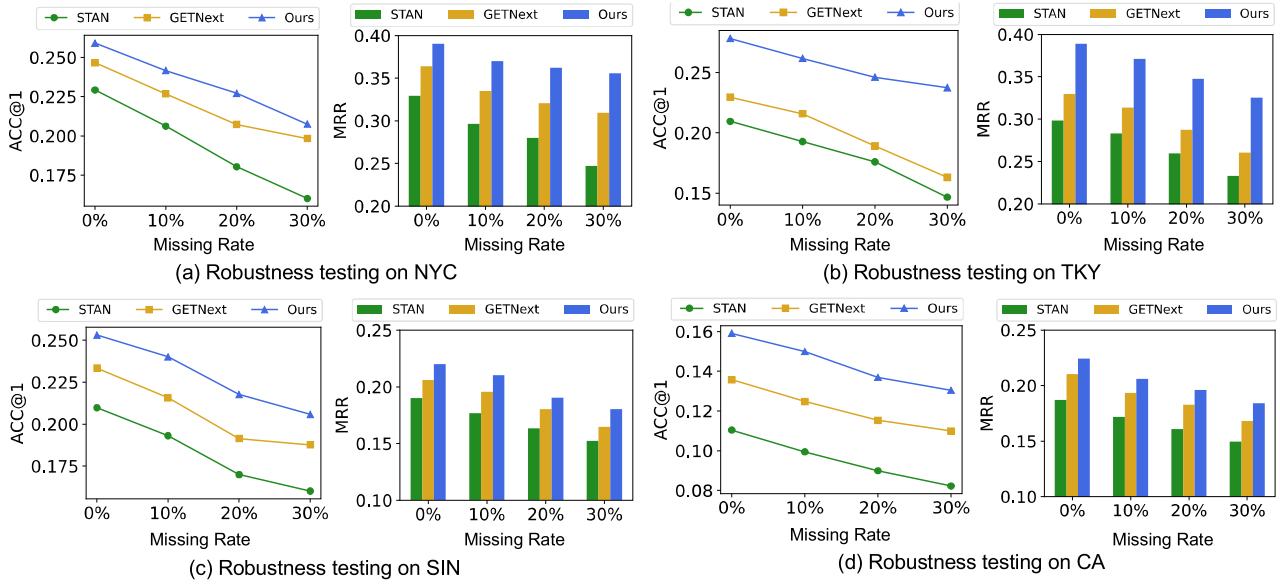


Fig. 10. Comparison of model performance with different missing ratios.

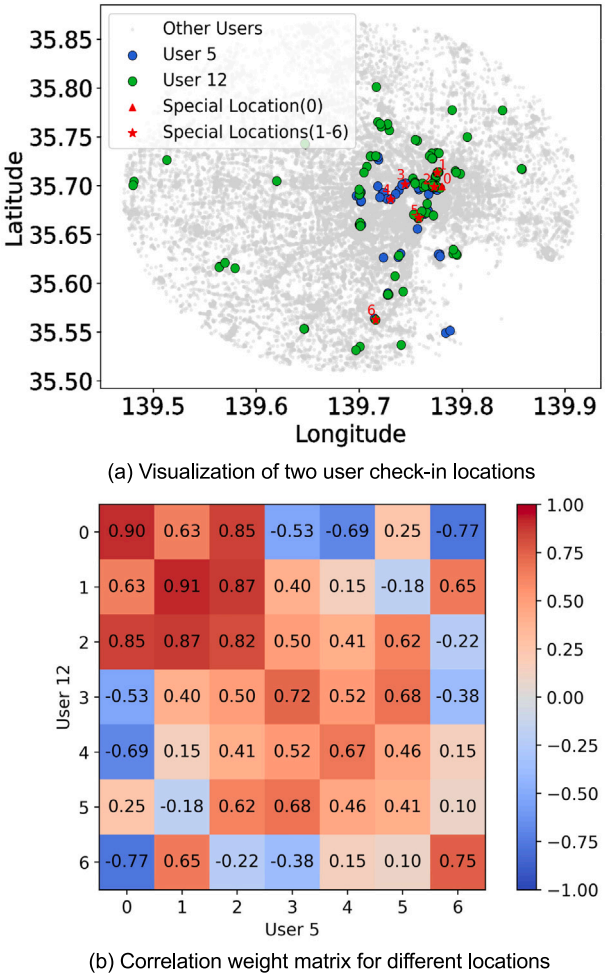


Fig. 11. The case study and learned correlation matrix.

accuracy and robustness compared to other baseline methods. Specifically, CTRNext model presented an average performance improvement of 8.99% in Acc@k, 7.76% in NDCG@k, and 9.65% in MRR. In addition, through case studies, we found that the captured collective movement patterns from a dual-driven user preference matching module are also helpful in alleviating cold start problems. Despite the promising results, our study has two limitations. Firstly, the CTRNext model's reliance on trajectory data may limit its applicability in scenarios where such data is unavailable. Secondly, while our model enhances diversity in recommendations, it may still struggle with extreme cases of user sparsity or where user preferences are highly volatile.

In the future, we will attempt to apply diversity penalties and proactive interactions to construct an exploratory-guided POI recommendation model, guiding users by leveraging both their historical check-in records and encouraging exploration of new locations or items that might inspire users' interest beyond their past preferences. Explore the use of large models, such as pre-trained large language models (LLMs) to further enhance POI recommendations, which can capture complex patterns and relationships from wide knowledge, providing deeper insights into user preferences.

CRediT authorship contribution statement

Jiankai Zuo: Formal analysis, Methodology, Software, Validation, Visualization, Writing – original draft. **Yaying Zhang:** Conceptualization, Investigation, Funding acquisition, Project administration, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The code and data are available at https://github.com/JKZuo/CTRNext.

Acknowledgments

This work was partly supported by the National Key Research and Development Program of China under Grant (2022YFB4501700), National Natural Science Foundation of China under Grant (72342026) and Fundamental Research Funds for the Central Universities under Grant (No. 2024-6-ZD-02).

References

Bok, Kyoungsoo, Lee, Chunghui, & Yoo, Jaesoo (2019). Recommending similar users using moving patterns in mobile social networks. *Computers & Electrical Engineering*, *77*, 47–60.

Chen, Lisi, Shang, Shuo, Jensen, Christian S., Yao, Bin, & Kalnis, Panos (2020). Parallel semantic trajectory similarity join. In *2020 IEEE 36th international conference on data engineering* (pp. 997–1008). IEEE.

Chen, Yudong, Wang, Xin, Fan, Miao, Huang, Jizhou, Yang, Shengwen, & Zhu, Wenwu (2021). Curriculum meta-learning for next POI recommendation. In *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining* (pp. 2692–2702).

Chen, Lei, Zhu, Guixiang, Liang, Weichao, & Wang, Youquan (2023). Multi-objective reinforcement learning approach for trip recommendation. *Expert Systems with Applications*, *226*, Article 120145.

Cheng, Chen, Yang, Haiqin, Lyu, Michael R., & King, Irwin (2013). Where you like to go next: Successive point-of-interest recommendation. In *Twenty-third international joint conference on artificial intelligence*.

Cho, Eunjoon, Myers, Seth A., & Leskovec, Jure (2011). Friendship and mobility: user movement in location-based social networks. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 1082–1090).

Clemens, Konstantin (2015). Geocoding with openstreetmap data. *Vol. 10*, In *GEOProcessing 2015*.

Fang, Jinfeng, & Meng, Xiangfu (2022). URPI-GRU: An approach of next POI recommendation based on user relationship and preference information. *Knowledge-Based Systems*, *256*, Article 109848.

Feng, Shanshan, Li, Xutao, Zeng, Yifeng, Cong, Gao, Chee, Yeow Meng, & Yuan, Quan (2015). Personalized ranking metric embedding for next new poi recommendation. In *Proceedings of the AAAI conference on artificial intelligence*.

Fu, Jiarun, Gao, Rong, Yu, Yonghong, Wu, Jia, Li, Jing, Liu, Donghua, & Ye, Zhiwei (2024). Contrastive graph learning long and short-term interests for POI recommendation. *Expert Systems with Applications*, *238*, Article 121931.

Furtado, Andre Salvaro, Kopanaki, Despina, Alvares, Luis Otavio, & Bogorny, Vania (2016). Multidimensional similarity measuring for semantic trajectories. *Transactions in GIS*, *20*(2), 280–298.

Islam, Md. Ashraful, Mohammad, Mir Mahathir, Das, Sarkar Snigdha Sarathi, & Ali, Mohammed Eunus (2022). A survey on deep learning based Point-of-Interest (POI) recommendations. *Neurocomputing*, *472*, 306–325.

Kazemi, Seyed Mehran, Goel, Rishab, Eghbali, Sepehr, Ramanan, Janahan, Sahota, Jaspreet, Thakur, Sanjay, Wu, Stella, Smyth, Cathal, Poupart, Pascal, & Brubaker, Marcus (2019). Time2vec: Learning a vector representation of time. arXiv preprint arXiv:1907.05321.

Li, Miao, Zheng, Wenguang, Xiao, Yingyuan, Zhu, Ke, & Huang, Wei (2021). Exploring temporal and spatial features for next POI recommendation in LBSNs. *IEEE Access*, *9*, 35997–36007.

Liu, Qiang, Wu, Shu, Wang, Liang, & Tan, Tieniu (2016). Predicting the next location: A recurrent model with spatial and temporal contexts. *Vol. 30*, In *Proceedings of the AAAI conference on artificial intelligence*.

Liu, An, Zhang, Yifan, Zhang, Xiangliang, Liu, Guanfeng, Zhang, Yanan, Li, Zhixu, Zhao, Lei, Li, Qing, & Zhou, Xiaofang (2020). Representation learning with multi-level attention for activity trajectory similarity computation. *IEEE Transactions on Knowledge and Data Engineering*, *34*(5), 2387–2400.

Long, Wangchen, Li, Tao, Xiao, Zhu, Wang, Dong, Zhang, Rui, Regan, Amelia C., Chen, Hongyang, & Zhu, Yongdong (2022). Location prediction for individual vehicles via exploiting travel regularity and preference. *IEEE Transactions on Vehicular Technology*, *71*(5), 4718–4732.

Luo, Yingtao, Liu, Qiang, & Liu, Zhaocheng (2021). STAN: Spatio-temporal attention network for next location recommendation. In *Proceedings of the web conference 2021* (pp. 2177–2185).

Meng, Xiangfu, & Fang, Jinfeng (2020). A diverse and personalized poi recommendation approach by integrating geo-social embedding relations. *IEEE Access*, *8*, 226309–226323.

Perifanis, Vasileios, Drosatos, George, Stamatelatos, Giorgos, & Efraimidis, Pavlos S. (2023). FedPOIRec: Privacy-preserving federated poi recommendation with social influence. *Information Sciences*, *623*, 767–790.

Rao, Xuan, Chen, Lisi, Liu, Yong, Shang, Shuo, Yao, Bin, & Han, Peng (2022). Graph-flashback network for next location recommendation. In *Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining* (pp. 1463–1471).

Rendle, Steffen, Freudenthaler, Christoph, & Schmidt-Thieme, Lars (2010). Factorizing personalized markov chains for next-basket recommendation. In *Proceedings of the 19th international conference on World wide web* (pp. 811–820).

Sánchez, Pablo, & Bellogín, Alejandro (2022). Point-of-interest recommender systems based on location-based social networks: a survey from an experimental perspective. *ACM Computing Surveys*, *54*(11s), 1–37.

Seyedhoseinzadeh, Kosar, Rahmani, Hossein A., Afsharchi, Mohsen, & Aliannejadi, Mohammad (2022). Leveraging social influence based on users activity centers for point-of-interest recommendation. *Information Processing & Management*, *59*(2), Article 102858.

Wang, Wei, Chen, Junyang, Wang, Jinzhong, Chen, Junxin, Liu, Jinquan, & Gong, Zhiguo (2019). Trust-enhanced collaborative filtering for personalized point of interests recommendation. *IEEE Transactions on Industrial Informatics*, *16*(9), 6124–6132.

Wang, Xinfeng, Fukumoto, Fumiyo, Cui, Jin, Suzuki, Yoshimi, Li, Jiyi, & Yu, Dongjin (2023). Eedn: Enhanced encoder-decoder network with local and global context learning for poi recommendation. In *Proceedings of the 46th international ACM SIGIR conference on research and development in information retrieval* (pp. 383–392).

Wang, Zhaobo, Zhu, Yanmin, Zhang, Qiaomei, Liu, Haobing, Wang, Chunyang, & Liu, Tong (2022). Graph-enhanced spatial-temporal network for next POI recommendation. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, *16*(6), 1–21.

Werneck, Heitor, Santos, Rodrigo, Silva, Nícollas, Pereira, Adriano C. M., Mourão, Fernando, & Rocha, Leonardo (2021). Effective and diverse POI recommendations through complementary diversification models. *Expert Systems with Applications*, *175*, Article 114775.

Wu, Yuxia, Li, Ke, Zhao, Guoshuai, & Qian, Xueming (2022). Personalized long-and short-term preference learning for next POI recommendation. *IEEE Transactions on Knowledge and Data Engineering*, *34*(4), 1944–1957.

Xia, Jiangnan, Yang, Yu, Wang, Senzhang, Yin, Hongzhi, Cao, Jiannong, & Philip, S. Yu (2023). Bayes-enhanced multi-view attention networks for robust POI recommendation. *IEEE Transactions on Knowledge and Data Engineering*.

Xiao, Zhu, Xu, Shenyuan, Li, Tao, Jiang, Hongbo, Zhang, Rui, Regan, Amelia C., & Chen, Hongyang (2020). On extracting regular travel behavior of private cars based on trajectory data analysis. *IEEE Transactions on Vehicular Technology*, *69*(12), 14537–14549.

Yan, Xiaodong, Song, Tengwei, Jiao, Yifeng, He, Jianshan, Wang, Jiaotuan, Li, Ruopeng, & Chu, Wei (2023). Spatio-temporal hypergraph learning for next POI recommendation. In *Proceedings of the 46th international ACM SIGIR conference on research and development in information retrieval* (pp. 403–412).

Yang, Song, Liu, Jiamou, & Zhao, Kaiqi (2022). GETNext: trajectory flow map enhanced transformer for next POI recommendation. In *Proceedings of the 45th international ACM SIGIR conference on research and development in information retrieval* (pp. 1144–1153).

Yang, Dingqi, Qu, Bingqing, Yang, Jie, & Cudre-Mauroux, Philippe (2019). Revisiting user mobility and social relationships in lbsns: a hypergraph embedding approach. In *The world wide web conference* (pp. 2147–2157).

Yang, Dingqi, Zhang, Daqing, Zheng, Vincent W., & Yu, Zhiyong (2014). Modeling user activity preference by leveraging user spatial temporal characteristics in LBSNs. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *45*(1), 129–142.

Yin, Feiyu, Liu, Yong, Shen, Zhiqi, Chen, Lisi, Shang, Shuo, & Han, Peng (2023). Next POI recommendation with dynamic graph and explicit dependency. *Vol. 37*, In *Proceedings of the AAAI conference on artificial intelligence* (pp. 4827–4834).

Yu, Dongjin, Yu, Ting, Wu, Yiyu, & Liu, Chengfei (2022). Personalized recommendation of collective points-of-interest with preference and context awareness. *Pattern Recognition Letters*, *153*, 16–23.

Yuan, Quan, Cong, Gao, Ma, Zongyang, Sun, Aixin, & Thalmann, Nadia Magnenat (2013). Time-aware point-of-interest recommendation. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval* (pp. 363–372).

Zang, Hongyu, Han, Dongcheng, Li, Xin, Wan, Zhifeng, & Wang, Mingzhong (2021). Cha: Categorical hierarchy-based attention for next poi recommendation. *ACM Transactions on Information Systems (TOIS)*, *40*(1), 1–22.

Zhao, Guoshuai, Lou, Peiliang, Qian, Xueming, & Hou, Xingsong (2020). Personalized location recommendation by fusing sentimental and spatial context. *Knowledge-Based Systems*, *196*, Article 105849.

Zhao, Pengpeng, Luo, Anjing, Liu, Yanchi, Xu, Jiajie, Li, Zhixu, Zhuang, Fuzhen, Sheng, Victor S., & Zhou, Xiaofang (2022). Where to go next: A spatio-temporal gated network for next poi recommendation. *IEEE Transactions on Knowledge and Data Engineering*, *34*(5), 2512–2524.

Zhao, Pengpeng, Zhu, Haifeng, Liu, Yanchi, Xu, Jiajie, Li, Zhixu, Zhuang, Fuzhen, Sheng, Victor S., & Zhou, Xiaofang (2019). Where to go next: A spatio-temporal gated network for next POI recommendation. In *33rd AAAI conference on artificial intelligence* (pp. 5877–5884).