Dual Contrastive Learning for Next POI Recommendation with Long and Short-Term Trajectory Modeling

Zhi Liu, Junhui Deng, Deju Zhang, Zhiyu Chen, Guojiang Shen, Xiangjie Kong*
College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China

Abstract

Next point-of-interest (POI) recommendation is a challenging task that aims to recommend the next location that a user may be interested in based on their check-in trajectories. Since users travel not only with long-term stable preferences but also with short-term dynamic interests, there is often a potential dependency between long-term and short-term preferences. Most existing works tend to mine the dependencies between long-term and short-term trajectories by contrastive learning but always ignore the negative impact of the learned dependencies on the accuracy of short-term trajectory modeling. Moreover, they often only utilize the context information of the user's trajectory, while neglecting the spatiotemporal dependencies between user trajectories. To address these issues, we proposed a novel dual contrastive learning framework DCLS. Specifically, we designed a novel dual contrastive learning scheme, for which we built two views: the first view is between the user's own long-term and short-term trajectories, and the second view is between the short-term trajectories of different users. We performed contrastive learning on both views, to learn the dependency between long-term and short-term trajectories, and improve the accuracy of trajectory modeling. We also designed a multi-class attention fusion module, which integrates the spatiotemporal influence of trajectory dependencies on user mobility, enhancing the recommendation performance. We conducted extensive experiments on three real-world datasets, which demonstrated that our model achieves advanced performance in the next POI recommendation.

Keywords: Next POI Recommendation; Contrastive Learning; Attention Mechanism

Introduction

Due to the development of transportation and modern networks, each user's daily travel choices have become more complex. In the face of complicated information, POI recommendation substantially improves the efficiency of users in travel decision-making and also provides a more scientific reference value for enterprises' location and advertising (Fu et al., 2024; Z. Liu et al., 2023).

The next POI recommendation differs from the POI recommendation in that it considers the user's real-time interests to make dynamic POI recommendations (Yang, Liu, & Zhao, 2022). Existing works found that users' interests consist of stable long-term preferences and short-term temporary interests, and considering these two aspects can improve the performance of the recommendation (Zheng et al., 2022; Duan, Fan, Zhou, Liu, & Wen, 2023). However, previous works often focused on the learning of either long-term or short-term trajectories (L. Huang, Ma, Wang, & Liu, 2019; Zhu et al., 2021). This could lead to the entanglement of long-term and short-term preferences, which may affect the recommendation accuracy. Recently, some work (Sun et al., 2020; J. Liu, Chen, Huang, Li, & Min, 2023) recognized the unique information in trajectories of different periods and attempted to use Long Short-Term Memory (LSTM), Recurrent Neural Networks (RNN), and other methods to learn this information. These works on the next POI recommendation have made some progress by modeling long and short-term trajectories. However, some aspects remain unaddressed.

Firstly, existing methods usually focus on learning users' long-term and short-term trajectories, while ignoring the potential dependencies between them. However, exploiting the dependencies between long-term and short-term trajectories can lead to higher-quality recommendations. Users' long-term preferences influence their short-term travel choices. Users' short-term preferences are also updating users' stable interests.

Secondly, a common method used by existing works to mine the dependencies between long-term and short-term trajectories is contrastive learning. When introducing trajectory dependencies to short-term trajectories, contrastive learning also brings noise to the trajectory modeling, thus affecting the accuracy of short-term trajectory modeling.

Finally, the same check-in segments in different trajectories reveal the spatiotemporal dependency among user trajectories (Yang et al., 2022). These shared trajectories indicate the similarity of individual mobility behavior, which can be used to predict the next POI. However, most existing models only fuse the user's own long-term and short-term trajectory information for recommendation, ignoring the spatiotemporal dependency among all trajectories. This leads to suboptimal recommendation performance.

In response to these challenges, this paper proposed a novel Dual Contrastive Learning framework with Long and Short-term trajectory modeling named **DCLS**. The goal of this framework is to capture the high-order dependency between long-term and short-term trajectories effectively while maintaining the accuracy of short-term trajectory modeling. For this purpose, we designed a novel dual contrastive learning scheme, for which we also designed two views. The first view is between the user's own long-term and short-term trajectories. The second view is between the short-term trajectories of different users. In the information fusion for the recommendation, we use the graph attention mechanism (GAT) to learn the spatiotemporal dependence hidden in all users' trajectories, thus providing a more comprehensive trajectory modeling. In summary, the primary contributions of our model are listed as follows:

- A dual contrastive learning framework (DCLS) for the next POI recommendation is proposed, which can mine the dependencies between long-term and short-term trajectories while considering the spatiotemporal dependencies of tra-

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

jectories.

- We designed a novel dual contrastive learning scheme and a novel information fusion strategy. The former can fully explore the dependencies between long-term and short-term trajectories while ensuring accurate trajectory modeling. The latter can adaptively fuse long-term preference, short-term preference, and spatiotemporal dependency.
- Extensive experiments conducted on three public benchmark datasets demonstrate that the performance of DCLS surpasses the state-of-the-art next POI recommendation methods.

Related Work

Long and Short-Term Trajectory Modeling

Despite the impressive development of work on the next POI recommendations, there is still a minority of work focused on studying long and short-term trajectory modeling for the next POI recommendations. Recently, some works have attempted to learn long and short-term trajectory preferences separately (Jiang, He, Cui, Xu, & Liu, 2023; Sun et al., 2020). Zhu et al. (Zhu et al., 2021) employed the Bidirectional Long Short-Term Memory (Bi-LSTM) neural networks with a package-level attention mechanism to study the long-term and short-term preferences of users. Liu et al. (J. Liu et al., 2023) introduced a Graph Neural Network (GNN) based model that transforms POIs into low-dimensional metrics, integrating both long and short-term preferences for a comprehensive representation of dynamic preferences. Additionally, Duan et al. (Duan et al., 2023) adopt a shared transformer encoder to encode both long and short-term behaviors, emphasizing behavioral commonalities over time. This innovative approach enables precise modeling and adaptive fusion of users' preferences, showcasing a significant advancement in POI recommendation methodologies. These studies exploited the relationship between long-term and short-term trajectories but ignored the protection of the accuracy of trajectory modeling.

Contrastive Sequential Recommendation

The self-supervised learning (SSL) paradigm, which has demonstrated remarkable success in learning representations from real data through pairs of positive and negative samples in diverse fields (Y. Zhang et al., 2022; B. Wu et al., 2023), recent works have explored its potential application in sequence recommendation (Zhou et al., 2020; Xie et al., 2022; Nizri, Azaria, & Hazon, 2023). Xie et al. (Xie et al., 2022) leveraged three random augmentation operators to enhance sequence augmentation techniques. Similarly, Zhou et al. (Zhou et al., 2020) employed item masking and item cropping while introducing four contrastive tasks to pre-train a bidirectional transformer for next-item prediction. Huang et al. (C. Huang, Wang, Wang, & Yao, 2023) innovatively devised a sequence attribute for next-item prediction, introducing a dual-transformer module and a dual contrastive learning scheme to learn users' low and high-level preferences discriminatively. Despite these works promoting the development of sequential recommendation, the potential of contrastive learning for the next POI recommendation has not been fully exploited.

Preliminaries And Problem Formulation

Let $P = \{p_1, p_2, \cdots, p_N\}$ be a set of POIs, $U = \{u_1, u_2, \cdots, u_M\}$ be a set of users, where N, M are the total number of POIs and users. $H = \{h_1, h_2, \cdots, h_{24}\}$ is used to indicate which part of the day it is (map a day to 24 slices), and $D = \{0, 1\}$ is used to indicate whether it is a weekend or not. Each POI $p \in P$ is denoted by a tuple $p = \langle lat, long, cat, num \rangle$ of latitude, longitude, POI categories, and visited-num, respectively.

Definition 1 (Trajectory) *Let $Q_i^u = \{q_1, q_2, \cdots, q_{|Q_i^u|}\}$ be a set of check-in recorders of user u on i-th day. Each check-in is denoted by $q = \langle u, p, h, w \rangle$, which indicates that user u visited POI p at time h and whether the day was a weekend.*

Definition 2 (POI-POI graph) *The POI-POI graph $\mathbb{G} = (P, A)$ shows the network of interactions and associations between POIs globally, where P denotes the set of POIs. $A \in R^{N \times N}$ denotes the adjacent matrix, and the value of $a_{i,j}$ in A represents the number of times p_i and p_j have been visited consecutively.*

Short-term trajectories react to the user's current movement state. When making recommendations, the sequence of check-in points before the recommendation is used as a short-term trajectory $T_s^u = Q_i^u$. And $T_l^u = \{Q_1^u, Q_2^u, \cdots, Q_{i-1}^u\}$ denotes the user u's long-term trajectories, consisting of check-in trajectories for several days (e.g.,7 days) before the recommendation task, where i is the total number of the user's trajectories. Specifically, for each user, given the long-term trajectories T_l^u, and the short-term trajectory T_s^u, we aim to give the user a set of sets of POIs S he wants to visit at the next timestamp.

Methodology

The framework of DCLS is shown in Figure 1, which is composed of three key components: (1) Trajectory Modeling Module is used to encode long and short-term trajectories separately (2) Dual Contrastive Learning Module is used to fully mine the dependencies of long-term and short-term trajectories and ensures the accuracy of short-term trajectory modeling, by conducting contrastive learning on two views (3) Multi-Class Attention Fusion Module is used to capture the importance of different time-span preferences and model the potential spatiotemporal dependencies between different trajectories, which will adaptively fuse long and short-term trajectories through one layer of attention and then fuses spatiotemporal dependencies through the graph attention mechanism.

Trajectory Modeling Module

Trajectory embedding It is widely known that the next POI recommendation is characterized by rich spatiotempo-

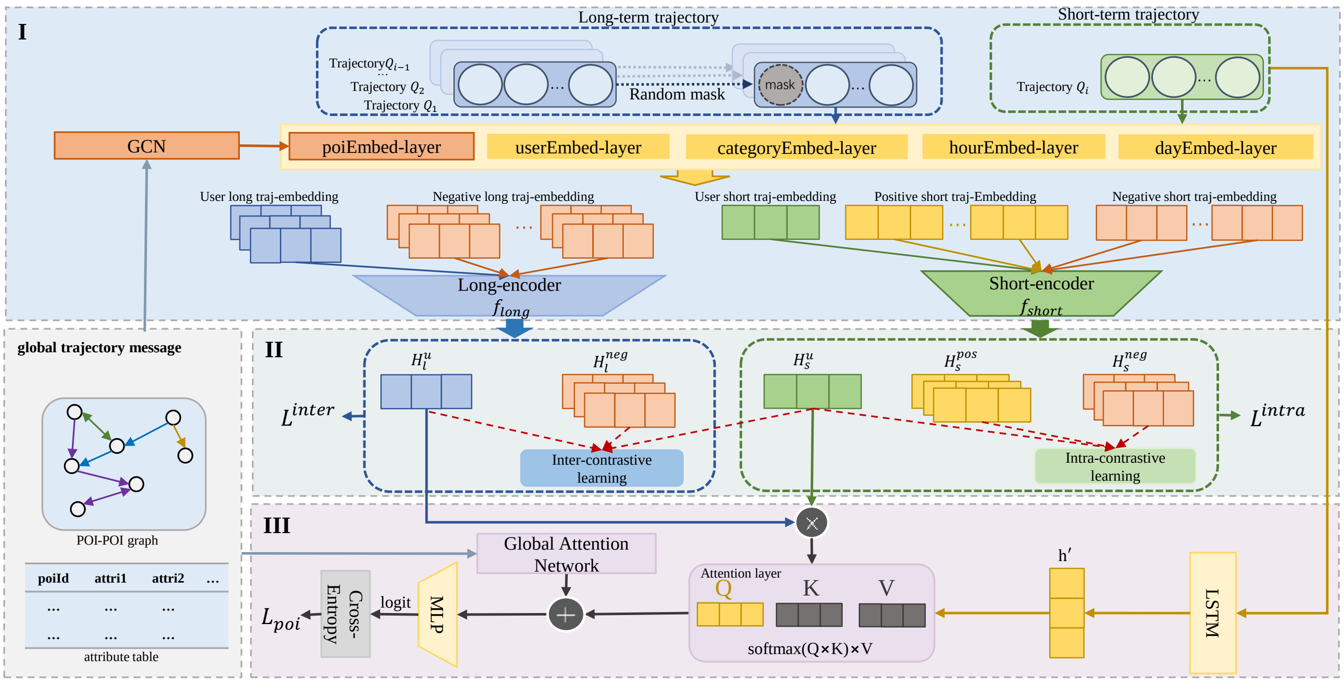


Figure 1: The overall framework of DCLS. This framework mainly consists of three modules: (I) Trajectory Modeling Module (II) Dual Contrastive Learning Module (III) Multi-Class Attention Fusion Module

ral contexts. Therefore, we take into account various factors when embedding the representation of trajectories, including categories, visit times, and weekend information. GCN (Kipf & Welling, 2016) is introduced to learn the dependencies relation of POIs. Let the output of the last GCN layer be the hidden representation $e_p \in R^{N \times d_p}$, which also is the learned global embedding matrix, where d_p refers to the dimension of the POI embedding and the i-th column of e_p is e_{p_i} denotes the embedding of POI p_i.

Then, the most basic embedding layer is used here for learning the embeddings for the remaining contexts. Finally, for the user's trajectory $Q_i^u = \{q_1, q_2, \cdots, q_{|Q_i^u|}\}$, each check-in point can be expressed as:

$$e_{q_i} = e_{p_i} \oplus e_{u_i} \oplus e_{c_i} \oplus e_{h_i} \oplus e_{w_i} \tag{1}$$

where $\oplus$ is the concatenation operation. And $e_u \in R^{d_u}$ is the embedding of the user, $e_c \in R^{d_c}$ is the embedding of the category, $e_h \in R^{d_h}$ is the embedding of the access time, and $e_w \in R^{d_w}$ is the embedding of the weekend indicator. The sum of all embedding dimensions is denoted as D for convenience. So the embedding of the trajectory Q_i^u is denoted as $E_{Q_i^u} = [e_{q_1}, e_{q_2}, \cdots, e_{q_{|Q_i^u|}}] \in R^{|Q_i^u| \times D}$.

Long-term trajectory modeling For the user's long-term trajectories T_l^u, each trajectory in the user's long-term sequence will be randomly masked according to 10% probability to get T_l^{u*}. Then the T_l^{u*} is fed into the trajectory embedding module to obtain the embedding of each trajectory. Finally concatenate the representations of each trajectory in the long-term trajectories to obtain the user's long-term trajectories embedding representation $E_{T_l^u} \in R^{X \times D}$, where X denotes the total number of check-in points in the long-term trajectories.

To model both long and short-term trajectories well, a bi-directional transformer encoder is introduced as long and short-term trajectory encoders to learn the long and short-term trajectories hidden representation respectively. The user's long-term trajectories representation $E_{T_l^u}$ is fed into the long-term trajectory encoder:

$$H_{T_l^u} = longEncoder(E_{T_l^u}) \tag{2}$$

where $H_{T_l^u} \in R^{X \times D}$ is the hidden representation of the long-term trajectories. In this equation, the long-term trajectory encoder is specified as follows:

$$H_{T_l^u} = LayerNorm(H_{T_l^u} + Dropout(FFN(U_i))) \tag{3}$$

$$U_i = MultiHead(Attention(E_{T_l^u} W_q, E_{T_l^u} W_k, E_{T_l^u} W_v)) \tag{4}$$

where $W_q, W_k, W_v \in R^{D \times \frac{D}{h}}$ are the learnable transformation weight matrices of query, key, and value, respectively. h denotes the number of heads.

Short-term trajectory modeling Firstly, the user's short-term trajectory is fed into the trajectory embedding module to get the embedding $E_{T_s^u} \in R^{Y \times D}$, where Y denotes the total number of check-in points in the short-term trajectory. Then, $E_{T_s^u}$ is fed into the short-term trajectory encoder:

$$H_{T_s^u} = shortEncoder(E_{T_s^u}) \tag{5}$$

$H_{T_s^u} \in R^{Y\times D}$ is the hidden representation that reflects the short-term preference of the user.

Dual Contrastive Learning

Inter-contrastive learning Typically, long-term trajectories represent the user's overall travel pattern, while a short-term trajectory represents the user's specific travel plan. For instance, users who prefer natural scenery are more likely to choose a park, while users who like urban life are more likely to go to a mall. However, the user's short-term travel choices may also reflect a change in their long-term preference. For example, a user who usually likes urban areas may start to visit more natural places, indicating a shift in their preference. Therefore, there is a potentially important relationship between the user's long-term and short-term preferences. To explore this potentially important relationship, contrastive learning is adopted to mine the information from the first view we designed. By comparing the preferences of different periods, the user's short-term choices can be optimized according to their long-term preferences, and long-term preferences can be updated based on their short-term choices. T_s^u is the embedding of the short-term trajectories of user u. Then, the long-term trajectories T_l^+ of the same user are positive samples, and the long-term trajectories T_l^- of different users are negative samples.

$$H_s^u = mean(shortEncoder(E_{T_s^u})) = mean(H_{T_s^u}) \tag{6}$$

$$H_l^u = mean(longEncoder(E_{T_l^u})) = mean(H_{T_l^u}) \tag{7}$$

$$H_l^- = mean(longEncoder(E_{T_l^-})) = \frac{1}{m}\sum_{i=0}^{m} H_i^{l-} \tag{8}$$

where m is the number of negative samples. $E_{T_s^u}, E_{T_l^u}, E_{T_l^-}$ are the embeddings of the user's short-term trajectory, long-term trajectories, and negative sample trajectories, respectively. Finally, the preferences of H_s^u, H_l^u, and H_l^- will be compared using a simplified InfoNCE loss.

$$\begin{aligned} L^{inter} = &-log(\sigma(\mu < H_s^u, H_l^u >))- \\ &log(1-\sigma(\mu < H_s^u, H_l^- >< H_l^u, H_l^- >)) \end{aligned} \tag{9}$$

where σ denotes the sigmoid function, and $<,>$ denotes the inner product of two embeddings regulated by temperature t. The μ is a parameter that controls the magnitude of the gradient of the loss function.

Intra-contrastive learning The second view is between the short-term trajectories of different users. Given that short trajectories with the same destination tend to have similar preferences, the trajectories T_s^- whose destination is different from the target user's destination are negative samples. The trajectories T_s^+ whose destination is consistent with the target user's destination are positive samples. Similarly, the intra-contrastive loss is calculated as follows:

$$H_s^+ = mean(shortEncoder(E_{T_s^+})) = \frac{1}{n}\sum_{i=0}^{n} H_i^{s+} \tag{10}$$

$$H_s^- = mean(shortEncoder(E_{T_s^-})) = \frac{1}{m}\sum_{i=0}^{m} H_i^{s-} \tag{11}$$

$$\begin{aligned} L^{intra} = &-log(\sigma(\mu < H_s^u, H_s^+ >))- \\ &log(1-\sigma(\mu < H_s^u, H_s^- >< H_s^+, H_s^- >)) \end{aligned} \tag{12}$$

where $E_{T_s^+}, E_{T_s^-}$ are the embeddings of the positive sample short trajectories and the negative sample short trajectories, respectively. n, m are the number of positive and negative samples, respectively.

Multi-Class Attention Fusion

Long and short-term preferences fusion The importance of long-term and short-term preferences largely depends on the sequence order of the current user's visits (Zheng et al., 2022). If the user visits several new types of points consecutively, then the user is mainly influenced by short-term preferences at this time. If the user visits a place they have been to in the past, then the user is likely to be affected by long-term preferences. For this reason, a separate LSTM is deployed to model the recently visited sequences explicitly. The final output of the LSTM is h_q. In the long-term and short-term fusion process, the hidden representation h_q is used as the query vectors for the attention mechanism. Then the user's long and short-term preferences are concatenated into h_k and h_v as the key vector and the value vector, respectively.

$$h' = \sum_{i=1}^{T} a_{q,k_i} h_{v_i}, h_{v_i} \in h_v \tag{13}$$

$$\alpha_{q,k_i} = \frac{exp(h_q^T h_{k_i})}{\sum_{i=1}^{T} exp(h_q^T h_{k_i'})} \tag{14}$$

where T represents the total number of long-term and short-term trajectories; α_{q,k_i} denotes the attention score for each latent representation h_{k_i} based on the user's present situation. Then the learned user preference h' passed through a multi-layer perception (MLP) to get the probability distribution $y' \in R^{1\times N}$ over the N POIs.

Spatiotemporal dependencies fusion A simplified graph attention mechanism is employed to learn potential dependencies between POIs from graph $\mathbb{G}$ (Yang et al., 2022). This mechanism can capture the spatiotemporal dependencies between global trajectory points. The node feature matrix $Z \in R^{N\times d_f}$ is first linearly transformed to get $W_z \in R^{N\times d_p}$ and then matrix-multiplied with the two parts of the attention weight tensor to get W_1 and $W_2 \in R^{N\times N}$, respectively.

$$W_1 = W_z a_1, W_2 = W_z a_2 \tag{15}$$

$$\hat{A} = A + i \tag{16}$$

where a_1 and a_2 are two learnable vectors in the attention matrix, respectively. $A \in R^{N\times N}$ is the adjacent matrix, $i \in R^{N\times N}$ is an all-ones matrix used to ensure that zeros in A do not affect the calculation. Next, W_1 and the transpositions of W_2 are summed and a LeakyRelu activation function is applied

Table 1: Performance comparison in HR@K and NDCG@K on three datasets.

	PHO				NYC				SIN			
	HR@5	HR@10	NDCG@5	NDCG@10	HR@5	HR@10	NDCG@5	NDCG@10	HR@5	HR@10	NDCG@5	NDCG@10
ATST-LSTM	0.1579	0.2377	0.1033	0.1385	0.1667	0.2031	0.0912	0.1638	0.1296	0.1933	0.1027	0.1476
PLSPL	0.1775	0.2569	0.1285	0.1538	0.1741	0.2413	0.0961	0.1825	0.1447	0.1719	0.1126	0.1384
iMTL	0.1830	0.2747	0.1301	0.1632	0.1798	0.2422	0.0989	0.1861	0.1505	0.1801	0.1051	0.1423
CFPRec	0.3421	0.4253	0.2432	0.2730	0.2771	0.3606	0.1971	0.2190	0.2310	0.3085	0.1588	0.1836
ContraRec	0.3381	0.3680	0.2843	0.2939	0.1951	0.2368	0.1425	0.1560	0.2047	0.2710	0.1454	0.1660
DisenPOI	0.4209	0.4988	0.3402	0.3652	0.2940	0.3634	0.2401	0.2594	0.2808	0.3420	0.2307	0.2506
CLSPRec	0.5368	0.6368	0.3811	0.4175	0.3545	0.4352	0.2653	0.2871	0.3544	0.4093	0.2794	0.2942
DCLS	**0.6842**	**0.7017**	**0.5255**	**0.5472**	**0.4357**	**0.4980**	**0.3385**	**0.3599**	**0.4382**	**0.4894**	**0.3592**	**0.3747**
Improvement(%)	27.4	10.1	37.8	31.0	22.9	14.4	27.5	25.3	23.6	19.5	28.5	27.3

Table 2: Ablation study on the PHO.

	HR@5	HR@10	NDCG@5	NDCG@10
Full Model	**0.6842**	**0.7017**	**0.5255**	**0.5472**
w/o G	0.6579	0.6930	0.5178	0.5314
w/o S	0.6754	0.7281	0.5230	0.5356
w/o att1	0.5439	0.6579	0.3998	0.4305
w/o att2	0.6404	0.6930	0.4959	0.5203

to get the alignment score. Then, the alignment score and $\hat{A}$ are multiplied to obtain the final attention matrix $w' \in R^{N\times N}$.

$$w' = relu(W_1 + W_2^T)\hat{A} \tag{17}$$

Finally, the POI recommendation probability distribution $\bar{y} \in R^{1\times N}$ under global attention weight w' fine-tuning can be expressed as :

$$\bar{y} = y' + w' \tag{18}$$

Training and Optimization

The cross-entropy is adopted as the loss function to calculate the POI prediction loss during the model training process.

$$L_{poi} = -\sum_{i=1}^{N} log(\bar{y}_i) \tag{19}$$

where $\bar{y}_i$ denotes the probability that POI p_i will be visited at the next time, and N is the number of POIs. Meanwhile, two contrastive learning tasks are introduced to optimize the model. The final loss function is expressed as follows:

$$L_{all} = L_{poi} + \alpha L^{inter} + \beta L^{intra} \tag{20}$$

where α, β are loss coefficients and they control the strengths of the two contrastive learning tasks respectively.

Experiments

Datasets and Experimental Settings

We conducted experiments on three real-world public datasets collected from Foursquare: Singapore (SIN), New York City (NYC), and Phoenix (PHO). These check-in datasets spanned about 18 months (April 2012 to September 2013). To ensure the reliability of our data, POIs with less than 10 interactions and trajectories with less than 3 check-ins are removed, and inactive users with less than 5 check-ins are filtered out.

The key hyperparameters in our model are set as follows. The embedding dimensions of POI and user are both $d_p, d_u = 64$. The embedding length of the POI category, the access time, and the weekend indicator are $d_c, d_h, d_w = 32$. The loss weights of the two contrastive learning tasks are set to $\alpha, \beta = 1$. Moreover, we employ the Adam optimizer with a learning rate in $\{1e-4, 1e-5\}$.

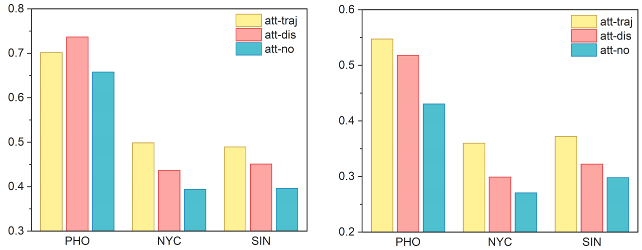
Baselines and Evaluation Metrics

To demonstrate the validity of the proposed model, we compared the performance of DCLS with the following recent models: an Attention-based model: ATST-LSTM (L. Huang et al., 2019); models considering the spatio-temporal context: PLSPL (Y. Wu, Li, Zhao, & Qian, 2020), iMTL (L. Zhang et al., 2021), DisenPOI (Qin et al., 2023); models based on contrastive learning: CFPRec (L. Zhang et al., 2022), ContraRec (Wang et al., 2023), CLSPRec (Duan et al., 2023).

In evaluating model performance, we adopt two widely used metrics: Hit Rate at K (HR@K) and Normalized Discounted Cumulative Gain at K (NDCG@K). HR@K mainly measures the accuracy of the recommendation results, that is, whether the ground truth item appears in the top-K list. On the other hand, NDCG@K places more emphasis on the ranking quality of the recommended sequences, taking into account the position of the ground truth item in the list. In general, the higher the values of these two metrics, the better the performance in the next POI recommendation.

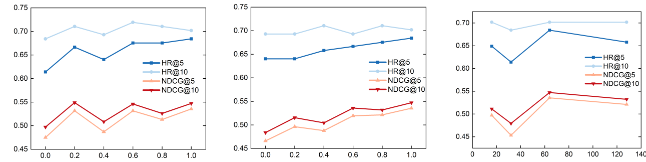
Results and Analysis

Overall performance comparison In Table 1, we show the performance of our proposed model and several baseline models. The results show that DCLS outperforms other baseline models on all datasets, which validates the advantages of our model. This is because (a) by designing two views of contrastive learning objectives, the dependencies between short and long trajectories are fully exploited and the accuracy of the trajectory modeling is guaranteed (b) spatiotemporal dependencies between trajectories are considered and adaptively integrated with users' long and short-term prefer-



(a) HR@10 (b) NDCG@10

Figure 2: Comparison of different learning methods to learn about spatiotemporal dependence of trajectories. (i) att-adj denotes learning on $\mathbb{G}$, (ii) att-dis denotes learning on $\mathbb{G}'$, (iii) att-no denotes not learning spatiotemporal dependencies



(a) Parameter of α (b) Parameter of β (c) Parameter of d_p

Figure 3: Parameter sensitivity analysis of DCLS

ences, resulting in a more comprehensive preference modeling.

Ablation study Due to space limitations, we only show the performance comparison on the PHO dataset, with similar results for the NYC and SIN datasets. To validate the effectiveness of the different modules in DCLS, we compared the performance of the full model with its four variants. 1) w/o G removed GCN and replaced it with a simple embedding layer; 2) w/o S removed the dual contrastive learning scheme; 3) w/o att1 removed the spatiotemporal dependence fusion; 4) w/o att2 disregarded the effect of the user's current state on the importance of long and short-term preferences. These experimental results are shown in Table 2. We can see that the performance of DCLS is generally higher than the rest of the variants, so we can say that each module in DCLS contributes to the performance improvement. First, w/o att1 has the lowest performance, indicating that incorporating spatio-temporal dependence is helpful for inferring user preferences. Second, w/o S is lower than the full model in most cases, suggesting that the dual-contrast scheme improves most of the performance of recommendations. DCLS consistently beats w/o G, which verifies that the GCN-based learning of POI embeddings is effective. w/o att2 also shows that the long and short-term trajectory fusion strategy that we designed enhances the overall performance.

Analysis of spatiotemporal dependency fusion We further analyze the learning of spatiotemporal dependencies. We construct two graphs: a POI-POI graph $\mathbb{G}$ and a geographic adjacency graph $\mathbb{G}'$. $\mathbb{G}'$ is an undirected graph, where the distance between two POIs is calculated by the haversine formula. If the distance is less than a predefined threshold Δd, they are connected; otherwise, they are not. Then we conduct experiments on these two graphs and also consider the case without spatio-temporal dependency learning. The experimental results are shown in Figure 2. From this figure, we can draw the following conclusions: First, learning the spatiotemporal dependencies of trajectories is very helpful for recommendation performance. Second, the spatiotemporal dependency information learned from the check-in sequences of all users is better than that learned from the geographic distance in most cases. It can also be said that the real-world adjacency relationship between POIs can be learned from the user's check-in sequences. This is understandable, as users tend to visit POIs that are geographically close when traveling.

Sensitivity and parametric analyses A parameter sensitivity analysis is also conducted to evaluate the impact of different hyperparameters on DCLS, including the contrastive loss weights α and β for the two views and the POI embedding size d_p. Due to space limitations, we only report the experimental results on PHO, as shown in Figure 3. When α increases from 0 to 1 (β is fixed at 1), we can see that all the performance metrics of DCLS show a zigzag upward trend. When β increases from 0 to 1 (α is fixed at 0), the model performance has a similar growth trend. From this, we can see that the dual contrastive tasks designed have both contributed to the improvement of the model performance. We also found that when the POI embedding size d_p ranges from 16 to 128, the performance of DCLS first increases and then gradually decreases. This is because the small embedding dimension lacks sufficient capacity to represent all relevant information. Consequently, the model fails to capture essential features, resulting in suboptimal performance. Conversely, an excessively large embedding dimension introduces redundancy in the model's parameters. This redundancy can hinder training efficiency and lead to overfitting, ultimately degrading the model's generalization ability.

CONCLUSION

In this paper, we proposed DCLS, a novel dual-contrast learning framework that can fully utilize long and short-term trajectory information to accurately recommend the next POI for users. A dual-contrast scheme has been designed to fully exploit the dependency relationship between long-term and short-term trajectories while ensuring the accuracy of trajectory modeling. Meanwhile, a multi-class attention fusion mechanism is developed to fuse long-term and short-term trajectories and spatiotemporal dependency adaptively. We conducted extensive experiments on three real-world datasets which demonstrate the effectiveness and superiority of our model over the state-of-the-art baselines.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 62072409 and Grant 62073295, in part by the Zhejiang Provincial Natural Science Foundation under Grant LR21F020003, and in part by the Zhejiang Province Basic Public Welfare Research Project under Grant LTGG24F020009.

References

Duan, C., Fan, W., Zhou, W., Liu, H., & Wen, J. (2023). Clsprec: Contrastive learning of long and short-term preferences for next poi recommendation. In *Proceedings of the 32nd acm international conference on information and knowledge management* (pp. 473–482).

Fu, J., Gao, R., Yu, Y., Wu, J., Li, J., Liu, D., & Ye, Z. (2024). Contrastive graph learning long and short-term interests for poi recommendation. *Expert Systems with Applications*, *238*, 121931.

Huang, C., Wang, S., Wang, X., & Yao, L. (2023). Dual contrastive transformer for hierarchical preference modeling in sequential recommendation. In *Proceedings of the 46th international acm sigir conference on research and development in information retrieval* (pp. 99–109).

Huang, L., Ma, Y., Wang, S., & Liu, Y. (2019). An attention-based spatiotemporal lstm network for next poi recommendation. *IEEE Transactions on Services Computing*, *14*(6), 1585–1597.

Jiang, S., He, W., Cui, L., Xu, Y., & Liu, L. (2023). Modeling long-and short-term user preferences via self-supervised learning for next poi recommendation. *ACM Transactions on Knowledge Discovery from Data*.

Kipf, T. N., & Welling, M. (2016). Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*.

Liu, J., Chen, Y., Huang, X., Li, J., & Min, G. (2023). Gnn-based long and short term preference modeling for next-location prediction. *Information Sciences*, *629*, 1–14.

Liu, Z., Zhang, D., Zhang, C., Bian, J., Deng, J., Shen, G., & Kong, X. (2023). Kdrank: Knowledge-driven user-aware poi recommendation. *Knowledge-Based Systems*, *278*, 110884.

Nizri, M., Azaria, A., & Hazon, N. (2023). Contrastive explanations for recommendation systems. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 45).

Qin, Y., Wang, Y., Sun, F., Ju, W., Hou, X., Wang, Z., ... Zhang, M. (2023). Disenpoi: Disentangling sequential and geographical influence for point-of-interest recommendation. In *Proceedings of the sixteenth acm international conference on web search and data mining* (pp. 508–516).

Sun, K., Qian, T., Chen, T., Liang, Y., Nguyen, Q. V. H., & Yin, H. (2020). Where to go next: Modeling long-and short-term user preferences for point-of-interest recommendation. In *Proceedings of the aaai conference on artificial intelligence* (Vol. 34, pp. 214–221).

Wang, C., Ma, W., Chen, C., Zhang, M., Liu, Y., & Ma, S. (2023). Sequential recommendation with multiple contrast signals. *ACM Transactions on Information Systems*, *41*(1), 1–27.

Wu, B., Liang, X., Zheng, X., Guo, Y., Zhang, X., Wang, J., et al. (2023). Mvrace: Multi-view graph contrastive encoding for graph neural network pre-training. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 45).

Wu, Y., Li, K., Zhao, G., & Qian, X. (2020). Personalized long-and short-term preference learning for next poi recommendation. *IEEE Transactions on Knowledge and Data Engineering*, *34*(4), 1944–1957.

Xie, X., Sun, F., Liu, Z., Wu, S., Gao, J., Zhang, J., ... Cui, B. (2022). Contrastive learning for sequential recommendation. In *2022 ieee 38th international conference on data engineering (icde)* (pp. 1259–1273).

Yang, S., Liu, J., & Zhao, K. (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).

Zhang, L., Sun, Z., Wu, Z., Zhang, J., Ong, Y. S., & Qu, X. (2022). Next point-of-interest recommendation with inferring multi-step future preferences. In *Proceedings of the 31st international joint conference on artificial intelligence (ijcai)* (pp. 3751–3757).

Zhang, L., Sun, Z., Zhang, J., Lei, Y., Li, C., Wu, Z., ... Klanner, F. (2021). An interactive multi-task learning framework for next poi recommendation with uncertain check-ins. In *Proceedings of the twenty-ninth international conference on international joint conferences on artificial intelligence* (pp. 3551–3557).

Zhang, Y., Shen, Z., Wu, C.-H., Xie, B., Hao, J., Wang, Y.-Y., ... Han, J. (2022). Metadata-induced contrastive learning for zero-shot multi-label text classification. In *Proceedings of the acm web conference 2022* (pp. 3162–3173).

Zheng, Y., Gao, C., Chang, J., Niu, Y., Song, Y., Jin, D., & Li, Y. (2022). Disentangling long and short-term interests for recommendation. In *Proceedings of the acm web conference 2022* (pp. 2256–2267).

Zhou, K., Wang, H., Zhao, W. X., Zhu, Y., Wang, S., Zhang, F., ... Wen, J.-R. (2020). S3-rec: Self-supervised learning for sequential recommendation with mutual information maximization. In *Proceedings of the 29th acm international conference on information & knowledge management* (pp. 1893–1902).

Zhu, G., Wang, Y., Cao, J., Bu, Z., Yang, S., Liang, W., & Liu, J. (2021). Neural attentive travel package recommendation via exploiting long-term and short-term behaviors. *Knowledge-Based Systems*, *211*, 106511.