Disentangled Contrastive Hypergraph Learning for Next POI Recommendation

Yantong Lai*
Institute of Information Engineering, Chinese Academy of Sciences
School of Cyber Security, University of Chinese Academy of Sciences
Beijing, China
laiyantong@iie.ac.cn

Yijun Su†
Institute of Information Engineering, Chinese Academy of Sciences
Beijing, China
suyijun.ucas@gmail.com

Lingwei Wei
Institute of Information Engineering, Chinese Academy of Sciences
Beijing, China
weilingwei@iie.ac.cn

Tianqi He
Meituan.com
Beijing, China
hetianqi@meituan.com

Haitao Wang†
Meituan.com
Beijing, China
wanghaitao13@meituan.com

Gaode Chen
Institute of Information Engineering, Chinese Academy of Sciences
School of Cyber Security, University of Chinese Academy of Sciences
Beijing, China
chengaode@iie.ac.cn

Daren Zha
Institute of Information Engineering, Chinese Academy of Sciences
Beijing, China
zhadaren@iie.ac.cn

Qiang Liu
Meituan.com
Beijing, China
liuqiang43@meituan.com

Xingxing Wang
Meituan.com
Beijing, China
wangxingxing04@meituan.com

ABSTRACT

Next point-of-interest (POI) recommendation has been a prominent and trending task to provide next suitable POI suggestions for users. Most existing sequential-based and graph neural network-based methods have explored various approaches to modeling user visiting behaviors and have achieved considerable performances. However, two key issues have received less attention: i) Most previous studies have ignored the fact that user preferences are diverse and constantly changing in terms of various aspects, leading to entangled and suboptimal user representations. ii) Many existing methods have inadequately modeled the crucial cooperative associations between different aspects, hindering the ability to capture complementary recommendation effects during the learning process. To tackle these challenges, we propose a novel framework Disentangled Contrastive Hypergraph Learning (DCHL) for next POI recommendation. Specifically, we design a multi-view disentangled hypergraph learning component to disentangle intrinsic aspects among collaborative, transitional and geographical views with adjusted hypergraph convolutional networks. Additionally, we propose an adaptive fusion method to integrate multi-view information automatically. Finally, cross-view contrastive learning is employed to capture cooperative associations among views and reinforce the quality of user and POI representations based on self-discrimination. Extensive experiments on three real-world datasets validate the superiority of our proposal over various state-of-the-arts. To facilitate future research, our code is available at https://github.com/icmpnorequest/SIGIR2024_DCHL.

CCS CONCEPTS

• **Information systems → Recommender systems**.

KEYWORDS

Next POI Recommendation, Disentangled Representation, Hypergraph Neural Networks, Contrastive Learning

ACM Reference Format:
Yantong Lai, Yijun Su, Lingwei Wei, Tianqi He, Haitao Wang, Gaode Chen, Daren Zha, Qiang Liu, and Xingxing Wang. 2024. Disentangled Contrastive Hypergraph Learning for Next POI Recommendation. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '24), July 14–18, 2024, Washington, DC, USA.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3626772.3657726

1 INTRODUCTION

Due to the prevalence of location-based social networks [5], people are increasingly willing to record and share their daily life and experience along with geographical information on location-based

*Work done as a research intern at Meituan. This work is supported by the Postdoctoral Fellowship Program of China Postdoctoral Science Foundation (No. GZC20232969).

†Corresponding author.

SIGIR '24, July 14–18, 2024, Washington, DC, USA

ACM ISBN 979-8-4007-0431-4/24/07
https://doi.org/10.1145/3626772.3657726

social networking applications (e.g., Facebook, Instagram, Yelp and Dianping). Consequently, personalized recommender systems have been widely utilized to help users discover point of interests (POIs) from such large amount information. In POI recommender systems, next POI recommendation is one of significant and fundamental tasks. Formally, next POI recommendation aims to provide suitable predictions for users in the next movement according to their historical trajectories [17, 18, 31, 34, 36, 37].

Generally, most existing methods treat next POI recommendation as a sequential prediction task and adopt sequential methods to model transition patterns, ranging from Markov Chains [4, 33] to recurrent neural networks (RNN) [9, 24, 35, 51] and recent self-attention mechanism [22, 25]. Among these sequential-based methods, some approaches [22, 24, 25, 35, 51] validate the importance of spatio-temporal contexts (e.g., spatial or temporal intervals and spatio-temporal gates), and incorporate such information in mining the regularity of users' trajectories. However, these sequential-based methods mainly focus on each user's intra-sequence exploiting but fail to explore collaborative information from other users. Inspired by the great success of graph neural networks (GNN) in capturing similarities between high-order neighbors and modeling complex relationships, some researchers [8, 15, 17–19, 21, 23, 31, 37, 45] leverage GNN-based or hypergraph neural network (HGNN) based methods to enrich POI and user representations. For example, Graph-Flashback [31] utilizes a spatial-temporal knowledge graph to endow POI representations and incorporates them into RNN-based methods to capture sequential transition patterns for next POI recommendation. Inspired by the flexible structure of hypergraph to represent high-order neighbors and to capture high-order collaborative signals, Lai et al. [17] proposed a multi-view spatial-temporal enhanced hypergraph network to capture spatial-temporal information and distill high-order collaborative signals simultaneously. The proposed model MSTHN alleviates the data sparsity issue and over-smoothing effects in GNN-based methods for next POI recommendation.

While the aforementioned approaches have achieved state-of-the-art performances for next POI recommendation, two key issues remain less explored.

First, most previous studies have ignored the fact that user preferences are diverse and constantly changing in terms of various aspects, leading to suboptimal and entangled user representations. In next POI recommendation, the rationale behind user-POI interactions is driven by several potential factors (e.g., spatial, temporal and categorical) [30], such as buying a cup of coffee near the office or being attracted by a far away restaurant for its specific cuisine. However, the learned user preferences in current graph or hypergraph based approaches are entangled. They only considered coarse-grained user-POI interactions, but ignored various potential aspects behind their behaviors, hindering the ability to capture diverse and accurate user preferences. Therefore, it is challenging to disentangle and model multi-aspect user representations that driven their behaviours.

Second, many existing methods have inadequately modeled the crucial cooperative associations between different aspects, hindering the ability to capture complementary recommendation effects during the learning process. Specifically, the complementary effect means that it can combine information from multiple views to gain a more comprehensive understanding of the underlying data and enhance recommendation performances. In next POI recommendation, some researchers [17, 20, 28, 29] have adopted multi-view learning or disentangled learning paradigm to learn view-specific or aspect-specific representations separately, and simply fuse them for next POI prediction. However, such methods fail to differentiate the similarities among views or aspects. Additionally, existing works [12, 42] only consider the cooperative associations at the level of prediction, resulting in no mutual enhancement among the views is guaranteed to be captured. Therefore, it is an urgent need to properly model the crucial cooperative associations and encourage the mutual enhancement among views.

In this paper, we propose a novel model Disentangled Contrastive Hypergraph Learning (DCHL) for next POI recommendation, to address above challenges. For addressing the first limitation, our model disentangles multi-view representations by considering complex collaborative, global transitional and geographical relationships between users and POIs, which have been proved important, effective and interpretable for next POI recommendation [17, 18, 28–30]. Inspired by the highly flexible ability to represent high-order neighbors, we innovatively design three distinct hypergraphs-collaborative hypergraph, transitional hypergraph, and geographical hypergraph, to represent nodes in global dependencies from different views. We then encode POIs into disentangled embeddings for decoupling intrinsic aspects among collaborative, transitional and geographical views by proposing adjusted method on aggregation and propagation of hypergraph convolutional network, respectively. After disentanglement, we can obtain decoupled user preferences on collaborative, transitional and geographical aspects according to specific hypergraph structure. To integrate and balance multi-view information, an adaptive fusion method is further utilized to adaptively synthesize the final user representation, leading to more interpretable and personalized recommendations. For addressing the second limitation, our model captures crucial cooperative associations among different views by designing a cross-view contrastive objective to consider complementary recommendation effects through self-augmentation. Experimental results on three real-world datasets have demonstrated the effectiveness of our DCHL for next POI recommendation.

In summary, the main contributions of our work are as follows:

- We explore two challenging yet practical problems in next POI recommendation and propose a novel framework disentangled contrastive hypergraph learning (DCHL) to solve them and enhance recommendation performances.
- We innovatively design three distinct hypergraphs among collaborative, transitional and geographical views and propose adjusted method on aggregation and propagation of hypergraph convolutional networks, to solve the limitation of entangled and suboptimal user representations.
- We adapt cross-view contrastive learning to cooperatively supervise each other between views via self-augmentation, to solve the limitation of hard to capture complementary recommendation effects during the learning process.
- Extensive experiments on three real-world datasets validate the effectiveness of our proposed DCHL over various state-of-the-art methods for next POI recommendation.

2 PRELIMINARY

In this section, we first formulate the task of next POI recommendation, and then introduce the definition of hypergraph.

2.1 Task Formulation

Let $\mathcal{U} = \{u_1, u_2, ..., u_{|\mathcal{U}|}\}$ and $\mathcal{L} = \{l_1, l_2, ..., l_{|\mathcal{L}|}\}$ be a set of users and POIs, respectively. Each POI $l \in \mathcal{L}$ has a unique geographical coordinates (*longitude, latitude*) tuple, i.e., (*lon, lat*). For each user $u \in \mathcal{U}$, we obtain her/his trajectory $s_u = \{(l_{u,i}, t_{l_{u,i}})|i = 1, 2, ...\}$, where each tuple $(l_{u,i}, t_{l_{u,i}})$ indicates user u visited POI $l_{u,i}$ at timestamp $t_{l_{u,i}}$.

Given a target user u and her/his trajectory sequence s_u, the goal of next POI recommendation is to recommend top-K POIs that u may visit in the next timestamp.

2.2 Hypergraph

Hypergraph [1, 2, 10, 11] is a generalization of graph, where an edge connects two or more vertices. Formally, a hypergraph can be represented by $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$, where $\mathcal{V}$ denotes the vertices set and $\mathcal{E}$ represents the hyperedges set. Incidence matrix $\mathbf{H} \in \mathbb{R}^{|\mathcal{V}|\times|\mathcal{E}|}$ is introduced to describe the topology structure of hypergraph. When node $v \in \mathcal{V}$ is in hyperedge $e \in \mathcal{E}$, $\mathbf{H}_{v,e} = 1$, otherwise 0.

3 METHODOLOGY

In this section, we present our proposed framework DCHL in detail. As illustrated in Figure 1, we first elaborate on the construction of multi-view disentangled hypergraphs from collaborative, transitional, and geographical views based on users' check-ins. Then, we perform disentangled hypergraph learning by designing hypergraph convolutional networks with adjusted aggregation and propagation method to decouple intrinsic aspects. After that, we learn and fuse multi-view user preferences according to hypergraph structure and a proposed adaptive fusion method. Additionally, we utilize cross-view contrastive learning to capture complementary recommendation effects among different views. Finally, we demonstrate our prediction and optimization approach.

3.1 Multi-View Disentangled Hypergraph Learning

3.1.1 Multi-View Disentangled Hypergraph Construction. In next POI recommendation, there exist some complex relationships between users and POIs, such as user-POI interactions, POI-POI transitional relationships and POI-POI geographical relationships. To represent such relationships, previous methods [28, 29, 38] leverage graph, where users and POIs can be regarded as nodes and the relationships between them are edges. However, conventional graph structure is limited to pairwise relationship, which could not connect higher-order neighbors within the same specific semantics. Motivated by the highly flexible structure of hypergraph, we innovatively design three distinct hypergraphs (Figure 1) as follows:

Collaborative View Hypergraph. The construction of collaborative view hypergraph involves building a hypergraph that captures high-order collaborative signals between users and POIs. Formally, we construct the collaborative view hypergraph $\mathcal{G}_C = (\mathcal{V}_C, \mathcal{E}_C)$, where $\mathcal{V}_C$ denotes the POIs set. In the hypergraph $\mathcal{G}_C$, we represent each user's trajectory s_u as a hyperedge, so the hyperedge set $\mathcal{E}_C$ consists of all users' trajectories. Additionally, incidence matrix $\mathbf{H}_C \in \mathbb{R}^{|\mathcal{L}|\times|\mathcal{U}|}$ is introduced to describe user-POI interactions. This collaborative view hypergraph provides valuable insights into both intra-sequence and inter-sequence relationships. By leveraging the hypergraph, our model can effectively discover similar users who have similar visiting patterns.

Transitional View Hypergraph. Since hyperedge in normal hypergraph structure is undirected, it is not fit for representing directed relationships (e.g., POI-POI transitional relationship). Therefore, we propose to model such transitional relationship with a directed hypergraph. Formally, we design a transitional view hypergraph $\mathcal{G}_T = (\mathcal{V}_T, \mathcal{E}_T)$, where nodes are POIs and hyperedges consist of directed transitional relationships between POIs in all trajectories. Incidence matrix $\mathbf{H}_T \in \mathbb{R}^{|\mathcal{L}|\times|\mathcal{E}_T|}$ denotes directed POI-POI transitional relationship, where POIs in rows represent source nodes and POIs in columns are target nodes. Transitional view focuses on mining transitional patterns and helps explore potential POIs from a global view.

Geographical View Hypergraph. The construction of geographical view hypergraph involves building a hypergraph that depicts POI-POI geographical relationships within some geographical constraints. Formally, we construct the geographical view hypergraph $\mathcal{G}_G = (\mathcal{V}_G, \mathcal{E}_G)$ where $\mathcal{V}_G$ denotes the POIs set. In the hypergraph $\mathcal{G}_G$, a hyperedge contains POIs within specific distance threshold Δ_d by calculating Haversine distance [7] between them. Incidence matrix $\mathbf{H}_G \in \mathbb{R}^{|\mathcal{L}|\times|\mathcal{E}_G|}$ depicts POI-POI geographical relationship. Specifically, if the Haversine distance between POI l_i and l_j is no larger than distance threshold Δ_d, we set $\mathbf{H}_G^{(i,j)} = 1$. The geographical view hypergraph takes geographical influence into consideration and reflects user geographical preferences.

After constructing hypergraphs from collaborative, transitional and geographical views, we can explicitly model and obtain richer POI representations in a disentangled learning way. Moreover, learning from multiple views leads to more holistic and accurate POI representations.

3.1.2 Disentangled Hypergraph Convolutional Networks. Aiming to learn multi-view disentangled POI representations from above three hypergraphs, we propose adjusted methods on aggregation and propagation of hypergrpah convolutional network, respectively. Before encoding, we initialize user embeddings $\mathbf{U} \in \mathbb{R}^{|\mathcal{U}|\times d}$ and POI embeddings $\mathbf{L} \in \mathbb{R}^{|\mathcal{L}|\times d}$ via look-up table, where d denotes embedding dimension. Next, we will introduce our proposed three hypergrpah neural networks separately.

Collaborative Hypergraph Convolutional Network. After constructing collaborative hypergraph $\mathcal{G}_C$, we develop collaborative hypergraph convolutional network with two-step information propagation scheme to capture high-order POIs iteratively. In the node-hyperedge-node propagation scheme, hyperedges serve as mediums for nodes aggregation within the hyperedge and propagation across hyperedges. To be more specific, for each node l of the hyperedge e in hypergraph $\mathcal{G}_C$, as shown in Figure 1, we perform the following two operations to update its representation:

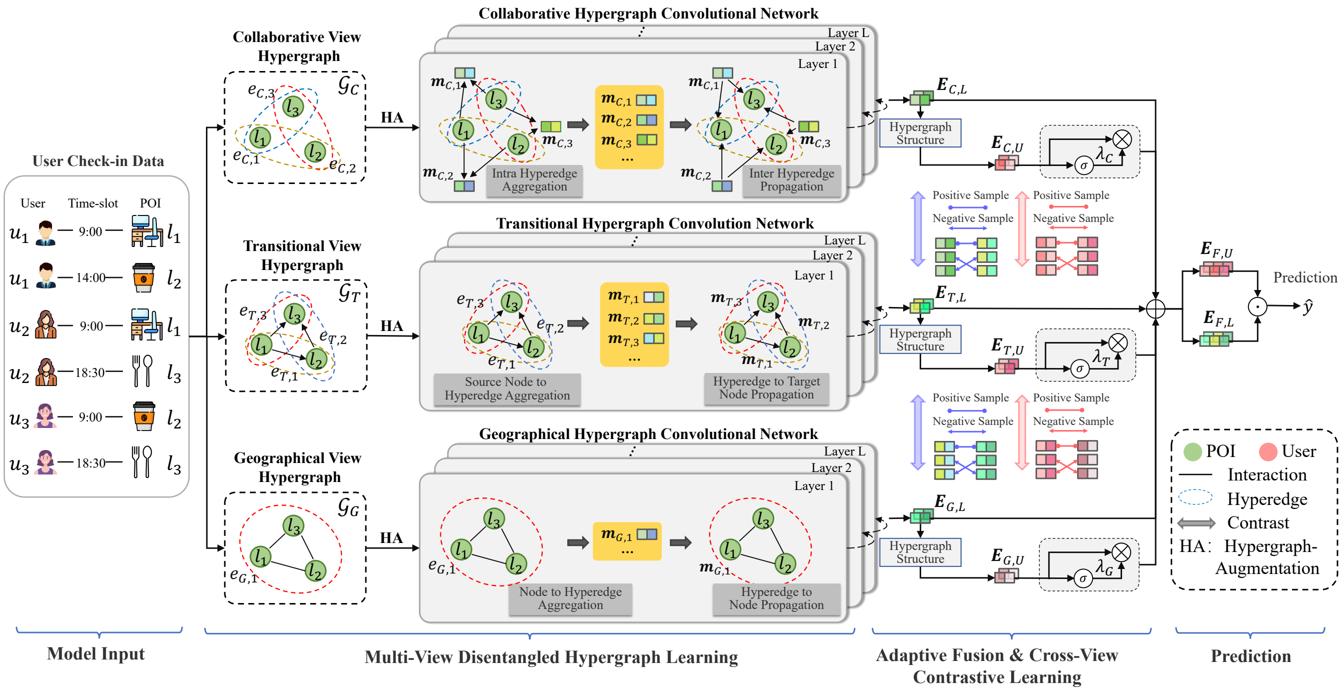

Figure 1: The overall framework of our proposed DCHL

Intra Hyperedge Aggregation. For hyperedge $e \in \mathcal{E}_C$, we aggregate its member embeddings to generate medium message as:

$$\mathbf{m}_{C,e} = \mathrm{AGG}_{n2e}(\{\mathbf{l} | l \in e\}) \tag{1}$$

where $\mathrm{AGG}_{n2e}(\cdot)$ denotes node to hyperedge aggregation function and $\mathbf{l} \in \mathbb{R}^d$ denotes the embedding of node l.

Inter Hyperedge Propagation. Since each node l may belong to several hyperedges, in this stage, we aggregate message from related hyperedges to refine the representation of node l as:

$$\bar{\mathbf{l}}_C = \mathrm{AGG}_{e2n}(\{\mathbf{m}_{C,e} | e \in \mathcal{E}_{C,l}\}) \tag{2}$$

where $\mathrm{AGG}_{e2n}(\cdot)$ denotes hyperedge to node propagation function, $\mathcal{E}_{C,l}$ denotes related hyperedges set of node l in hypergraph $\mathcal{G}_C$ and $\bar{\mathbf{l}}_C \in \mathbb{R}^d$ represents the refined embedding of node l in hypergraph $\mathcal{G}_C$.

Through above two steps, our collaborative hypergraph convolutional operation can capture collaborative signals that motivates users on choosing POIs. By stacking multiple layers, we can explore higher-order relationships, and the message passing from the $(\ell-1)$-th layer to the ℓ-th layer of node l is defined as:

$$\mathbf{e}_{C,l}^{(\ell)} = \bar{\mathbf{l}}_C^{(\ell)} + \bar{\mathbf{l}}_C^{(\ell-1)} \tag{3}$$

where $\mathbf{e}_{C,l}$ denotes the ℓ-th layer embedding of node l in collaborative view. We apply residual connections to alleviate the over-smoothing issue of GNN. Finally, we average the embeddings obtained at each layer to generate the final representation for node l:

$$\mathbf{e}_{C,l} = \frac{1}{L+1}\sum_{\ell=0}^{L} \mathbf{e}_{C,l}^{(\ell)} \tag{4}$$

where L denotes the total number of collaborative hypergraph convolutional layers. After that, we can get POI representations $\mathbf{E}_{C,L} \in \mathbb{R}^{|\mathcal{L}| \times d}$ of collaborative view. Besides, we implement $\mathrm{AGG}_{n2e}(\cdot)$ and $\mathrm{AGG}_{e2n}(\cdot)$ with mean pooling for its effectiveness and efficiency in collaborative, transitional and geographical views.

Transitional Hypergraph Convolutional Network. Since collaborative hypergraph convolutional network cannot deal with directed hypergraph, we propose transitional hypergraph convolutional network for transitional view hypergraph $\mathcal{G}_T$. Similarly, it adopts the two-step aggregation and propagation scheme but differs from above. To be more specific, for source node l_i, target node l_j and the hyperedge e in hypergraph $\mathcal{G}_T$, as shown in Figure 1, we introduce our proposed directed node-hyperedge-node scheme:

Source Node to Hyperedge Aggregation. Similar to intra hyperedge aggregation in collaborative hypergraph convolutional network, we aggregation source node embeddings to hyperedge $e \in \mathcal{E}_T$ to generate medium message $\mathbf{m}_{T,e} = \mathrm{AGG}_{n2e}(\{\mathbf{l}_i | l_i \in e\})$.

Hyperedge to Target Node Propagation. Since transitional view hypergraph $\mathcal{G}_T$ is directed, we can only propagate related hyperedge embeddings to target nodes to refine its representation:

$$\mathbf{l}_{T,j} = \mathrm{AGG}_{e2n}(\{\mathbf{m}_{T,e} | e \in \mathcal{E}_{l_j}\}) \tag{5}$$

where $\mathcal{E}_{l_j}$ denotes related hyperedges set that transfers transitional information from related source nodes to target node l_j, and $\mathbf{l}_{T,j} \in \mathbb{R}^d$.

Through the directed hypergraph convolutional operation, our transitional view hypergraph can capture POI-POI transitional relationships from the global view. Analogous to collaborative hypergraph convolutional network for propagation through L layers, we finally obtain POI representations $\mathbf{E}_{T,L} \in \mathbb{R}^{|\mathcal{L}|\times d}$ of transitional view.

Geographical Hypergraph Convolutional Network. In geographical view hypergraph $\mathcal{G}_G$, a hyperedge aggregates POIs within a specific distance threshold Δ_d. As illustrated in Fig. 1, for each POI l of hyperedge e in $\mathcal{G}_G$, we perform the following node-hyperedge-node paradigm to update POI representations of geographical view:

Node to Hyperedge Aggregation. Similar to intra hyperedge aggregation described in collaborative hypergraph convolutional network, we aggregate embeddings of POIs in hyperedge e to generate its medium message as $\mathbf{m}_{G,e} = \mathbf{AGG}_{n2e}(\{\mathbf{l}|l \in e\})$, where $\mathbf{l} \in \mathbb{R}^d$ denotes the embedding of node l.

Hyperedge to Node Propagation. Since each hyperedge only contains POIs that satisfy physical distance, the aggregated message $\mathbf{m}_{G,e}$ should not propagate across hyperedges unlimitedly. Specifically, hyperedge to node operation propagates aggregated message from other nodes within physical distance to update representation of node l as $\bar{\mathbf{l}}_G = \mathbf{AGG}_{e2n}(\{\mathbf{m}_{G,e}|e \in \mathcal{E}_l\})$, where $\mathcal{E}_l$ denotes related hyperedges set that meets geographical constraints.

Similar to above two hypergraph convolutional networks, we also stack L layers for higher-order neighbors information and apply residual connections to alleviate the over-smoothing issue. Finally, we obtain POI representations $\mathbf{E}_{G,L} \in \mathbb{R}^{|\mathcal{L}|\times d}$ of geographical view.

By designing three distinct adjusted aggregation and propagation methods of hypergraph convolutional networks, we achieve the goal to disentangle POI representations from collaborative, transitional and geographical views.

3.2 Adaptive Fusion for User Representation

As mentioned above, we obtain decoupled POI representations $\mathbf{E}_{C,L}$, $\mathbf{E}_{T,L}$ and $\mathbf{E}_{G,L} \in \mathbb{R}^{|\mathcal{L}|\times d}$ from collaborative, transitional and geographical views. Based on the structure of user-POI interactions, we learn disentangled user representation as follows:

$$\mathbf{E}_{X,U} = \mathbf{H}_C^T \cdot \mathbf{E}_{X,L} \tag{6}$$

where $X \in \{C, T, G\}$ and $\mathbf{H}_C^T \in \mathbb{R}^{|\mathcal{U}|\times|\mathcal{L}|}$ is the transpose of incidence matrix $\mathbf{H}_C \in \mathbb{R}^{|\mathcal{L}|\times|\mathcal{U}|}$ in collaborative view. Consequently, we obtain disentangled user representations $\mathbf{E}_{C,U}$, $\mathbf{E}_{T,U}$ and $\mathbf{E}_{G,U} \in \mathbb{R}^{|\mathcal{U}|\times d}$ that driven their behavior from learned POI intrinsic collaborative, transitional and geographical aspects.

Then the core problem comes to how to fuse decoupled multi-view user representations for their final preferences. Common fusion ways are element-wise addition, soft attention or concatenation operations, but they either ignore different importance of each view or lack of interpretability. To solve the limitation, we propose an adaptive fusion method to fuse three view-specific user representations by designing three different gates as:

$$\mathbf{E}_{F,U} = \lambda_C \mathbf{E}_{C,U} + \lambda_T \mathbf{E}_{T,U} + \lambda_G \mathbf{E}_{G,U} \tag{7}$$

where $\lambda_{X,U} = \sigma(\mathbf{E}_{X,U}\mathbf{W}_X)$, $X \in \{C, T, G\}$ and $\mathbf{E}_{F,U} \in \mathbb{R}^{|\mathcal{U}|\times d}$. $\mathbf{W}_X \in \mathbb{R}^d$ are trainable weights respectively and σ is the activation function. We choose *Sigmoid* here, for *ReLU* may cause information loss when the embedding is negative. Subsequently, our model can automatically discriminate the importance of collaborative, transitional and geographical views for user preference.

For final POI representation, we perform addition operation to fuse them as $\mathbf{E}_{F,L} = \mathbf{E}_{C,L} + \mathbf{E}_{T,L} + \mathbf{E}_{G,L}$, $\mathbf{E}_{F,L} \in \mathbb{R}^{|\mathcal{L}|\times d}$. The reason for not applying the same adaptive fusion but simple addition on final POI representation is to reduce complexity.

3.3 Cross-View Contrastive Learning

In this subsection, aiming to capture crucial cooperative associations among collaborative, transitional and geographical views, we design cross-view contrastive learning to augment view-specific user and POI representations with self supervision signals. Our contrastive learning component maximizes the agreement between views, allowing them work cooperatively to capture complementary recommendation effects. In particular, we take the same user/POI of different views as positive pairs (e.g., $(\mathbf{e}_{C,u}, \mathbf{e}_{T,u})$) and treat views of different users/POIs as negative pairs. Formally, we define our contrastive loss between collaborative and transitional views for user representations with InfoNCE [27] as:

$$\mathcal{J}_{C,T}^U = \frac{1}{|\mathcal{U}|}\sum_{u\in\mathcal{U}} -\log \frac{\exp(s(\mathbf{e}_{C,u}, \mathbf{e}_{T,u})/\tau)}{\Sigma_{u'\in\mathcal{U}} \exp(s(\mathbf{e}_{C,u}, \mathbf{e}_{T,u'})/\tau)} \tag{8}$$

where $s(\cdot,\cdot)$ is cosine similarity function and τ is a temperature hyer-parameter. Analogously, we can define contrastive loss for users between collaborative and geographical views as $\mathcal{J}_{C,G}^U$, and contrastive loss between transitional and geographical views as $\mathcal{J}_{T,G}^U$. After that, we sum contrastive loss between any two views to obtain final contrastive loss for user presentations as:

$$\mathcal{J}_{SSL}^U = \mathcal{J}_{C,T}^U + \mathcal{J}_{C,G}^U + \mathcal{J}_{T,G}^U \tag{9}$$

Similarly, we can obtain final contrastive loss for POI representations based on Equation 8-9 as $\mathcal{J}_{SSL}^L$. By averaging contrastive losses of users and POIs, we obtain the final contrastive loss as:

$$\mathcal{J}_{SSL} = \mathcal{J}_{SSL}^U + \mathcal{J}_{SSL}^L \tag{10}$$

To further alleviate the overfitting issue during cross-view contrastive learning, we perform hypergraph augmentation operation on three constructed hypergraphs with hyperedge dropout (dropout ratio μ) method. It helps improve robustness of learned representations to counter certain noise.

3.4 Prediction and Optimization

With the fused user and POI embeddings $\mathbf{E}_{F,U}$, $\mathbf{E}_{F,L}$, we compute the interaction score between user u and target POI l via dot product as $\hat{\mathbf{y}}_{u,l} = \text{softmax}(\mathbf{e}_{F,u}^T \mathbf{e}_{F,l})$. Here, $\mathbf{e}_{F,u}$ and $\mathbf{e}_{F,l}$ denote the final embedding of user u and POI l.

We formulate the learning objective as a cross-entropy loss function, which has been largely used in next POI recommendation:

$$\mathcal{J}_{Rec} = -\sum_{u\in\mathcal{U}}\sum_{l\in\mathcal{L}} \left(\mathbf{y}_{u,l}\log(\hat{\mathbf{y}}_{u,l}) + (1-\mathbf{y}_{u,l})\log(1-\hat{\mathbf{y}}_{u,l})\right) \tag{11}$$

where $\mathbf{y}_{u,l}$ equals to 1 if user u visits the POI l and 0 otherwise.

Finally, we integrate the self-supervised loss with our recommendation loss into a multi-task learning objective as follows:

$$\mathcal{J} = \mathcal{J}_{Rec} + \lambda_1 \mathcal{J}_{SSL} + \lambda_2 \|\Theta\|_2 \tag{12}$$

where $\|\Theta\|_2$ represents the $L2$ regularization of all parameters for preventing over-fitting issue under the control of λ_2 and λ_1 denotes the weight of self-supervised signals.

Table 1: Dataset statistics

	#Users	#POIs	#Check-ins	#Sessions	Sparsity
NYC	834	3,835	44,686	8,841	98.61%
TKY	2,173	7,038	308,566	41,307	97.82%
Meituan	18,451	49,143	677,609	140,069	99.93%

4 EXPERIMENTS

In this section, we present our experimental setup and results on three real-world datasets.

4.1 Experimental Setting

4.1.1 Datasets. We conduct experiments on three real-world LBSN datasets: **Foursquare-NYC** (NYC for abbreviation), **Foursquare-TKY** (TKY) [48] and **Meituan**. The NYC and TKY datasets are separately collected from New York city and Tokyo city over 11 months from Foursquare. The Meituan dataset is an industrial dataset collected from one of the largest life service platforms in China, and we sample the data of sponsored search advertising delivery service in Beijing from September to December in 2023.

Following previous works [17, 35], we first sort the recorded user interactions in each dataset in chronological order and eliminate unpopular POIs that are visited by less than 5 users. Then, we split each user's complete check-ins into sessions within 24 hours and remove those which includes fewer than 3 records. Furthermore, inactive users with less than 3 sessions are filtered out. According to [35], the first 80% sessions of each user are used for training and the rest for testing. To avoid data leakage when predicting next POI on testing dataset, we choose POIs that are later than all check-ins in training dataset. The statistics of pre-processed datasets are shown in Table 1.

4.1.2 Evaluation Metrics. Following most existing works in next POI recommendation, we adopt two widely used evaluation metrics: Recall@K and Normalized Discounted Cumulative Gain (NDCG@K). Recall@K measures the rate of the label within top-K recommendations and NDCG@K reflects the quality of ranking lists. For fairness, we repeat experiments on each metric for 10 times and report the averaged Recall@K and NDCG@K with the $K \in \{5, 10\}$.

4.1.3 Baselines. We compare our DCHL with following representative methods for next POI recommendation, including 1) statistical-based method UserPop; 2) RNN-based methods STGN and LSTPM; 3) self-attention-based method STAN; 4) GNN-based or hypergraph-based methods LightGCN, SGRec, GETNext, MSTHN and STHGCN; 6) graph or hypergraph contrastive learning based method DisenPOI and HCCF:

UserPop: It ranks the most popular POIs according to each user's visiting frequency.

STGN [51]: It is an LSTM-based model that introduces spatial and temporal gates for users' long- and short-term preferences.

LSTPM [35]: It is an LSTM-based model that captures long- and short-term preferences with a non-local network and geo-dilated LSTM.

STAN [25]: It is a self-attention based model that explicitly considers spatial-temporal influences within a user's check-in sequence.

LightGCN [14]: It is a GNN-based collaborative filtering model that omits the non-linear activation and feature transformation during propagation.

SGRec [19]: A GNN-based method, which proposes Seq2Graph augmentation and captures collaborative signals among one-hop neighbors.

GETNext [49]: A state-of-the-art GNN enhanced Transformer method, which utilizes global transition patterns and captures collaborative signals for next POI prediction.

MSTHN [17]: A state-of-the-art multi-view spatial-temporal hypergraph method that jointly learns representations of users and POIs from local and global views with hypergraph for capturing high-order collaborative signals.

STHGCN [45]: A state-of-the-art spatio-temporal hypergraph method that integrates complex high-order information and global collaborative relations among trajectories.

DisenPOI [29]: It is a state-of-the-art graph disentangled contrastive learning based method that extracts disentangled representations of both sequential and geographical influences with contrastive learning.

HCCF [40]: A state-of-the-art hypergraph contrastive learning based method that jointly captures local and global collaborative relations with a hypergraph enhanced cross-view contrastive learning architecture.

For fairness comparison, we remove the POI categorical information in SGRec, GETNext and STHGCN, when comparing with other methods that do not use.

4.1.4 Parameter Settings. Our experiments are conducted with PyTorch 1.12.0 on 80 GB Nvidia A100 GPU. For baselines, we firstly preserve the settings as provided in original papers and fine-tune each model's hyperparameters on three datasets. For our DCHL, we adopt Adam [16] as optimizer with a learning rate of $1e^{-3}$, weight decay of $5e^{-4}$ and hyperedge dropout rate of $\{0.25, 0.5, 0.75, 1\}$. We apply the same dimension size $d = 128$ for user and POI embeddings. In each batch, we pad sessions which do not meet the maximum session length in batch. Furthermore, we empirically choose 2.5km (for NYC and TKY) and 0.15km for Meituan as distance threshold. The layer number of hypergraph convolutional network is chosen from $\{1, 2, 3, 4, 5\}$. The temperature parameter τ is searched from the range $\{0.1, 0.3, 0.5, 1, 3, 5, 10\}$ to control the strength of gradients in our contrastive learning. The regularization weight λ_1 and λ_2 are tuned from the range $\{1e^{-5}, 1e^{-4}, 1e^{-3}, 1e^{-2}, 1e^{-1}\}$ for loss balance.

4.2 Performance Comparison

The results of all the methods are reported in Table 2. From the results, we have the following observations.

Our proposed DCHL achieves the best results on all datasets. Our DCHL consistently outperforms all baselines on three datasets in terms of all evaluation metrics. We contribute the improvements to the following aspects: i) By considering collaborative, transitional and geographical views via disentangled hypergraph learning, our DCHL disentangles multi-view user preferences that driven their behaviors and alleviates the data sparsity issue by proposing adjusted

Table 2: Performances comparison on three datasets in terms of Recall and NDCG. The best and the second best performances are bolded and underlined, respectively. The relative improvements are calculated between the best and the second best scores

Method	NYC				TKY				Meituan			
	R@5	R@10	N@5	N@10	R@5	R@10	N@5	N@10	R@5	R@10	N@5	N@10
UserPop	0.2866	0.3297	0.2283	0.2423	0.2229	0.2668	0.1718	0.1861	0.2028	0.2489	0.1961	0.2016
STGN	0.2371	0.2594	0.2261	0.2307	0.2112	0.2587	0.1482	0.1589	0.2004	0.2441	0.1891	0.1933
LSTPM	0.2495	0.2668	0.2425	0.2483	0.2203	0.2703	0.1556	0.1734	0.2021	0.2615	0.2132	0.2287
STAN	0.3523	0.3827	0.3025	0.3137	0.2621	0.3317	0.2074	0.2189	0.2879	0.3168	0.2337	0.2419
LightGCN	0.3221	0.3488	0.2958	0.3042	0.2213	0.2594	0.1977	0.2098	0.2556	0.2979	0.2211	0.2307
SGRec	0.3451	0.3723	0.3052	0.3178	0.2537	0.3213	0.2221	0.2447	0.2859	0.3112	0.2362	0.2431
GETNext	0.3572	0.3866	0.3079	0.3094	0.2686	0.3282	0.2212	0.2242	0.3125	0.3478	0.2413	0.2509
MSTHN	0.4076	<u>0.4398</u>	0.3612	0.3702	0.3378	<u>0.3927</u>	0.2567	<u>0.2721</u>	0.3641	<u>0.3977</u>	<u>0.2842</u>	<u>0.2963</u>
STHGCN	<u>0.4081</u>	0.4366	<u>0.3626</u>	<u>0.3703</u>	<u>0.3392</u>	0.3924	<u>0.2592</u>	0.2693	<u>0.3653</u>	0.3944	0.2838	0.2959
DisenPOI	0.3577	0.3831	0.2979	0.3071	0.2692	0.3314	0.2263	0.2332	0.3275	0.3516	0.2435	0.2521
HCCF	0.3534	0.3745	0.3025	0.3134	0.2689	0.3253	0.2325	0.2429	0.3260	0.3495	0.2523	0.2638
DCHL	**0.4385**	**0.4861**	**0.3859**	**0.4017**	**0.3662**	**0.4083**	**0.2951**	**0.3078**	**0.3957**	**0.4286**	**0.3113**	**0.3220**
%Improv	+7.45	+10.53	+6.43	+8.48	+7.96	+3.97	+13.85	+13.12	+8.32	+7.77	+9.54	+8.67

aggregation and propagation methods of hypergraph convolutional network. ii) Benefiting from our designed cross-view contrastive learning schema, our DCHL can self-augment learned view-specific representations and distill supervision signals for capturing complementary recommendation effects during the learning process.

Among all baselines, methods that leverage spatial-temporal information perform better than those do not use. For example, GNN-based method SGRec, DisenPOI and GETNext also surpasses LightGCN on three datasets, especially by 23.86% in terms of Recall@10 on TKY dataset. DCHL, MSTHN and STHGCN under hypergraph neural network paradigm outperform HCCF that only focuses on user-item interactions modeling on three datasets. Additionally, our DCHL outperforms MSTHN and STHGCN, for modeling potential decoupled factors behind interactions. Moreover, DisenPOI which disentangles geographical and sequential influences also outperforms SGRec on both recall and ndcg, and surpasses GETNext on recall but weak on ndcg. The reason is due to the utilization of global transitional influence. They prove the importance and necessity of modeling multi-view representations for users and POIs in an explicit disentangled learning way.

From Table 2, methods that leverage non-consecutive POIs information perform better than those mainly focus on sequential modeling. For instance, STAN outperforms RNN-based methods STGN and LSTPM for learning from non-adjacent POIs intra sequence. HGNN-based methods (e.g., MSTHN, STHGCN and DCHL) perform better than GNN-based methods (e.g., SGRec and GETNext) for capturing higher-order collaborative signals. They can alleviate the data sparsity issue and over-smoothing issue of GNN. Moreover, most graph or hypergraph contrastive learning based methods perform better than GNN-based approaches (e.g., LightGCN and SGRec), indicating the effectiveness of capturing cooperative associations between views for performance enhancements. The utilization of cross-view contrastive learning encourages self-supervision between each view and captures implicit complementary effects.

Table 3: Ablation study on key components of DCHL w.r.t. Recall@10 and NDCG@10

Method	NYC		TKY		Meituan	
	R@10	N@10	R@10	N@10	R@10	N@10
w/o C	0.4741	0.3941	0.3925	0.3007	0.3980	0.3050
w/o T	0.4839	0.4011	0.4008	0.2990	0.3916	0.2990
w/o G	0.4817	0.3969	0.3982	0.3048	0.4148	0.3143
w/o CL	0.4841	0.4006	0.4021	0.2983	0.3598	0.2721
DCHL	0.4861	0.4017	0.4044	0.3078	0.4286	0.3220

4.3 Ablation Study

4.3.1 Effectiveness of Key Components of DCHL. To investigate the effectiveness of each component of our DCHL, we conduct an ablation study to examine component-specific benefits from following: i) w/o C that removes collaborative view; ii) w/o T that removes transitional view; iii) w/o G that removes geographical view; iv) w/o CL that removes cross-view contrastive learning. The performance results are reported in Table 3 and we have the following observations:

First, when removing collaborative view of DCHL, performances drop clearly. It strongly indicates the importance of modeling user-POI interactions, for it can distill higher-order collaborative signals. Second, when removing transitional view of DCHL, performances decrease slightly compared with other variants on NYC and TKY datasets. It proves that collaborative view and geographical view achieve complementary effects for performances and user-POI interactions could reflect part of user preference. Since Meituan dataset is much sparser than NYC and TKY datasets, capturing global transitional relationship helps alleviate the data sparsity issue. The result is consistent with GETNext and STHGCN, which also consider global transitional influence. Third, when removing geographical view of DCHL, performances drop on NYC and TKY, for not capturing disentangled mutual effects of geographical view. However, on Meituan dataset, it has least effect. The reason is due to the adoption

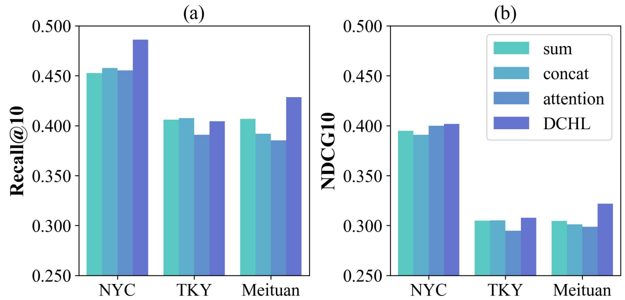

Figure 2: Performance comparison among various information fusion methods (i.e., sum, concatenation and attention)

of data from delivery service and users are not so sensitive to geographical influence compared with check-ins by themselves. Fourth, when removing cross-view contrastive learning, performances have least impact on both NYC and datasets, but the most impact on Meituan dataset. It indicates the important complementary effects between disentangled three views when user considers delivery services. Moreover, cross-view contrastive learning can enhance performances via self-supervision augmentation.

4.3.2 Effectiveness of Adaptive Fusion for User Representation. After we obtain three disentangled user preferences, we compare our adaptive fusion method with various information fusion approaches (i.e., concatenation, element-wise sum and attention) for comprehensive user representations (Figure 2). From Figure 2, the proposed adaptive fusion method outperforms other variants on the three datasets, especially on Recall@10. The reason owes to the learned weights can adaptively balance the contributions of collaborative, geographical and transitional information, respectively. Noticeably, trainable fusion method attention play worse than our proposed adaptive fusion method on three datasets and even worse than element-wise sum and concatenation. It heavily depends on the quality of learnable weights, while our adaptive fusion learns the weight based on three disentangled user preferences.

4.4 User Cold-Start Performance Analysis

We further verify if our DCHL could alleviate the data sparsity issue. We then divide users into different groups based on the number of their interactions, e.g., the top 15% as the most active users, the bottom 15% as inactive users and the rest are normal users. Each group are separately measured by recall and ndcg metrics. Here, we evaluate different user group performances based on trained DCHL (Figure 3 (a)(b)) and trained DCHL model with learned users and POIs embeddings from corresponding training datasets (Figure 3 (c)(d)). From Figure 3, active user group performances outperform normal and inactive groups on NYC and TKY datasets, but are worst on Meituan dataset. The reason is that active users would hang on delivery service platforms even without strong purchase intentions and their behaviors will introduce noise and lead to performance declination. It is worth noting that inactive user group outperforms other two groups on NDCG@10 and achieve great performance on Recall@10. It validates the importance of high-order collaborative signals and alleviates the data sparsity issue in user cold-start

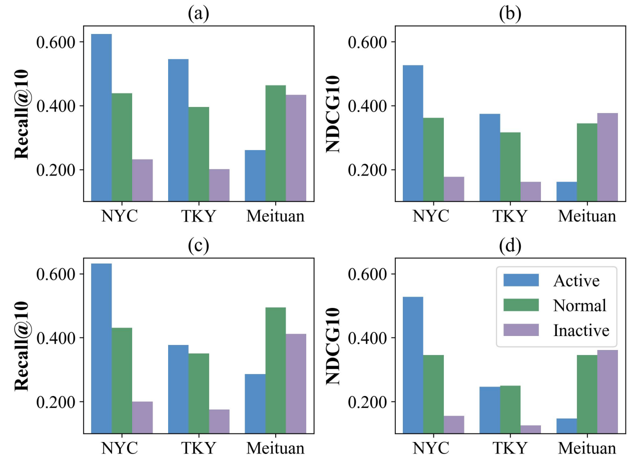

Figure 3: User cold-start performance comparison based on trained DCHL (a)(b) and trained DCHL + corresponding trained users and POIs embeddings (c)(d)

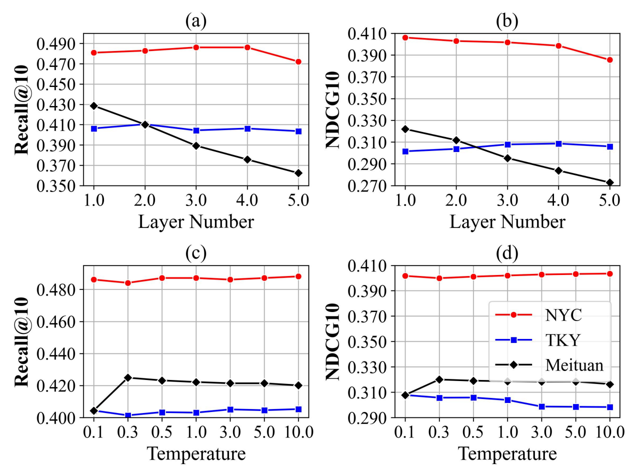

Figure 4: Hyperparameters study of the DCHL

scenario. From Figure 3(c)(d), benefiting from learned embeddings, normal user group achieves better results compared with that without trained embeddings on sparser and larger Meituan dataset. It provides some solutions when dealing with large-scale graphs or hypergraphs in practical industrial scenarios.

4.5 Hyperparameter Analysis

We further qualitatively analyze the impacts of layer number and temperature in DCHL.

Impact of Layer Number. To investigate the impact of stacking hypergraph convolutional layers, we conduct experiments with number of layer in $\{1, 2, 3, 4, 5\}$. As illustrated in Figure 4(a)(b), our DCHL balances Recall@10 and NDCG@10 when stacking 3 layers on NYC and TKY dataset, respectively. It proves that our DCHL can capture high-order collaborative signals effectively. When aggregating and propagating on 1 layer, our DCHL achieves the best results on Meituan dataset. The possible cause of dropping would be the introducing of noise.

Impact of Temperature. To investigate the impact of temperature τ in controlling the strength of gradients in our contrastive learning, we conduct experiments with temperature τ in $\{0.1, 0.3, 0.5, 1, 3, 5, 10\}$. From Figure 4 (c)(d), our DCHL balances Recall@10 and NDCG@10 with $\tau = 0.5$ on three datasets. When $\tau = 0.1$, the results on Meituan dataset are obviously worse than others, for the gradients are sharp and lead to performance degradation.

4.6 In-Depth Analysis of DCHL

To explore the effect of our adjusted aggregation and propagation method of hypergraph convolutional network, we maintain other parts of DCHL and replace each hypergraph convolutional network with LightGCN [14] and HGNN+ [11] (e.g., C-LightGCN and C-HGNN for replacement in collaborative views, and other two views are similar to that). From Table 4, when replacing specific hypergraph convolutional network of each view with HGNN+, performances degrade in varying degrees. Performance drop clearly on both Recall@10 and NDCG@10 when replacing collaborative hypergraph convolutional network with LightGCN. The reason may be due to the lack of high-order collaborative signals among users and over-smoothing issue. When replacing transitional hypergraph convolutional network with HGNN+ or LightGCN, performances also drop and even worse with HGNN+. The two variants are designed based on undirected message passing scheme and cannot be perfectly adapted into directed transitional relationship.

Table 4: Performance comparison of different graph or hypergraph convolutional methods w.r.t. Recall@10 and NDCG@10

Method	NYC		TKY		Meituan	
	R@10	N@10	R@10	N@10	R@10	N@10
C-HGNN+	0.4752	0.4002	0.3954	0.2997	0.4222	0.3188
T-HGNN+	0.4761	0.3979	0.3891	0.2982	0.4239	0.3180
G-HGNN+	0.4754	0.3985	0.3906	0.2967	0.4231	0.3115
C-LightGCN	0.4663	0.3968	0.3665	0.2958	0.4039	0.3098
T-LightGCN	0.4754	0.3997	0.3969	0.3004	0.4251	0.3181
G-LightGCN	0.4732	0.3989	0.3966	0.2997	0.4214	0.3203
DCHL	0.4861	0.4017	0.4044	0.3078	0.4286	0.3220

5 RELATED WORK

5.1 Next POI Recommendation

Next POI recommendation aims to suggest next suitable location for users based on their recent visiting behaviours. Most existing methods treat it as a sequential prediction task and adopt sequential methods to solve, ranging from Markov chain [4] to RNN and its variants [6, 35, 51] and recent self-attention mechanism [22, 25]. However, these sequential based methods mainly focus on modeling each user's trajectory and overlook non-consecutive POIs in the trajectory or among users. The rapid development of graph learning have gained great attention in next POI recommendation, ranging from graph learning [42] to hypergraph embedding [46, 47] and more recent graph or hypergraph neural networks [8, 15, 17–19, 21, 23, 31, 37, 38]. For example, Graph-Flashback [31] utilizes a spatial-temporal knowledge graph to endow POI representations and incorporates them into RNN-based methods to capture sequential transition patterns. Lai et al. [17] leverage multi-view spatial-temporal enhanced hypergraph network to capture spatial-temporal information and high-order collaborative signals, validating the strong ability of HGNN for next POI recommendation. Nonetheless, most graph or hypergraph based methods ignore the fact that user preferences are diverse in terms of various aspects, resulting in suboptimal and entangled user representations. Though few works [20, 28, 29] have adopted disentangled learning to learn aspect-specific representations separately, they only fuse them simply, failing to differentiate the importance of each aspect. To tackle the challenge, we propose a multi-view disentangled hypergraph learning method, combined with a novel adaptive fusion approach, to disentangle and adaptively fuse user representations from collaborative, transitional and geographical views.

5.2 Graph or Hypergraph Contrastive Learning for Recommendation

Recently, contrastive learning [32, 50] has been proven effective in addressing data sparsity issue and received considerable attention in various recommendation scenarios, such as collaborative filtering [3, 39, 40], sequential or session recommendation [41, 43], click-through rate prediction [13] and bundle recommendation [26]. One research line is to adapt various data augmentations (e.g., node dropout and edge dropout) via stochastic operations on the original data, leading to random noise perturbation [13, 39, 44]. Another branch aims to create self-supervision signals across views under multi-view learning or multi-channel learning paradigm [3, 29, 40]. For example, Qin et al. [29] disentangles sequential and geographical influences in a self-supervised way. Different from them, our DCHL designs disentangled collaborative, transitional and geographical hypergraphs and creates contrastive signals across views to capture complementary effects for next POI recommendation.

6 CONCLUSION

This paper mainly focuses on disentangling multi-view user preferences that driven their behaviors and capturing crucial cooperative associations between views to enhance complementary recommendation effects for next POI recommendation. To achieve our goal, we present a novel framework Disentangled Contrastive Hypergraph Learning (DCHL). Our DCHL performs disentangled hypergraph learning to decouple intrinsic aspects among collaborative, transitional and geographical views with adjusted aggregation and propagation methods. Additionally, we introduce cross-view contrastive learning with self-augmentation for capturing complementary effects. Experimental results on three datasets demonstrate the effectiveness of our DCHL. In future, we may explore implicit disentangled learning method to model intents behind user-POI interactions and interpretability of user decisions and recommendations for next POI recommendation.

REFERENCES

[1] Song Bai, Feihu Zhang, and Philip HS Torr. 2021. Hypergraph convolution and hypergraph attention. *Pattern Recognition* 110 (2021), 107637.

[2] Austin R Benson, David F Gleich, and Jure Leskovec. 2016. Higher-order organization of complex networks. *Science* 353, 6295 (2016), 163–166.

[3] Mengru Chen, Chao Huang, Lianghao Xia, Wei Wei, Yong Xu, and Ronghua Luo. 2023. Heterogeneous graph contrastive learning for recommendation. In *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining*. 544–552.

[4] Chen Cheng, Haiqin Yang, Michael R Lyu, and Irwin King. 2013. Where you like to go next: Successive point-of-interest recommendation. In *Twenty-Third international joint conference on Artificial Intelligence*.

[5] Eunjoon Cho, Seth A Myers, and Jure Leskovec. 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*. 1082–1090.

[6] Kyunghyun Cho, Bart van Merrienboer, Çaglar Gülçehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. In *EMNLP*. ACL, 1724–1734.

[7] Nitin R Chopde and Mangesh Nichat. 2013. Landmark based shortest path detection by using A* and Haversine formula. *International Journal of Innovative Research in Computer and Communication Engineering* 1, 2 (2013), 298–302.

[8] Weizhen Dang, Haibo Wang, Shirui Pan, Pei Zhang, Chuan Zhou, Xin Chen, and Jilong Wang. 2022. Predicting Human Mobility via Graph Convolutional Dual-attentive Networks. In *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*. 192–200.

[9] Jie Feng, Yong Li, Chao Zhang, Funing Sun, Fanchao Meng, Ang Guo, and Depeng Jin. 2018. Deepmove: Predicting human mobility with attentional recurrent networks. In *Proceedings of the 2018 world wide web conference*. 1459–1468.

[10] Yifan Feng, Haoxuan You, Zizhao Zhang, Rongrong Ji, and Yue Gao. 2019. Hypergraph neural networks. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 33. 3558–3565.

[11] Yue Gao, Yifan Feng, Shuyi Ji, and Rongrong Ji. 2023. HGNN$^+$: General Hypergraph Neural Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 45, 3 (2023), 3181–3199. https://doi.org/10.1109/TPAMI.2022.3182052

[12] Haoyu Han, Mengdi Zhang, Min Hou, Fuzheng Zhang, Zhongyuan Wang, Enhong Chen, Hongwei Wang, Jianhui Ma, and Qi Liu. 2020. STGCN: a spatial-temporal aware graph learning method for POI recommendation. In *2020 IEEE International Conference on Data Mining (ICDM)*. IEEE, 1052–1057.

[13] Tianqi He, Kaiyuan Li, Shan Chen, Haitao Wang, Qiang Liu, Xingxing Wang, and Dong Wang. 2023. DMBIN: A Dual Multi-behavior Interest Network for Click-Through Rate Prediction via Contrastive Learning. In *SIGIR*. 1366–1375.

[14] Xiangnan He, Kuan Deng, Xiang Wang, Yan Li, Yongdong Zhang, and Meng Wang. 2020. Lightgcn: Simplifying and powering graph convolution network for recommendation. In *Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval*. 639–648.

[15] Zheng Huang, Jing Ma, Yushun Dong, Natasha Zhang Foutz, and Jundong Li. 2022. Empowering Next POI Recommendation with Multi-Relational Modeling. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2034–2038.

[16] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[17] Yantong Lai, Yijun Su, Lingwei Wei, Gaode Chen, Tianci Wang, and Daren Zha. 2023. Multi-view Spatial-Temporal Enhanced Hypergraph Network for Next POI Recommendation. In *International Conference on Database Systems for Advanced Applications*. Springer, 237–252.

[18] Yantong Lai, Yijun Su, Lingwei Wei, Tianci Wang, Daren Zha, and Xin Wang. 2024. Adaptive Spatial-Temporal Hypergraph Fusion Learning for Next POI Recommendation. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 7320–7324.

[19] Yang Li, Tong Chen, Yadan Luo, Hongzhi Yin, and Zi Huang. 2021. Discovering Collaborative Signals for Next POI Recommendation with Iterative Seq2Graph Augmentation. In *Proceedings of the 30th IJCAI*. 1491–1497.

[20] Yinfeng Li, Chen Gao, Quanming Yao, Tong Li, Depeng Jin, and Yong Li. 2022. DisenHCN: Disentangled Hypergraph Convolutional Networks for Spatiotemporal Activity Prediction. *Proceedings of IEEE 38th International Conference on Data Engineering (ICDE)* (2022).

[21] Zeyu Li, Wei Cheng, Haiqi Xiao, Wenchao Yu, Haifeng Chen, and Wei Wang. 2021. You are what and where you are: Graph enhanced attention network for explainable poi recommendation. In *Proceedings of the 30th ACM International conference on information & knowledge management*. 3945–3954.

[22] Defu Lian, Yongji Wu, Yong Ge, Xing Xie, and Enhong Chen. 2020. Geography-aware sequential location recommendation. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*. 2009–2019.

[23] Nicholas Lim, Bryan Hooi, See-Kiong Ng, Yong Liang Goh, Renrong Weng, and Rui Tan. 2022. Hierarchical multi-task graph recurrent network for next poi recommendation. In *Proceedings of the 45th international ACM SIGIR conference on Research and development in Information Retrieval*.

[24] Qiang Liu, Shu Wu, Liang Wang, and Tieniu Tan. 2016. Predicting the next location: A recurrent model with spatial and temporal contexts. In *Thirtieth AAAI conference on artificial intelligence*.

[25] Yingtao Luo, Qiang Liu, and Zhaocheng Liu. 2021. Stan: Spatio-temporal attention network for next location recommendation. In *Proceedings of the Web Conference 2021*. 2177–2185.

[26] Yunshan Ma, Yingzhi He, An Zhang, Xiang Wang, and Tat-Seng Chua. 2022. CrossCBR: cross-view contrastive learning for bundle recommendation. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 1233–1241.

[27] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. 2018. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748* (2018).

[28] Yingrong Qin, Chen Gao, Yue Wang, Shuangqing Wei, Depeng Jin, Jian Yuan, and Lin Zhang. 2022. Disentangling Geographical Effect for Point-of-Interest Recommendation. *IEEE Transactions on Knowledge and Data Engineering* (2022).

[29] Yifang Qin, Yifan Wang, Fang Sun, Wei Ju, Xuyang Hou, Zhe Wang, Jia Cheng, Jun Lei, and Ming Zhang. 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*. 508–516.

[30] Hossein A Rahmani, Mohammad Aliannejadi, Mitra Baratchi, and Fabio Crestani. 2022. A systematic analysis on the impact of contextual information on point-of-interest recommendation. *ACM Transactions on Information Systems (TOIS)* 40, 4 (2022), 1–35.

[31] Xuan Rao, Lisi Chen, Yong Liu, Shuo Shang, Bin Yao, and Peng Han. 2022. Graph-flashback network for next location recommendation. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 1463–1471.

[32] Xubin Ren, Wei Wei, Lianghao Xia, and Chao Huang. 2024. A Comprehensive Survey on Self-Supervised Learning for Recommendation. *arXiv preprint arXiv:2404.03354* (2024).

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

[34] Yijun Su, Xiang Li, Wei Tang, Ji Xiang, and Yuanye He. 2018. Next check-in location prediction via footprints and friendship on location-based social networks. In *2018 19th IEEE International Conference on Mobile Data Management (MDM)*. IEEE, 251–256.

[35] Ke Sun, Tieyun Qian, Tong Chen, Yile Liang, Quoc Viet Hung Nguyen, and Hongzhi Yin. 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. 214–221.

[36] Tianci Wang, Yantong Lai, Gaode Chen, Ruohan Wang, Jiahui Shen, and Ji Xiang. 2023. A Dynamic-aware Heterogeneous Graph Neural Network for Next POI Recommendation. In *Pacific Rim International Conference on Artificial Intelligence*. Springer, 313–326.

[37] Zhaobo Wang, Yanmin Zhu, Haobing Liu, and Chunyang Wang. 2022. Learning Graph-based Disentangled Representations for Next POI Recommendation. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1154–1163.

[38] Zhaobo Wang, Yanmin Zhu, Qiaomei Zhang, Haobing Liu, Chunyang Wang, and Tong Liu. 2022. Graph-Enhanced Spatial-Temporal Network for Next POI Recommendation. *ACM Transactions on Knowledge Discovery from Data (TKDD)* 16, 6 (2022), 1–21.

[39] Jiancan Wu, Xiang Wang, Fuli Feng, Xiangnan He, Liang Chen, Jianxun Lian, and Xing Xie. 2021. Self-supervised graph learning for recommendation. In *Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval*. 726–735.

[40] Lianghao Xia, Chao Huang, Yong Xu, Jiashu Zhao, Dawei Yin, and Jimmy Huang. 2022. Hypergraph contrastive collaborative filtering. In *Proceedings of the 45th International ACM SIGIR conference on research and development in information retrieval*. 70–79.

[41] Xin Xia, Hongzhi Yin, Junliang Yu, Qinyong Wang, Lizhen Cui, and Xiangliang Zhang. 2021. Self-supervised hypergraph convolutional networks for session-based recommendation. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 35. 4503–4511.

[42] Min Xie, Hongzhi Yin, Hao Wang, Fanjiang Xu, Weitong Chen, and Sen Wang. 2016. Learning graph-based poi embedding for location-based recommendation. In *Proceedings of the 25th ACM international on conference on information and knowledge management*. 15–24.

[43] Xu Xie, Fei Sun, Zhaoyang Liu, Shiwen Wu, Jinyang Gao, Jiandong Zhang, Bolin Ding, and Bin Cui. 2022. Contrastive learning for sequential recommendation. In *ICDE*. IEEE, 1259–1273.

[44] Dongkuan Xu, Wei Cheng, Dongsheng Luo, Haifeng Chen, and Xiang Zhang. 2021. Infogcl: Information-aware graph contrastive learning. *Advances in Neural Information Processing Systems* 34 (2021), 30414–30425.

[45] Xiaodong Yan, Tengwei Song, Yifeng Jiao, Jianshan He, Jiaotuan Wang, Ruopeng Li, and Wei Chu. 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*. 403–412.

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

[47] Dingqi Yang, Bingqing Qu, Jie Yang, and Philippe Cudré-Mauroux. 2020. Lbsn2vec++: Heterogeneous hypergraph embedding for location-based social networks. *IEEE Transactions on Knowledge and Data Engineering* (2020).

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

[49] Song Yang, Jiamou Liu, and Kaiqi Zhao. 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*. 1144–1153.

[50] Junliang Yu, Hongzhi Yin, Xin Xia, Tong Chen, Jundong Li, and Zi Huang. 2023. Self-supervised learning for recommender systems: A survey. *IEEE Transactions on Knowledge and Data Engineering* (2023).

[51] Pengpeng Zhao, Anjing Luo, Yanchi Liu, Fuzhen Zhuang, Jiajie Xu, Zhixu Li, Victor S Sheng, and Xiaofang Zhou. 2020. Where to go next: A spatio-temporal gated network for next poi recommendation. *IEEE Transactions on Knowledge and Data Engineering* (2020).