IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Traitement du Signal
Vol. 43, No. 1, February, 2026, pp. 275-301
Journal homepage: http://iieta.org/journals/ts

Visual Perception–Driven Point-of-Interest Prediction via Multi-View Disentanglement and Spatial Encoding

Chenxuan Lai, Huai Lin, Dewen Seng*

School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou 310018, China

Corresponding Author Email: sengdw@hdu.edu.cn

https://doi.org/10.18280/ts.430120

Received: 30 September 2025
Revised: 25 December 2025
Accepted: 23 February 2026
Available online: 28 February 2026

Keywords:
spatial representation learning, graph convolutional networks, multi-view disentanglement, contrastive learning, spatial perception, positional encoding

ABSTRACT

The effective modeling of complex multidimensional interactions in spatial environments is a critical challenge in intelligent sensing and urban computing. Existing approaches often fail to simultaneously capture global structural dependencies and local spatial attributes when modeling spatial interaction intensity and location distribution patterns. Furthermore, representation entanglement during multi-source information fusion limits the capacity to characterize complex scenarios. To address these limitations, a graph-based representation learning framework that integrates spatial frequency information and positional encoding was proposed from the perspective of spatial perception modeling. Furthermore, a spatial representation model based on multi-view disentanglement and adaptive fusion was developed. Spatial interaction frequency was introduced to weight node relationships, thereby enhancing the expressiveness of graph structures in capturing significant interaction patterns. In parallel, positional encoding was incorporated into node representations to improve sensitivity to spatial distribution characteristics. Building upon this foundation, a multi-view disentangled representation learning framework was constructed, in which global interaction dependencies, temporal behavioral patterns, and spatial proximity features were modeled separately. A contrastive learning strategy was employed to enforce consistency across views while promoting complementary information extraction. Subsequently, an adaptive fusion mechanism was utilized to generate unified spatial representations. Experimental results on real-world spatial interaction datasets demonstrate that superior performance was achieved across multiple evaluation metrics compared to baseline methods, indicating that prediction accuracy for spatial targets is significantly improved. These findings suggest that the integration of graph representation learning with multi-view disentanglement enhances the model's capability to capture complex spatial scenarios and offers a promising approach for visual perception–driven spatial intelligence applications.

1. INTRODUCTION

1.1 Research background and significance

With the rapid proliferation of mobile internet technologies and intelligent terminals, location-based spatial interaction data have increased substantially. Data derived from check-in behaviors tightly couple user activities with geographic locations and are often accompanied by multimodal information, including textual reviews, ratings, and images, thereby forming large-scale spatial behavioral datasets. Under such circumstances, the extraction of stable behavioral patterns from massive data and the realization of personalized spatial prediction have become critical research challenges. Spatial target prediction and recommendation exhibit substantial practical value across a wide range of applications. These approaches are not only fundamental to local service platforms and location-based service systems but also provide essential support for route planning, urban resource allocation, and human mobility analysis. However, compared with traditional recommendation tasks, spatial behavior modeling is subject to more complex constraints. Specifically, spatial distance exerts a significant influence, leading to a distance decay effect in user behavior [1]. Meanwhile, interaction data are typically highly sparse, with limited user visit records available [2]. In addition, user behavior demonstrates pronounced temporal dynamics and periodic patterns, which further increase the complexity of modeling.

To address these challenges, methodological development has evolved from rule-based approaches to representation learning paradigms [3]. Early methods relied primarily on distance-based or popularity-based ranking strategies, which were insufficient for capturing individual preferences. Subsequently, collaborative filtering and matrix factorization techniques introduced latent representation spaces; however, their effectiveness was constrained by data sparsity and limited expressive capacity under linear assumptions. More recently, graph neural networks have enabled the learning of high-quality representations from complex relational structures, thereby providing a promising technical foundation

for spatial modeling.

1.2 Related work

To address the challenges of sparsity and heterogeneity in spatial interaction data, a relatively well-structured technical framework has gradually been established in existing studies [4, 5]. Current approaches can be broadly categorized into traditional methods, deep learning–based methods, graph neural network–based methods, and disentanglement and contrastive learning–based methods [6-8].

1.2.1 Traditional spatial modeling methods

Early approaches were primarily developed based on collaborative filtering paradigms, in which predictions were generated by exploiting user similarity [9]. However, performance was observed to be unstable under sparse data conditions. To mitigate this limitation, matrix factorization techniques were introduced, where low-dimensional latent representations were learned to alleviate sparsity issues. Representative methods such as the ranking-based geographical factorization model [10] incorporated geographical factors into the modeling process, while the factorized personalized Markov chain [11] enhanced representational capacity by integrating multiple contextual features. In addition, spatial semantic information has been explored through representation learning techniques [12], and spatiotemporal factors have been incorporated into ranking-based modeling frameworks [13]. Nevertheless, these approaches remain fundamentally constrained by their linear modeling capacities.

1.2.2 Deep learning–based spatial representation learning methods

Deep learning–based approaches have been developed to enhance the modeling capacity of spatial behavioral patterns through neural network architectures. In sequence modeling, recurrent neural networks and their variants have been widely adopted to capture temporal dependencies in user behavior [14, 15], and have been applied extensively to spatial prediction tasks [16]. Long short-term memory and gated recurrent unit architectures have further improved the capability to model long-term dependencies. For instance, the long- and short-term preference modeling method integrates both long-term and short-term user interests, enabling multi-scale modeling [17]. The introduction of attention mechanisms has allowed critical behavioral segments to be selectively emphasized. Representative models such as the Spatio-temporal attention network incorporate spatiotemporal attention to enhance representation performance [18]. Subsequently, Transformer-based architectures [19] have demonstrated superior performance in sequence modeling tasks. Models such as the self-attentive sequential recommendation have been shown to effectively capture long-range dependencies [20, 21].

1.2.3 Graph neural network–based spatial relationship modeling methods

Graph neural networks have been widely recognized as effective tools for modeling complex relational structures in spatial interaction data [22, 23]. Compared with sequence-based models, graph structures enable more natural representations of multi-relational interactions and enhance feature learning through neighborhood aggregation mechanisms [24]. As illustrated in Figure 1, a typical user–point-of-interest bipartite graph structure is adopted, in which users are connected exclusively to points of interest, and no direct connections are established among user nodes. Similarly, point-of-interest nodes are only connected to user nodes, without direct links to other points of interest.

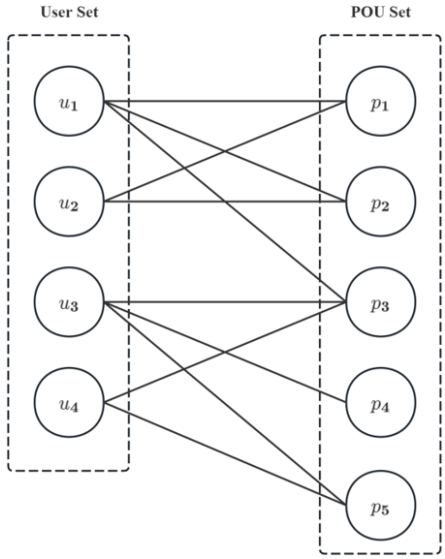


Figure 1. User–point-of-interest bipartite graph

The graph convolutional network proposed by Kipf et al. [25] laid the foundational framework for this research direction. Subsequently, the light graph convolutional network improved computational efficiency through structural simplification [26]. Building upon these advances, a variety of graph structure extensions have been introduced, including geographical proximity graphs and semantic graphs [27-29], as well as graph-based sequential modeling approaches such as the session-based recommendation with graph neural networks [30, 31]. In addition, hypergraph-based methods have been developed to capture higher-order relational dependencies, such as disentangled contrastive hypergraph learning [32, 33]. Collectively, these approaches have significantly enhanced the capability of spatial relationship modeling.

1.2.4 Disentangled representation learning and contrastive learning methods

With the continuous improvement of model capacity, research focus has gradually shifted toward semantic disentanglement. Disentangled representation learning aims to map different underlying factors into independent latent subspaces, thereby improving representation quality and interpretability [34, 35]. As illustrated in Figure 2, the disentanglement task in point-of-interest recommendation is conceptually presented, where factors originally entangled within a shared embedding space are separated into distinct subspaces. Specifically, the blue component represents geographical preference, the orange component corresponds to sequential preference, and the purple component denotes collaborative preference.

Representative approaches, such as the macro-micro disentangled variational auto-encoder [36] and the disentangled graph convolutional network [37], model user preferences through multi-dimensional latent representations. Building upon these methods, a disentangled dual-graph

framework for point-of-interest recommendation (DisenPOI) further decomposes user behavior into multiple preference factors, enabling finer-grained modeling of spatial interactions [38]. In parallel, contrastive learning has been widely adopted to enhance representation robustness by constructing positive and negative sample pairs to constrain the embedding space [39, 40]. For example, self-supervised graph learning [41] introduces graph augmentation strategies to formulate contrastive learning tasks, while extremely simple graph contrastive learning [42] reduces computational complexity through noise-based perturbation mechanisms. These approaches have been shown to effectively reduce information redundancy across different semantic subspaces [34].

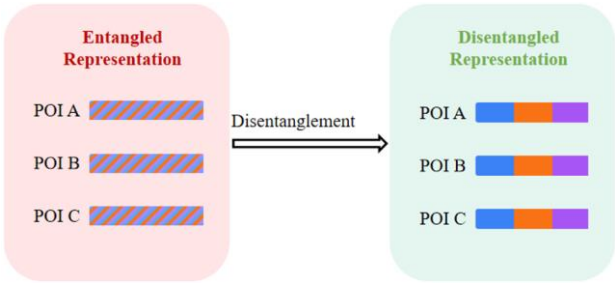


Figure 2. Conceptual illustration of disentangled representation learning

1.3 Research content

Spatial behavior is characterized by pronounced regional concentration and temporal regularity, while heterogeneous dependencies on spatial, temporal, and preference-related factors are observed across different users [43-45]. In this context, the following research problems are addressed:

(a) Existing graph-based collaborative filtering methods insufficiently exploit interaction intensity and positional information in spatial scenarios;

(b) Disentangled representation learning models lack personalized mechanisms during the fusion stage, limiting their ability to capture user-specific differences;

(c) Collaborative signals are inadequately modeled in multi-view learning frameworks, thereby affecting the completeness of learned representations.

To address these challenges, the following contributions are made:

(a) A Spatial Frequency–aware Light Graph Convolutional Network (SF-LightGCN) model integrating spatial interaction frequency and positional encoding is proposed. By incorporating weighted graph structures and spatial encoding mechanisms, the expressive capacity of baseline representations is enhanced;

(b) A three-view disentangled representation learning model with adaptive fusion, termed DisenPOI++, is developed. A collaborative view is introduced, and a user-level dynamic fusion mechanism is designed to improve the capability of personalized representation learning;

(c) Extensive experiments are conducted on real-world datasets, including Foursquare and Gowalla. The effectiveness of the proposed models is validated through comprehensive comparative evaluations and ablation studies.

1.4 Organization of the study

The remainder of the study is organized into six chapters, structured as follows:

Chapter 1: Introduction. The research background and significance are presented, the related work is systematically reviewed, and the research problems and primary contributions are clearly defined.

Chapter 2: Theoretical Foundations and Technical Background. The problem formulation of spatial behavior modeling is introduced, along with key methodologies, including graph convolutional networks, disentangled representation learning, and contrastive learning.

Chapter 3: SF-LGCN Model. A spatial representation learning model integrating spatial interaction intensity and positional encoding is proposed. The model architecture, information modeling strategy, and optimization process are described in detail.

Chapter 4: DisenPOI++ Model. A multi-view disentangled representation learning model with adaptive fusion is developed. The three-view disentanglement mechanism, gated fusion strategy, and training procedure are elaborated.

Chapter 5: Experiments and Analysis. Experimental settings are introduced, and model performance is evaluated through comparative experiments, ablation studies, and parameter sensitivity analysis.

Chapter 6: Conclusion and Future Work. The research findings are summarized, existing limitations are analyzed, and potential future research directions are discussed.

2. THEORETICAL FOUNDATIONS AND TECHNICAL BACKGROUND

In this chapter, fundamental theories and key methodologies closely related to the present research are introduced, providing the theoretical basis for the design and analysis of the proposed models in subsequent sections.

2.1 Problem formulation of spatial behavior modeling

2.1.1 Notation and definitions

To facilitate subsequent formulation and analysis, the fundamental notations and definitions used in point-of-interest recommendations are first introduced. Let $U=\{u_1,u_2,\ldots,u_M\}$ denote the set of users, where M represents the total number of users. Let $P=\{p_1,p_2,\ldots,p_N\}$ denote the set of points of interest, where N is the total number of points of interest. Each point of interest p_I is associated with a geographical coordinate (lat_I,lon_i), representing its latitude and longitude, respectively. The interaction between users and points of interest is represented by an interaction matrix $R\in\mathrm{R}^{M\times N}$. In implicit feedback scenarios, each element $r_{ui}\in\{0,1\}$ indicates whether user u has visited the point of interest p_i. Furthermore, the check-in frequency of user u at p_I is denoted as $f(u,p_i)$, representing the total number of visits made by user u to p_i during the observation period. It is defined that $f(u,p_i)\geq 1$ if and only if $r_{ui}=1$. The set of points of interest visited by user u constitutes the positive sample set, denoted as $P_u=\{p_i\in P \mid r_{ui}=1\}$. When temporal information is considered, the historical check-in sequence of user u is represented as an ordered sequence $S_u=\left\{p_1^u,p_2^u,\ldots,p_{|S_u|}^u\right\}$, where p_t^u denotes the point of interest visited by user u at the t-th check-in, and $|S_u|$ represents the total number of check-ins for user. The check-in sequence can be further segmented into multiple sessions

based on temporal intervals, denoted as $\{s_1^u, s_2^u, ...\}$, where each session s_k^u consists of a consecutive subsequence of check-ins within a specific time window.

2.1.2 General top-K recommendation and next point-of-interest recommendation

Point-of-interest recommendations can be broadly categorized into two paradigms depending on whether temporal information in user check-in behavior is explicitly considered: general Top-K recommendation and next point-of-interest recommendation. The two proposed models correspond to these respective paradigms.

(a) General top-K point-of-interest recommendation

The objective of general top-K recommendation is to identify the K points of interest that are most likely to be of interest to a user from the entire set of candidate points of interest, based on historical check-in data. In this paradigm, the temporal order of check-in behavior is not explicitly considered; instead, all historical interactions are treated as an unordered set, with emphasis placed on modeling the user's global and static preferences. Formally, given the historical interaction set P_u of user u, along with auxiliary information (e.g., check-in frequency $f(u;)$, and geographical coordinates of points of interest), a scoring function F_{score} is learned to estimate the preference score for each candidate point of interest $p_i \notin P_u$, as defined in Eq. (1):

$$\hat{y}_{ui} = F_{score}(u, p_i \mid P_U, C) \tag{1}$$

where, C denotes the set of auxiliary information. Subsequently, all candidate points of interest are ranked in descending order according to the predicted preference scores $\hat{y}_{ui}$, and the top K points of interest are selected as the recommendation result $TopK(u)=\{p_{i_1}, p_{i_2}, ..., p_{i_K}\}$. A schematic illustration of this recommendation paradigm is provided in Figure 3.

General top-K recommendation represents one of the most classical paradigms in recommendation systems. Its core technical foundation lies in collaborative filtering, where user preferences and point-of-interest characteristics are inferred from the user–point-of-interest interaction matrix. Both light graph convolutional network and the proposed SF-LGCN model belong to this paradigm. The primary advantage of this approach lies in its relatively simple model structure and its ability to fully exploit global interaction information. However, a key limitation is that the temporal dynamics of check-in behavior are neglected, resulting in an inability to capture the evolution of user preferences over time as well as the transition dependencies between consecutive check-ins.

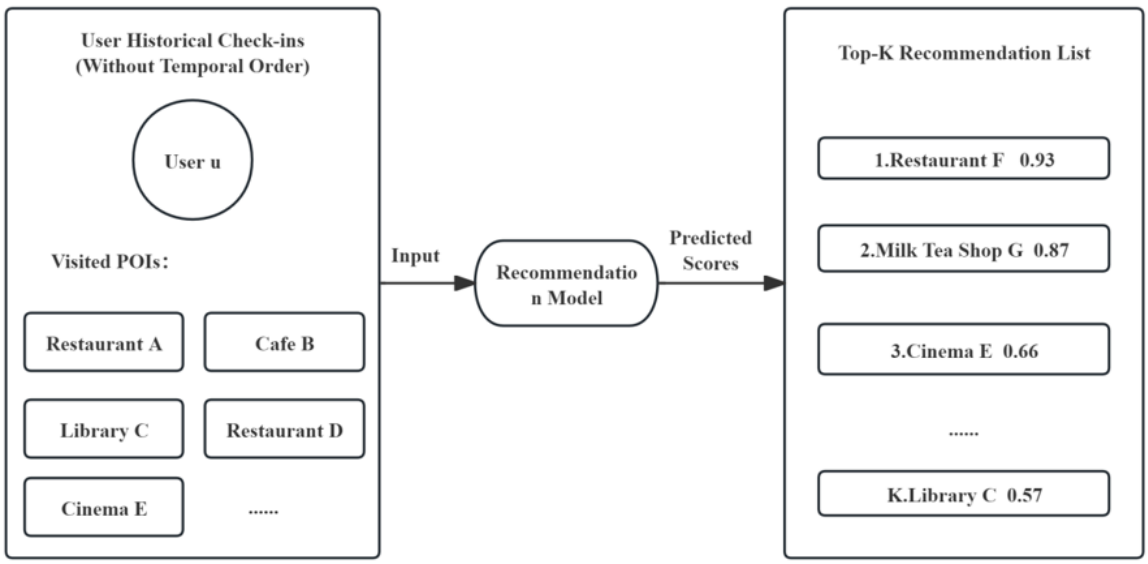


Figure 3. Illustration of general top-K point-of-interest recommendation

(b) Next point-of-interest recommendation

Next point-of-interest recommendation constitutes a more fine-grained recommendation paradigm, in which the objective is to predict the most likely point of interest to be visited by a user at the next time step, given the historical check-in sequence. In contrast to general top-K recommendation, the temporal order of check-in behavior is explicitly incorporated, with emphasis placed on modeling dynamic user preferences and sequential transition patterns. Formally, given the historical check-in sequence of user u, denoted as $S_u=\{p_1^u, p_2^u, ..., p_{|S_u|}^u\}$, or the current session s_k^u, together with relevant contextual information, a sequential recommendation function F_{next} is learned to predict the next point of interest to be visited, as defined in Eq. (2):

$$\hat{p}_{|S_u|+1}^u = F_{next}(u, S_u, C) \tag{2}$$

In practical implementations, F_{next} assigns a probability or preference score to each candidate point of interest, and the top K points of interest with the highest scores are returned as the recommendation list. An illustrative example of this paradigm is provided in Figure 4.

Compared with general top-K recommendations, next point-of-interest recommendations enable the modeling of richer behavioral patterns. For instance, sequential transition regularities such as a tendency to visit a cinema after lunch can be captured, along with the dynamic evolution of user preferences over time. Models such as DisenPOI and the proposed DisenPOI++ belong to this paradigm, where session graphs are constructed to model transition relationships among points of interest as well as geographical proximity relationships.

(c) Relationship between the two paradigms

It should be emphasized that the two paradigms described

above are not mutually exclusive but instead provide complementary perspectives for modeling user preferences. General top-K recommendation focuses on identifying the types of points of interest that a user prefers in a global and static sense, whereas next point-of-interest recommendation aims to predict the point of interest that a user is most likely to visit at a specific moment. Within the present study, SF-LGCN enhances information utilization in point-of-interest recommendation from the perspective of global collaborative filtering, while DisenPOI++ achieves multi-factor personalized modeling from the perspective of sequential disentangled representation learning. Consequently, both approaches contribute complementary improvements to point-of-interest recommendations from different modeling levels.

2.2 Graph convolutional networks

2.2.1 Fundamentals of graph neural networks

Graph neural networks constitute a class of deep learning models specifically designed for processing graph-structured data [46]. Traditional deep learning models are primarily developed for data in Euclidean space with regular grid structures, such as images and speech signals, where translation invariance is typically assumed and neighborhood structures are fixed. In contrast, a substantial amount of real-world data is represented in non-Euclidean domains, including social networks and knowledge graphs, where node connections are irregular and topological structures are highly complex. Graph neural networks are well suited to handle such data. In point-of-interest recommendation systems, the interaction relationships between users and points of interest naturally form a bipartite graph structure, which provides a suitable foundation for the application of graph neural network-based methods in recommendation tasks. Formally, a graph G=(V,E) consists of a set of nodes V and a set of edges E. The structural information of the graph can be represented by an adjacency matrix $A \in R^{|V|\times|V|}$. The core mechanism of graph neural networks is message passing, which can be summarized as follows: information is aggregated from neighboring nodes (message aggregation), and node representations are subsequently updated by integrating the aggregated information with the node's own features (state update). Figure 5 illustrates the information propagation process in graph convolutional networks. Among various graph neural network architectures, the graph convolutional network [25] is one of the most representative models. The layer-wise propagation rule of the graph convolutional network is defined as:

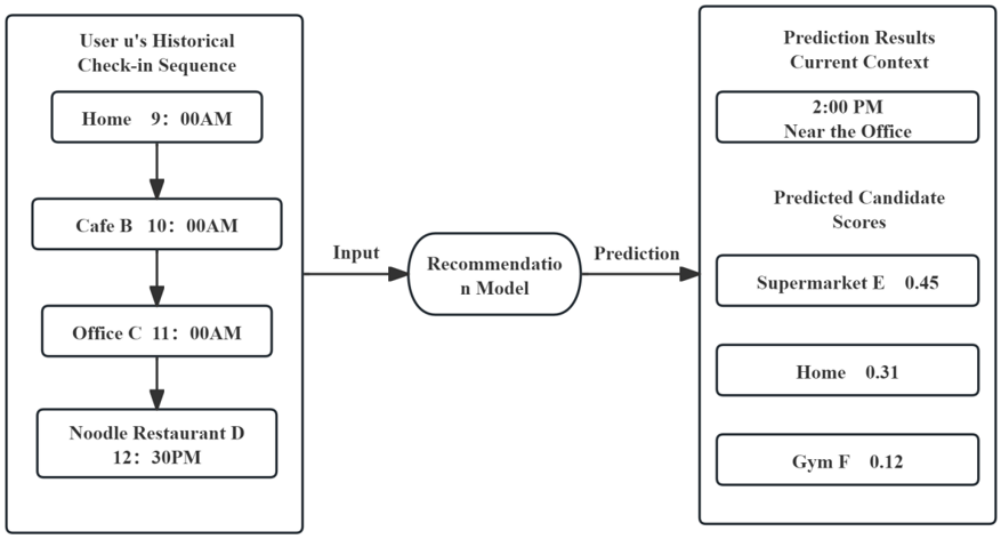


Figure 4. Illustration of next point-of-interest recommendation

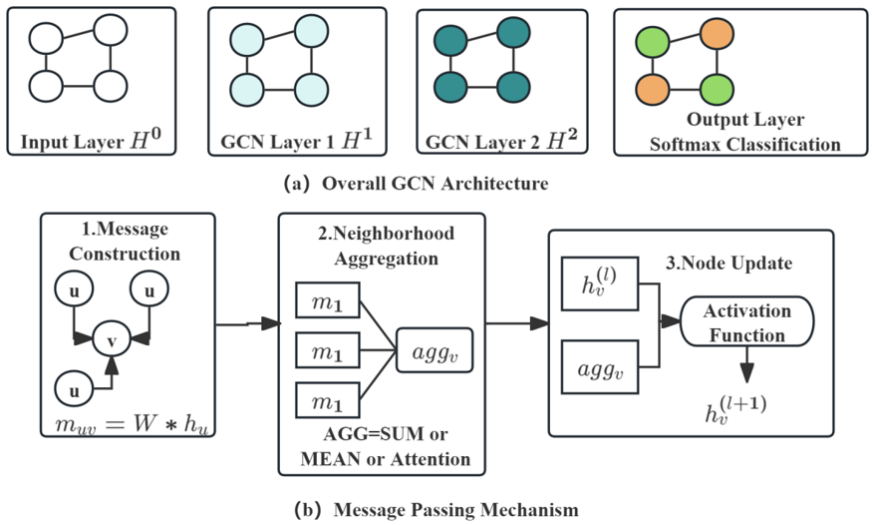


Figure 5. Message passing mechanism in graph convolutional networks

$$H^{(l+1)}=\sigma\left(\widetilde{D}^{-\frac{1}{2}}\widetilde{A}\widetilde{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}\right) \tag{3}$$

where, $\widetilde{A}$=A+I denotes the adjacency matrix with self-loops added, $\widetilde{D}$ is the corresponding degree matrix, $H^{(l)}$ represents the node embedding matrix at the l-th layer, $W^{(l)}$ is the learnable weight matrix, and σ denotes a nonlinear activation function.

2.2.2 Lightweight graph convolutional network architecture

He et al. [26] observed that, when graph convolutional networks are applied to recommendation systems, the feature transformation matrices W and nonlinear activation functions σ do not contribute to performance improvement in collaborative filtering tasks. Instead, these components may degrade performance due to increased model complexity. Based on this insight, the light graph convolutional network was proposed, with the design principle of retaining only the most essential operation in graph convolutional networks for recommendation tasks, namely neighborhood aggregation. Accordingly, the graph convolution operations in the light graph convolutional network are simplified as follows:

$$e_u^{(l+1)}=\sum_{i\in N_u}\frac{1}{\sqrt{|N_u|}\sqrt{|N_i|}}e_i^{(l)} \tag{4}$$

$$e_i^{(l+1)}=\sum_{u\in N_i}\frac{1}{\sqrt{|N_i|}\sqrt{|N_u|}}e_u^{(l)} \tag{5}$$

where, $e_u^{(l)}$ and $e_i^{(l)}$ denote the embeddings of user u and item i at the l-th layer, respectively, and N_u and N_i represent the neighbor sets of user u and item i. After L layers of graph convolution, the final embeddings are obtained by aggregating representations from all layers through weighted summation:

$$e_u=\sum_{l=0}^{L}\alpha_l\cdot e_u^{(l)} \tag{6}$$

$$e_i=\sum_{l=0}^{L}\alpha_l\cdot e_i^{(l)} \tag{7}$$

where, α_l denotes the weighting coefficient for each layer, which is typically set as $\alpha_l=1/(L+1)$. Owing to its simplified yet effective design, the light graph convolutional network has become a benchmark model in graph-based collaborative filtering. It also serves as the foundational architecture for the SF-LGCN model introduced in Chapter 3.

2.2.3 Point-of-interest recommendation based on graph convolutional networks

The application of graph neural networks in recommendation systems has evolved from complex architectures to more simplified and efficient designs. Early approaches, such as the neural graph collaborative filtering [47], introduced graph convolution into user–item bipartite graphs, where feature transformation and nonlinear activation were retained at each layer. Although performance improvements were achieved, these models suffered from high computational complexity and a tendency toward overfitting. Subsequently, the light graph convolutional network [26] demonstrated through empirical analysis that such operations contribute minimally to collaborative filtering performance. As a result, only the neighborhood aggregation mechanism was preserved, leading to both structural simplification and improved performance. This advancement highlights that the core of recommendation tasks lies in effectively leveraging interaction structures rather than relying on complex feature transformations.

Building upon this foundation, subsequent studies have further enhanced the expressive capacity of graph-based models. The graph sample and aggregate method [48] improves scalability in large-scale scenarios through sampling-based aggregation, while graph attention networks [49] enable differentiated modeling of neighboring nodes via attention mechanisms. In the context of spatial recommendation, additional efforts have been made to capture complex spatiotemporal dependencies. For example, multi-graph structures and geographically aware convolutional operations have been employed to model spatial periodicity and interaction patterns [50]. Moreover, multi-granularity contrastive learning strategies have been introduced to enhance the representations of users and points of interest [51]. These advancements provide a solid technical foundation for integrating check-in frequency and geographical information in the proposed approach.

2.3 Disentangled representation learning

2.3.1 Fundamental concepts of disentangled representation learning

Disentangled representation learning aims to learn data representations in which independent generative factors are separated into distinct dimensions or subspaces within the representation space [35]. An ideal disentangled representation is expected to satisfy the following property: each subspace is sensitive to only one underlying factor of variation, such that changes in a specific factor affect only the corresponding subspace while leaving other subspaces unaffected. For example, consider that an observed data instance x is generated by K independent latent factors $\{z_1,z_2,...,z_K\}$. The objective of disentangled representation learning is to learn an encoder function f, defined as:

$$f(x)=\left[f_1(x),f_2(x),...,f_K(x)\right] \tag{8}$$

where, $f_k(x)$ is associated exclusively with the k-th generative factor z_k. As illustrated in Figure 6, a typical application of disentangled representation learning in recommendation systems involves decomposing user preferences into multiple independent aspects, such as brand preference, price sensitivity, and functional requirements. This decomposition enables more fine-grained user profiling and recommendation [36].

2.3.2 Multi-factor disentanglement in point-of-interest recommendation

In point-of-interest recommendation scenarios, the selection of target locations by users is typically influenced by multiple factors, which imposes higher requirements on model expressiveness and representation capacity. In recent years, disentangled representation learning has been widely adopted to address such multi-factor modeling challenges. Its core principle lies in decomposing user preferences into multiple mutually independent semantic dimensions, each

corresponding to a latent influencing factor [52]. In spatial recommendation tasks, three primary types of disentangled factors are commonly identified. First, geographical preference reflects the spatial concentration of user activities and represents a key characteristic distinguishing spatial recommendation from traditional recommendation systems. Second, sequential preference captures temporal dependencies in user behavior, enabling the modeling of dynamic transition patterns in consecutive visits. Third, collaborative preference is derived from collective behavioral patterns across users and is used to uncover latent similarities among different users.

Existing methods exhibit varying emphases in disentangled modeling. For instance, DisenPOI [38] primarily focuses on the separation of geographical and sequential factors. More recent studies have further incorporated collaborative signals and introduced adaptive mechanisms to achieve dynamic multi-factor fusion, thereby enhancing both the expressive capacity of the model and the level of personalization.

2.4 Contrastive learning

2.4.1 Fundamental principles of contrastive learning

Contrastive learning is a self-supervised learning paradigm in which representations are learned by constructing positive and negative sample pairs, enabling the encoder to distinguish between similar and dissimilar instances [39]. Unlike traditional supervised learning, which relies on manually annotated labels, contrastive learning derives supervision signals directly from the intrinsic structure of data or through data augmentation strategies. As a result, it has achieved significant success across various domains, including image recognition, natural language processing, and graph representation learning. Figure 7 shows the application of contrastive learning in image recognition.

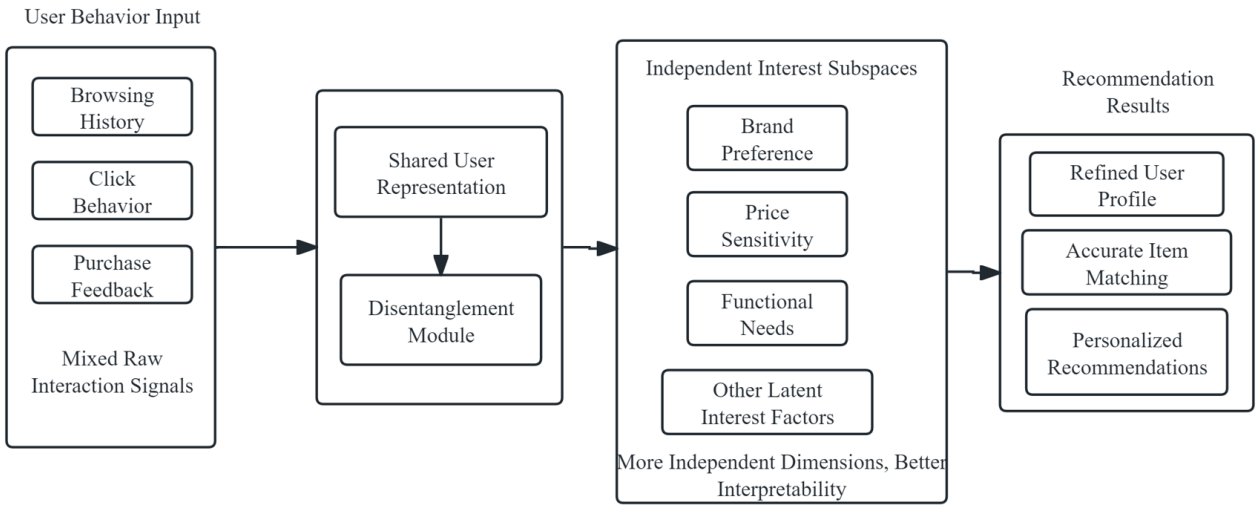


Figure 6. Role of disentangled representation learning in recommendation systems

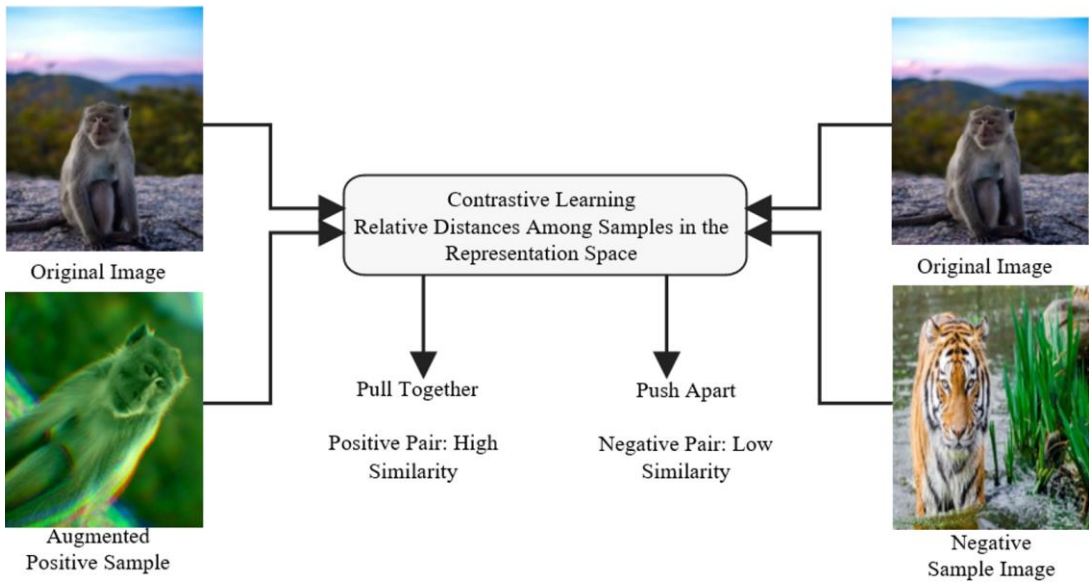


Figure 7. Application of contrastive learning in image recognition

The general framework of contrastive learning is defined as follows. Given an anchor sample x, a semantically similar sample x^+ forms a positive sample pair, while a semantically dissimilar sample x^- constitute negative sample pairs. The objective of contrastive learning is to train an encoder $f(\cdot)$ such that the representations of the anchor and its positive sample are pulled closer in the embedding space, while those of the anchor and all negative samples are pushed farther apart. This process can be interpreted as learning relative similarities among samples within a predefined metric space. The

construction of positive sample pairs typically relies on transformations applied to the original data. In computer vision, common data augmentation techniques—such as random cropping, rotation, and color jittering—are employed to generate multiple views of the same image as positive pairs. In natural language processing, semantically consistent sentence pairs can be constructed through techniques such as back-translation and token masking. In graph representation learning, positive pairs may be formed by generating different subgraph views of the same node or by applying perturbations to the graph structure. Negative samples are typically drawn either from other samples within the current mini-batch or from a memory bank containing historical representations.

The most commonly used loss function in contrastive learning is the information noise-contrastive estimation loss [53], defined as:

$$L_{CL}=-\log\frac{\exp(\mathrm{sim}(z,z^{+})/\tau)}{\exp(\mathrm{sim}(z,z^{+})/\tau)+\sum_{j=1}^{K}\exp\left(\mathrm{sim}\left(z,z_{j}^{-}\right)/\tau\right)} \tag{9}$$

where, z, z^{+} and z_{j}^{-} denote the representations of the anchor sample, the positive sample, and the j-th negative sample, respectively. The function $\mathrm{sim}(\cdot,\cdot)$ represents a similarity measure, typically implemented as cosine similarity. The parameter τ is a temperature hyperparameter, and K denotes the number of negative samples. From an intuitive perspective, this loss function encourages the similarity of positive sample pairs to dominate among all candidate samples (i.e., one positive sample and K negative samples), thereby maximizing the probability that the positive pair is correctly identified. By continuously pulling semantically consistent samples closer and pushing inconsistent samples apart, contrastive learning enables the extraction of invariant features that are strongly correlated with semantic information, while suppressing irrelevant variations introduced by data augmentation or noise. Consequently, the learned representations exhibit strong generalization and discriminative capabilities, making them particularly suitable for pretraining or as auxiliary objectives in downstream tasks.

2.4.2 Applications of contrastive learning in recommendation systems

In recommendation systems, data sparsity has long posed a significant challenge to the learning of high-quality representations. Contrastive learning addresses this limitation by constructing self-supervised training signals, enabling latent structural information to be extracted directly from the data and thereby improving both the robustness and generalization capability of learned representations [54]. Self-supervised graph learning [41] represents one of the earliest attempts to incorporate contrastive learning into graph-based collaborative filtering. In this approach, multiple augmented views of the user–item interaction graph are generated through stochastic perturbations, such as node dropping and edge perturbation. The resulting node representations from different views are treated as positive sample pairs for contrastive learning, thereby enhancing the model's robustness to structural perturbations. Subsequently, the extremely simple graph contrastive learning [42] introduced a simplified framework in which contrastive views are constructed by injecting noise directly into the embedding space, thus avoiding complex graph augmentation procedures. Empirical results demonstrate that the effectiveness of contrastive learning primarily depends on representation consistency rather than the specific form of augmentation.

In the context of disentangled representation learning, contrastive learning plays an even more critical role. Since explicit supervision signals for different latent factors are typically unavailable, contrastive learning provides indirect constraints by constructing proxy positive and negative samples, thereby facilitating the separation of different semantic subspaces and reducing information redundancy among representations. Furthermore, the application of contrastive learning in recommendation systems has continued to expand. On the one hand, it can be integrated with pretraining strategies to learn general-purpose representations from unlabeled data. On the other hand, it can be incorporated as an auxiliary loss and jointly optimized with the main task to improve overall representation quality. In summary, contrastive learning has evolved from a technique for alleviating data sparsity to a versatile framework for enhancing representation robustness and supporting disentangled modeling. This evolution provides essential technical support for the proposed models, such as DisenPOI++.

3. PROPOSED MODEL INTEGRATING SPATIAL INTERACTION FREQUENCY AND GEOGRAPHICAL INFORMATION

The light graph convolutional network has demonstrated strong performance in collaborative filtering tasks, where its simplified neighborhood aggregation mechanism enables effective modeling of user–point-of-interest interaction relationships. However, when directly applied to point-of-interest recommendation scenarios, two notable limitations remain. First, variations in interaction intensity between users and points of interest are not explicitly considered. Second, the geographical information associated with points of interest is not incorporated into the model. To address these limitations, an enhanced model, termed SF-LGCN, is proposed. Without introducing additional structural complexity, the model is improved from two complementary perspectives: edge weight modeling and node representation.

3.1 Overall model architecture

The overall architecture of the proposed SF-LGCN model is illustrated in Figure 8.

The proposed model is composed of three main components:

(a) Embedding layer

User embeddings are retained as standard learnable representations, while point-of-interest embeddings are enhanced by incorporating geographical information. Specifically, a geographical encoder is employed to encode latitude and longitude coordinates, and the resulting geographical features are integrated into the initial point-of-interest embeddings through learnable parameters.

(b) Graph convolutional layer (light graph convolutional network)

Multi-layer propagation is performed over the user–point-of-interest bipartite graph using the light graph convolutional network framework. In addition, spatial interaction frequency (i.e., check-in frequency) is incorporated as edge weights to participate in the message passing process.

(c) Prediction layer

Preference scores are computed using the inner product of

the final user and point-of-interest embeddings. Based on these scores, the top-K points of interest are ranked and returned as the recommendation results.

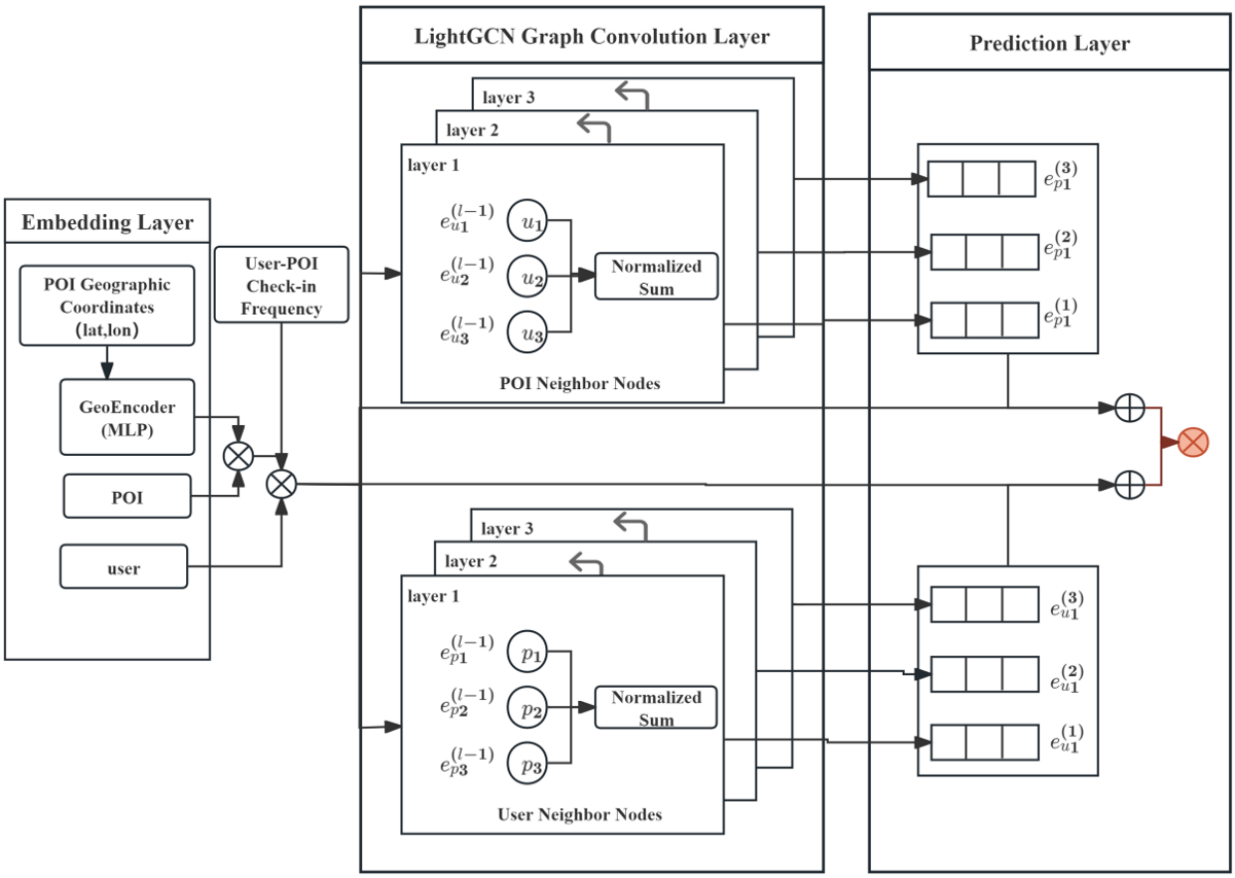
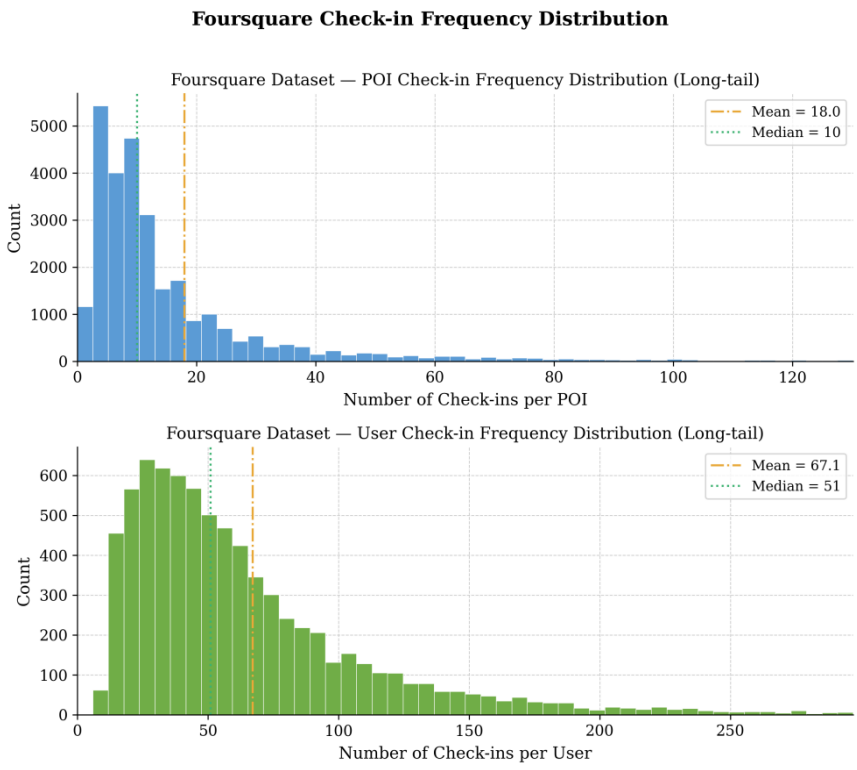


Figure 8. Architecture of the Spatial Frequency–aware Light Graph Convolutional Network (SF-LightGCN) model

3.2 Construction of frequency-weighted graphs

3.2.1 Analysis of frequency characteristics

The distribution of check-in frequency is characterized by a pronounced long-tail pattern (Figure 9), in which the majority of interactions occur only once, while a small number of high-frequency interactions reflect stable and persistent user preferences.



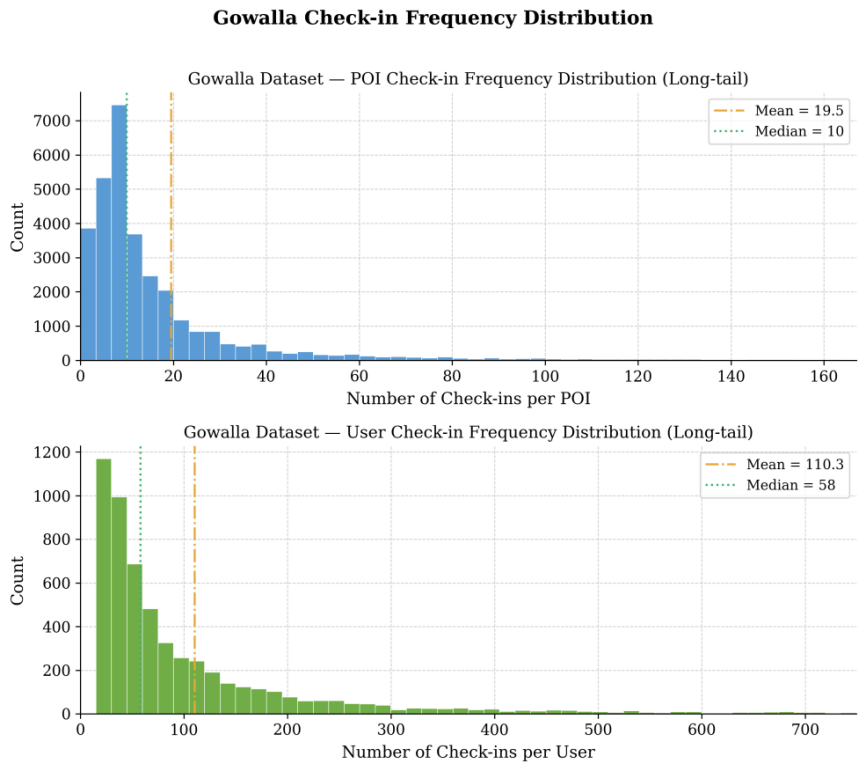


Figure 9. Histograms of check-in frequency distributions for users and points of interest in the Foursquare and Gowalla datasets

3.2.2 Edge weight definition

The check-in frequency is directly incorporated as the edge weight and defined as:

$$w_{up}=f(u,p) \tag{10}$$

where, $f(u,p)$ denotes the historical check-in frequency of user u at point-of-interest p. To prevent highly connected nodes from exerting excessive influence during the message passing process, symmetric normalization based on node weighted degrees is further applied. The normalized edge weight is defined as:

$$\widehat{w}_{up}=\frac{w_{up}}{\sqrt{D_u D_p}}=\frac{f(u,p)}{\sqrt{D_u D_p}} \tag{11}$$

$$D_u=\sum_{p\in N(u)} w_{up} \tag{12}$$

$$D_p=\sum_{u\in N(p)} w_{up} \tag{13}$$

where, D_u denotes the weighted degree of user node u, and D_p denotes the weighted degree of the point-of-interest node p. Through this design, the intensity of user preferences encoded by check-in frequency is effectively preserved, while numerical stability is maintained during graph convolutional propagation. Consequently, more reliable structural signals are provided for learning user and point-of-interest representations.

3.2.3 Construction of the weighted adjacency matrix

In the SF-LGCN model, check-in frequency is not merely treated as an auxiliary statistical feature; instead, it is directly incorporated into both the construction of the user–point-of-interest bipartite graph and the graph convolutional propagation process. This design enables more accurate modeling of the relationships between user preferences and point-of-interest characteristics. After obtaining the edge weights, a weighted user–point-of-interest interaction matrix R^w is first constructed, where each element is defined as $r^w_{up}=w_{up}$. Subsequently, this interaction matrix is embedded into the full adjacency matrix of the user–point-of-interest bipartite graph, resulting in a weighted adjacency matrix A^w. Following this, the standard symmetric normalization strategy adopted in the light graph convolutional network is applied to the weighted adjacency matrix, yielding the normalized matrix used for graph convolutional propagation:

$$\widehat{A}^w=D^{-\frac{1}{2}}A^wD^{-\frac{1}{2}} \tag{14}$$

where, D denotes the diagonal degree matrix computed from A^w, and each diagonal element corresponds to the sum of the edge weights connected to the respective node. Through this normalization process, the message passing mechanism accounts not only for differences in interaction frequency between users and points of interest but also mitigates scale bias introduced by high-frequency nodes. As a result, more balanced information propagation across nodes is achieved.

3.3 Geographical information integration module

3.3.1 Encoding of latitude and longitude coordinates

Each point of interest is associated with a set of geographical coordinates. To map the low-dimensional latitude and longitude coordinates into the same latent space as the point-of-interest embeddings, a multilayer perceptron is employed as a geographical encoder. The normalized coordinate vector is transformed through nonlinear operations

to obtain the geographical embedding of each point of interest. The encoding process is defined as:

$$g_p = tanh\left(W_2 \cdot Dropout\left(ReLU\left(W_1 x_p + b_1\right)\right) + b_2\right) \tag{15}$$

$$\boldsymbol{x}_p = [lat_p, lon_p] \in R^2 \tag{16}$$

where, x_p denotes the latitude–longitude coordinate vector associated with point of interest p. The terms $W_1 \in R^{32\times 2}$, $b_1 \in R^{32}$, $W_2 \in R^{d\times 32}$, and $b_2 \in R^d$ are learnable parameters, and d denotes the embedding dimension. Through this encoding process, the geographical embedding $g_p \in R^d$ is obtained. To further enhance nonlinear expressive capacity and improve training stability, a Dropout layer is introduced after the hidden layer, and a hyperbolic tangent (Tanh) activation function is applied at the output layer. In addition, to mitigate the influence of scale differences in the raw coordinate values, z-score normalization is applied to the point-of-interest coordinates during data preprocessing, and the normalized coordinates are subsequently fed into the geographical encoder.

3.3.2 Learnable weight fusion mechanism

After obtaining the geographical embeddings of points of interest, a globally shared learnable parameter is introduced to weight the contribution of geographical information, which is then directly added to the base point-of-interest identifier embedding. This design preserves structural simplicity while enabling adaptive control over the influence of geographical information within the overall representation. For a given point of interest p, the initial enhanced embedding is defined as:

$$e_p^{(0)} = e_p^{base} + \alpha_{geo} g_p \tag{17}$$

where, $e_p^{base} \in R^d$ denotes the base identifier embedding of point-of-interest p; $g_p \in R^d$ represents the geographical embedding generated by the geographical encoder; and $\alpha_{geo} \in R$ is a globally shared learnable fusion coefficient. In the implementation, α_{geo} is treated as a trainable scalar parameter, initialized to 0.3 and updated automatically during training via backpropagation. In SF-LGCN, geographical information is not incorporated as an independent feature applied only at the prediction stage. Instead, it is integrated into the point-of-interest representation prior to graph convolutional propagation, ensuring that geographical preference signals can continuously influence higher-order neighborhood aggregation. After the fusion process, the enhanced point-of-interest embeddings are used together with user embeddings as the initial input to the light graph convolutional network model. Specifically, user embeddings remain unchanged, while point-of-interest embeddings are updated to $e_p^{(0)}$. Subsequently, multi-layer propagation is performed over the weighted user–point-of-interest bipartite graph, allowing geographical information to be further integrated into the final node representations.

3.3.3 Design considerations for the fusion strategy

In point-of-interest recommendation tasks, several strategies have been developed for incorporating geographical information, including post-fusion, embedding concatenation, independent-channel modeling, and additive fusion. Post-fusion methods typically adjust prediction results at the final stage, which limits the effective utilization of spatial information during representation learning. Embedding concatenation introduces additional features but increases the dimensionality of representations and incurs additional computational cost. Independent-channel approaches require the construction of more complex architectures, thereby reducing model simplicity. By jointly considering expressive capacity and computational efficiency, an additive fusion strategy with a learnable weight is adopted. In this approach, the geographically encoded representation is directly added to the initial point-of-interest embedding. This method preserves the original embedding dimensionality while enabling the contribution of geographical information to be adaptively controlled through the parameter α_{geo}. As a result, geographical information is continuously incorporated during subsequent graph convolutional propagation, allowing spatial modeling capability to be effectively enhanced without compromising model simplicity.

3.4 Graph convolution and prediction layer

3.4.1 Light graph convolutional network-based graph convolution

After incorporating geographical information into the point-of-interest embeddings, the enhanced initial point-of-interest embeddings are concatenated with the initial user embeddings to form the overall node representation. Multi-layer light graph convolutional network propagation is then performed on the frequency-weighted user–point-of-interest bipartite graph. Let $E^{(l)}$ denote the nodes at the l-th layer. The graph convolutional propagation process is defined as:

$$E^{(0)} = \left[E_u^{(0)}; E_p^{(0)}\right] \tag{18}$$

$$E^{(l+1)} = \widehat{A}^w E^{(l)},\ l=0,1,\ldots,L-1 \tag{19}$$

where, $\widehat{A}^w$ denotes the frequency-weighted and symmetrically normalized adjacency matrix. At the node level, the propagation rules for user nodes and point-of-interest nodes can be expressed as:

$$e_u^{(l+1)} = \sum_{p\in N(u)} \frac{w_{up}}{\sqrt{D_u D_p}} e_p^{(l)} \tag{20}$$

$$e_p^{(l+1)} = \sum_{u\in N(p)} \frac{w_{up}}{\sqrt{D_u D_p}} e_u^{(l)} \tag{21}$$

After L layers of propagation, the representations from all layers are aggregated through mean pooling to obtain the final user and point-of-interest embeddings:

$$e_u^* = \frac{1}{L+1} \sum_{l=0}^{L} e_u^{(l)} \tag{22}$$

$$e_p^* = \frac{1}{L+1} \sum_{l=0}^{L} e_p^{(l)} \tag{23}$$

This design is consistent with the core principle of the light graph convolutional network, in which no additional feature

transformations or nonlinear activation functions are introduced during the propagation stage. Instead, higher-order collaborative relationships are effectively captured through a simplified neighborhood aggregation mechanism.

3.4.2 Preference prediction

After obtaining the final user and point-of-interest embeddings, the preference strength of a user for a candidate point of interest is computed using the inner product of their representations. The predicted preference score is defined as:

$$s_{up}=\left(e_u^*\right)^T e_p^* \tag{24}$$

3.5 Loss function and model optimization

The Bayesian personalized ranking loss [55] is adopted as the optimization objective. The core idea of the Bayesian personalized ranking is that, for any given user u, the predicted preference score of a visited (positive) point of interest should be higher than that of an unvisited (negative) point of interest. Let the training triplet set be defined as $D=\{(u,i,j)|i\in I_u^+, j\notin I_u^+\}$, where i denotes a positive point of interest that has been visited by user u, and j denotes a negative point of interest sampled from the unvisited set. The predicted preference scores are defined as:

$$s_{ui}=\left(e_u^*\right)^T e_i^*, s_{uj}=\left(e_u^*\right)^T e_j^* \tag{25}$$

According to the implementation, the Bayesian personalized ranking loss over the current batch is formulated as:

$$L_{BPR}=\frac{1}{|B|}\sum_{(u,i,j)\in B} softplus\left(s_{uj}-s_{ui}\right) \tag{26}$$

where, B denotes a training mini-batch, and $\mathrm{softplus}(x)=\log(1+e^x)$. This formulation is equivalent to the classical Bayesian personalized ranking objective $-\log\sigma(s_{ui}-s_{uj})$, but adopts the numerically more stable softplus function.

To mitigate overfitting, an L2 regularization term is further introduced. In the implementation, the regularization is applied to the initial embeddings of users, positive points of interest, and negative points of interest within the current batch, defined as:

$$L_{reg}=\frac{1}{2|B|}\left(\left\|e_u^{(0)}\right\|_2^2+\left\|e_i^{(0)}\right\|_2^2+\left\|e_j^{(0)}\right\|_2^2\right) \tag{27}$$

When the geographical information integration module is enabled, the initial point-of-interest embeddings $e_i^{(0)}$ and $e_j^{(0)}$ already incorporate the output of the geographical encoder as well as the influence of the learnable fusion coefficient α_{geo}. Consequently, the parameters of the geographical encoder and the fusion coefficient are also indirectly regularized through the regularization term. The overall loss function is therefore defined as:

$$L=L_{BPR}+\lambda L_{reg} \tag{28}$$

The Adam optimizer [56] is employed for parameter updates. According to the implementation, the optimizer is applied to all trainable parameters of the recommendation model. The main hyperparameter settings of the SF-LGCN model are summarized in Table 1.

Table 1. Key hyperparameter settings of the Spatial Frequency–aware Light Graph Convolutional Network (SF-LightGCN) model

Hyperparameter	Value
Embedding dimension d	64
Number of graph convolutional layers L	3
Learning rate	0.001
L2 regularization coefficient λ_r	1e-4
Hidden dimension of the geographical encoder d_{h}	32
Batch size	2048
Optimizer	Adam

3.6 Time and space complexity analysis

In this section, the time and space complexity of the SF-LGCN model (i.e., Algorithm 1) are analyzed. Compared with the original light graph convolutional network, the proposed model incorporates spatial interaction frequency as edge weights in the user–point-of-interest bipartite graph and further introduces a geographical coordinate encoder along with a learnable fusion coefficient α_{geo}. The incorporation of check-in frequency modifies only the numerical values of existing edges and does not introduce additional nodes, edges, or trainable parameters. Therefore, it does not affect the overall time or space complexity of the model. The additional computational overhead arises primarily from three components: the encoding of point-of-interest geographical coordinates, the fusion of geographical embeddings with point-of-interest representations, and the repeated computation of full-graph representations during both training and testing.

Algorithm 1. SF-LGCN model algorithm

Input: User check-in data; point-of-interest geographical coordinate data; number of graph convolutional layers K; embedding dimension d; learning rate η; regularization coefficient λ

Output: Predicted preference scores of users for candidate points of interest; model loss value; learned user and point-of-interest embeddings

1: All model parameters are initialized; user embedding matrix Eu and point-of-interest embedding matrix Ep are randomly initialized;

2: A geographical feature matrix is constructed based on point-of-interest latitude and longitude coordinates, and is input into the geographical encoder to obtain geographical representations G;

3: A frequency-weighted user–point-of-interest bipartite graph is constructed based on the check-in frequency matrix, yielding the weighted adjacency matrix Aw;

4: The normalized propagation matrix is computed according to Eq. (14);

5: **while** the model has not converged **do**

6: Training triplets (u,i,j) are randomly sampled;
7: **for** k = 1 to K **do**
8: The initial point-of-interest embeddings are computed by incorporating geographical information according to Eq. (17);
9: User and point-of-interest embeddings at the k-th layer are updated according to the graph convolutional propagation equations;
10: **end for**
11: Layer-wise embeddings are aggregated via mean pooling to obtain final user and point-of-interest embeddings;
12: Preference scores for positive and negative samples are computed according to Eq. (24);
13: The ranking loss is computed using the Bayesian personalized ranking loss function (Eq. (25));
14: The regularization term is incorporated to obtain the total loss (Eq. (27));
15: All model parameters are updated using the Adam optimizer;
16: **end while**
17: Preference scores for all candidate points of interest are computed according to the prediction equation;
18: The top-K recommendation results are generated by ranking the predicted scores in descending order;
19: **return**

3.6.1 Time complexity analysis

During a single full-graph forward propagation, the geographical coordinates of all points of interest are first processed by the geographical encoder to obtain geographical embeddings, which are then fused with the original point-of-interest embeddings using a learnable weighting mechanism. The time complexity of this stage is $O(Pd)$, where P denotes the number of points of interest and d denotes the embedding dimension. Subsequently, the fused node representations are propagated through L layers of the light graph convolutional network over the frequency-weighted sparse graph. At each layer, the core operation consists of a sparse matrix–dense matrix multiplication, with a time complexity of $O(|E|d)$, where $|E|$ denotes the number of edges. Therefore, the total time complexity of the L-layer graph convolution is $O(L|E|d)$. The overall time complexity of a single full-graph embedding computation can thus be expressed as:

$$O(Pd+L|E|d) \tag{29}$$

During the training phase, the Bayesian personalized ranking loss requires one forward propagation, and an additional geographical encoder computation over all point-of-interest coordinates is performed to construct positive and negative samples. Consequently, for a mini-batch of size B, the main computational cost can be summarized as:

$$O(L|E|d+2Pd+Bd) \tag{30}$$

Furthermore, considering the negative sampling process, approximately $|R|$ triplets are generated at the beginning of each epoch, incurring a sampling cost of $O(|R|)$. Therefore, the overall time complexity per training epoch can be expressed as:

$$O(|R|+[\frac{|R|}{B}]*(L|E|d+2Pd+Bd)) \tag{31}$$

From the above analysis, it can be observed that the dominant computational cost remains concentrated in the full-graph sparse propagation of the light graph convolutional network. The incorporation of frequency-based edge weighting only modifies edge values and does not alter the leading term $O(L|E|d)$. The additional overhead introduced by the geographical module is $O(Pd)$, while the learnable parameter α_{geo} is a scalar and thus introduces negligible computational cost. Consequently, the increase in time complexity relative to the original light graph convolutional network remains well controlled.

3.6.2 Space complexity analysis

In terms of space complexity, the model is required to maintain user embeddings, point-of-interest embeddings, the frequency-weighted sparse adjacency matrix, the point-of-interest coordinate matrix, and the parameters of the geographical encoder. The storage cost of user and point-of-interest embeddings is $O((U+P)d)$. The storage cost of the frequency-weighted sparse graph is $O(|E|)$. The storage cost of the point-of-interest coordinate matrix is $O(I)$. The geographical encoder consists of two linear transformation layers with a fixed hidden dimension; therefore, its parameter size scales linearly with d and can be expressed as $O(d)$. The parameter size of α_{geo} is $O(1)$. Accordingly, the overall space complexity of model parameters and static data can be expressed as:

$$O((U+P)d+|E|+P+d) \tag{32}$$

Overall, the space complexity of the SF-LGCN model with learnable spatial weights remains dominated by the embedding matrices and the sparse graph storage. The additional memory overhead introduced by the geographical module is relatively modest, thereby preserving scalability while enhancing the modeling of geographical information.

In this chapter, the limitation of the standard light graph convolutional network model in effectively utilizing auxiliary information for point-of-interest recommendation tasks has been addressed. To this end, the SF-LGCN model has been proposed. Improvements have been introduced from two perspectives: edge weight modeling and spatial information integration. While preserving the advantages of the light graph convolutional network—such as structural simplicity, a clear propagation mechanism, and high training efficiency—the proposed model further enhances the capability to capture user behavioral intensity (i.e., check-in frequency) and point-of-interest spatial attributes. As a result, more expressive and informative representations are obtained, providing effective support for subsequent improvements in recommendation performance. Building upon this foundation, the core model of this study, DisenPOI++, is introduced in the next chapter.

4. A SPATIAL PERCEPTION MODEL BASED ON MULTI-VIEW DISENTANGLEMENT AND ADAPTIVE FUSION

In the previous chapter, the spatial representation capability of the SF-LGCN model was enhanced to a certain extent by

incorporating check-in frequency and geographical information. However, multiple influencing factors are still jointly entangled within a unified representation, making it difficult to distinguish the contributions of different semantic sources to user behavior. To further improve representation capability, a spatial perception model based on multi-view disentangled representation learning and adaptive fusion, referred to as DisenPOI++, is proposed in this chapter.

4.1 Overall framework of the model

The overall architecture of the DisenPOI++ model is illustrated in Figure 10.

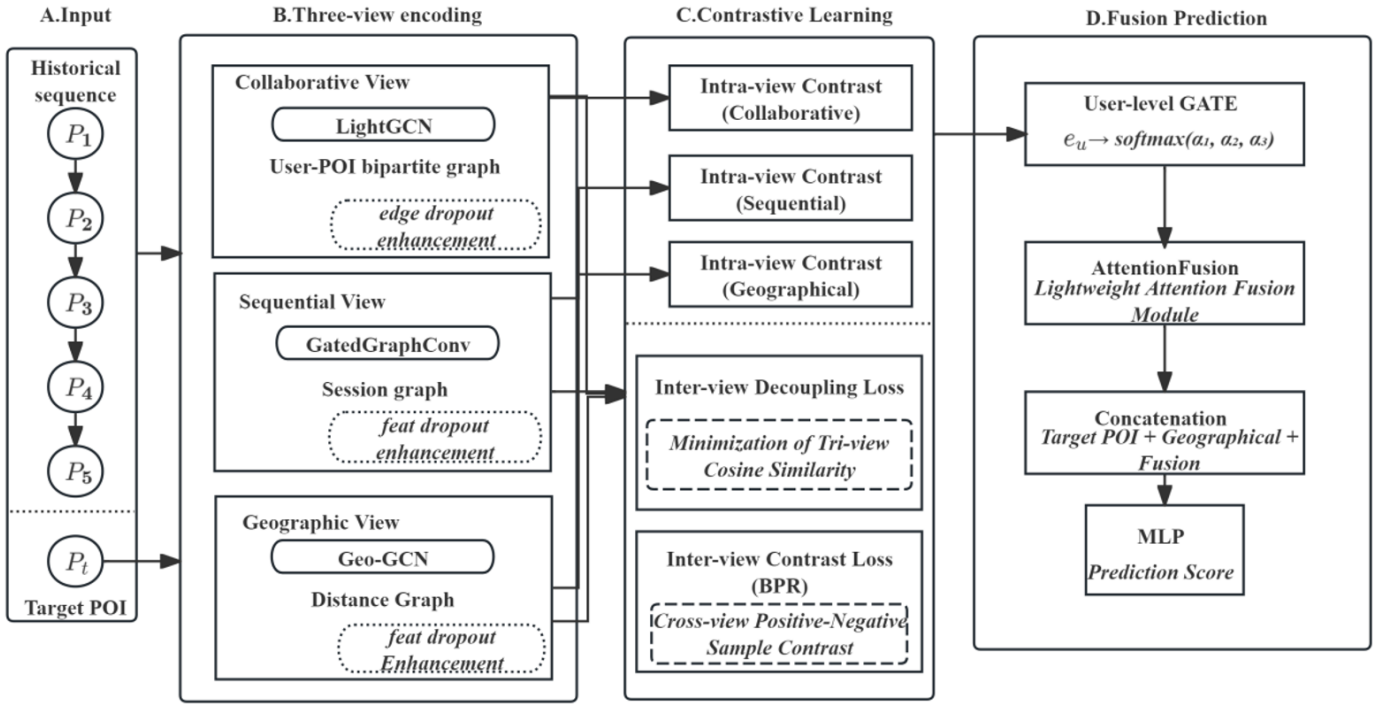


Figure 10. Overall architecture of the disentangled dual-graph framework for point-of-interest recommendation (DisenPOI)++ model

Multiple semantic views are constructed over a unified graph structure, and disentangled representation learning is employed to obtain distinct preference representations. These representations are subsequently integrated through an adaptive mechanism and utilized for spatial target prediction. The overall framework consists of three key components: multi-view construction, disentangled representation learning, and adaptive fusion.

4.2 Multi-view modeling

Spatial behavior is typically driven by multiple factors, and a single-view representation is insufficient to fully capture the underlying decision mechanisms governing user selection of spatial nodes. To address this limitation, a multi-view modeling framework is constructed from different semantic perspectives, in which complex behaviors are decomposed into multiple relatively independent representation spaces. In this manner, the expressive capacity of the model with respect to multi-source information is enhanced. Specifically, under a unified user–point-of-interest interaction framework, three complementary semantic views are constructed to characterize spatial structural relationships, behavioral dynamics, and group-level collaborative patterns, respectively.

First, the geographical view is designed to capture the spatial proximity relationships among points of interest. In real-world scenarios, points of interest that are geographically close often exhibit similarities in functionality or attributes, and user preferences formed within a specific region can frequently be transferred to other nearby locations. Therefore, in this view, geographical information is incorporated to construct a spatial correlation structure, enabling the model to capture local spatial distribution patterns and thereby improve the modeling of regional preferences.

Second, the sequential view is introduced to model the temporal dependencies in user behavior. Spatial visit patterns typically exhibit strong trajectory characteristics, where actions at different time steps are contextually related. For example, consecutively visited points of interest often reflect the current activity trajectory or short-term interests of a user. In this view, historical behavioral sequences are modeled to capture dynamically evolving interest patterns, thereby enhancing the predictive capability for subsequent actions.

Third, the collaborative view is introduced to model group-level preference relationships among users. Users exhibiting similar behavioral patterns tend to share common interests in similar types of points of interest, and such information can provide effective complementary signals under data sparsity conditions. In this view, collaborative signals embedded in the user–point-of-interest interaction structure are leveraged to learn latent associations between users and points of interest, thereby enhancing the generalization capability of the model.

The three views described above model user behavior from the perspectives of spatial structure, behavioral dynamics, and group-level information, respectively, and exhibit strong complementarity. The geographical view emphasizes spatial proximity, the sequential view focuses on behavioral evolution, and the collaborative view enables cross-user

information sharing. By jointly incorporating multi-view modeling, the issue of information deficiency inherent in single-view representations can be effectively mitigated, providing richer inputs for subsequent disentangled representation learning. It should be noted that the three views remain structurally independent, while a unified user–point-of-interest representation space is shared during parameter learning. This design enables different semantic information to be collaboratively optimized during the subsequent fusion stage. Based on this framework, the next section introduces the disentangled representation learning approach built upon the multi-view structure.

4.3 Disentangled representation learning

Based on the multi-view modeling framework, different views capture distinct semantic sources of user behavior. However, if multi-view information is directly fused, mutual interference among different semantic components may occur, thereby reducing the discriminative power of the learned representations. To address this issue, a disentangled representation learning mechanism is introduced, in which different semantic factors are mapped into relatively independent representation subspaces, thereby enhancing the model's ability to characterize complex behavioral patterns. Specifically, for the k-th view, user and point-of-interest embeddings, denoted as $e_u^{(k)}$ and $e_p^{(k)}$, are learned separately. Each view corresponds to an independent representation subspace that is dedicated to modeling a specific semantic factor, such as geographical, sequential, or collaborative information. During the representation learning process, a unified input data structure is shared across all views, while parameter updates remain relatively independent, thereby preventing semantic entanglement across views.

To further enhance disentanglement, cross-view constraints are introduced during representation learning to enforce diversity among representations from different views. Specifically, the similarity between embeddings from different views is constrained, encouraging each view to focus on its own semantic characteristics and thereby reducing redundant information propagation. This constraint mechanism improves model interpretability and strengthens the expressive capacity of each subspace.

During training, each view is updated based on its corresponding structural characteristics, while joint optimization is performed under a unified objective function. Since different views model user behavior from distinct perspectives, the disentangled representations are able to capture user preference structures at a finer granularity. For instance, the geographical view emphasizes spatial distribution patterns, the sequential view reflects dynamic behavioral evolution, and the collaborative view captures shared group-level preferences. It should be emphasized that disentangled representation learning does not imply complete isolation among views. Instead, semantic distinctions are preserved while more informative and structured representations are provided for subsequent fusion. In this manner, different semantic components can be adaptively integrated in later stages according to user-specific characteristics, thereby enabling more precise personalized modeling. In summary, the multi-view disentangled representation learning mechanism effectively reduces semantic interference, improves representation quality, and provides a reliable foundation for subsequent adaptive fusion.

4.4 Adaptive fusion mechanism

After completing multi-view disentangled representation learning, different views characterize user preferences from the perspectives of geographical structure, behavioral sequences, and collaborative relationships. However, users exhibit significant heterogeneity in their reliance on these types of information. For example, some users are more influenced by spatial proximity, whereas others depend more heavily on historical behavioral patterns or group-level preferences. Consequently, the use of fixed fusion weights is insufficient to capture such individual differences. To address this issue, an adaptive fusion mechanism is introduced, in which the importance weights of different views are dynamically assigned based on user-specific characteristics, thereby enabling personalized information integration. Specifically, for a given user u, a fusion weight vector is first generated based on the base embedding representation:

$$\boldsymbol{\alpha}_u=\mathrm{softmax}(W_f \cdot e_u+b_f) \tag{33}$$

where, $\boldsymbol{\alpha}_u=[\alpha_u^{(1)},\alpha_u^{(2)},\ldots,\alpha_u^{(K)}]$ denotes the weight distribution of users across different views, W_f and b_f are learnable parameters, and K represents the number of views.

Based on these weights, the final user representation is obtained through a weighted aggregation:

$$e_u=\sum_{k=1}^{K} \alpha_u^{(k)}\, e_u^{(k)} \tag{34}$$

For point-of-interest representations, a fusion strategy consistent with that of users is adopted, or alternatively, a shared-weight mechanism is employed for simplification, thereby ensuring structural stability and computational efficiency. The proposed adaptive mechanism enables the contribution of different views to be automatically adjusted according to user behavioral characteristics. For instance, for users exhibiting strong regional preferences, higher weights are assigned to the geographical view; conversely, for users with more stable behavioral trajectories, the influence of the sequential view is enhanced. Through such dynamic weight allocation, the model is better aligned with diverse user behavior patterns, leading to improved overall prediction performance. In addition, to prevent excessive concentration of weights on a single view, which may adversely affect model stability, normalization constraints are imposed during training to ensure that the weights across views remain within a reasonable range. This design facilitates balanced information fusion.

In summary, the adaptive fusion mechanism enables personalized integration of multi-view information based on disentangled representations. As a result, semantic distinctions are preserved while flexible adjustment according to user heterogeneity is achieved, thereby further enhancing the model's capability for spatial behavior modeling.

4.5 Model optimization

After multi-view disentangled representations and adaptive fusion have been obtained, the model is trained under a unified optimization objective to achieve joint optimization of user and point-of-interest representations. The Bayesian personalized ranking loss is adopted as the primary

optimization objective. Based on a pairwise ranking framework, this method enforces that points of interest previously visited by a user (positive samples) should be assigned higher predicted scores than unvisited points of interest (negative samples). For any training triplet (u, i, j), the predicted scores are defined as:

$$s_{ui}=e_u^T e_i, s_{uj}=e_u^T e_j \tag{35}$$

Based on these scores, the Bayesian personalized ranking loss function is formulated as:

$$L_{BPR} = -\sum_{(u,i,j)} \log\sigma\left(s_{ui} - s_{uj}\right) \tag{36}$$

where, $\sigma(\cdot)$ denotes the Sigmoid function. This objective directly optimizes the ranking order by emphasizing relative preference differences rather than absolute score values.

To mitigate overfitting, an L2 regularization term is further introduced to constrain user embeddings, POI embeddings, and other learnable parameters. The final optimization objective is formulated as:

$$L=L_{BPR}+\lambda\|\Theta\|_2^2 \tag{37}$$

where, λ denotes the regularization coefficient, and Θ represents the set of all learnable model parameters.

During training, parameters are updated using mini-batch stochastic gradient descent, and the Adam optimizer is employed to improve convergence efficiency. Within each mini-batch, positive and negative sample pairs are constructed through random sampling, and backpropagation is performed based on the aforementioned loss function. In this manner, joint optimization of multi-view representations and fusion weights is achieved. It should be noted that, since both view-specific representations and fusion weights are updated within a unified optimization framework, the model is able to automatically balance the contributions of different semantic components during training. As a result, more stable and discriminative representations are obtained.

To address the limitation that single-representation models are unable to effectively distinguish multi-source semantic information, the DisenPOI++ model is proposed based on the SF-LGCN framework. By constructing three types of semantic views—geographical, sequential, and collaborative—user behavior is modeled from multiple perspectives, including spatial structure, behavioral dynamics, and group-level preferences, thereby substantially enriching the sources of representation. On this basis, a disentangled representation learning mechanism is introduced, through which different semantic factors are mapped into relatively independent representation subspaces. This design reduces semantic interference and enhances the discriminative capability of the learned representations. Furthermore, an adaptive fusion mechanism is developed to dynamically assign view-specific weights according to user characteristics, enabling personalized integration of multi-source information and improving the model's adaptability to diverse user behavior patterns. Through these designs, effective utilization and joint optimization of multi-view information are achieved while preserving a relatively simple model structure. As a result, a more fine-grained and scalable solution for spatial behavior modeling is provided.

5. EXPERIMENTS AND ANALYSIS

In this chapter, systematic experimental evaluations are conducted for the SF-LGCN and DisenPOI++ models proposed in Chapters 3 and 4. First, the experimental setup is introduced, including the datasets, evaluation metrics, baseline methods, and parameter configurations. Subsequently, comparative experiments and ablation studies are performed for both models to verify their effectiveness and stability. Before presenting the detailed experimental results, it is necessary to clarify the complementary roles of the two models. SF-LGCN and DisenPOI++ improve point-of-interest recommendation from different perspectives and operate at different levels of modeling, thereby exhibiting a certain degree of complementarity. Specifically, SF-LGCN primarily focuses on information enhancement. Within the graph collaborative filtering framework, check-in frequency and geographical information are incorporated to improve model expressiveness through enhanced edge weighting and node representation. Owing to its relatively simple structure, this approach demonstrates strong generalizability and is well suited for recommendation models based on user–item graphs.

In contrast, DisenPOI++ emphasizes preference modeling. Through a multi-view disentangled representation learning framework combined with an adaptive fusion mechanism, multi-factor user preferences are modeled at a fine-grained level. Within this disentangled representation paradigm, collaborative information is further incorporated, and personalized fusion is achieved via dynamic weighting, thereby enhancing the model's ability to capture complex behavioral patterns.

5.1 Experimental setup

5.1.1 Datasets

Experimental evaluations were conducted on two widely used real-world location-based social network datasets:

Foursquare dataset: Collected from the Foursquare platform, this dataset contains user check-in records across different points of interest. Each record includes user identifier, point-of-interest identifier, geographical coordinates, and timestamp information.

Gowalla dataset: Collected from the Gowalla platform, this dataset has a structure similar to that of Foursquare and also contains user check-in data with geographical information.

To ensure data quality, preprocessing was performed on both datasets. In the Gowalla dataset, users with fewer than 15 visited points of interest and points of interest visited by fewer than 10 users were removed. In the Foursquare dataset, users with fewer than 10 visited points of interest and points of interest visited by fewer than 10 users were filtered out. Considering the differences in recommendation paradigms between the two models, distinct data splitting strategies were adopted. For the SF-LGCN model, which targets the general top-K recommendation task, the dataset was partitioned chronologically: the earliest 70% of interactions were used for training, the subsequent 10% for validation, and the most recent 20% for testing. For the DisenPOI++ model, which focuses on the next point-of-interest recommendation task, a time-aware leave-2-out strategy was employed. Specifically, the second-to-last interaction was used for validation, the last interaction was used for testing, and the remaining interactions were used for training.

During both training and evaluation, a negative sampling

strategy was incorporated. In the training phase, one negative sample was paired with each positive sample. In the validation and testing phases, 99 negative samples were paired with each user, thereby constructing a candidate ranking task with 100 items. The statistical characteristics of the two datasets are summarized in Table 2.

Table 2. Statistics of the Gowalla and Foursquare datasets

Dataset	Number of Users	Number of Points of Interest	Number of Check-ins	Sparsity
Gowalla	5,628	31,803	620,683	99.78%
Foursquare	7,642	28,483	512,523	99.87%

5.1.2 Evaluation metrics

The following evaluation metrics were adopted to assess recommendation performance:

Recall@K: Recall@K measures the proportion of ground-truth points of interest that are successfully retrieved within the top-K recommendation list. It is defined as:

$$\text{Recall@K}=\frac{|\text{Recommended list}\cap\text{Ground-truth visits}|}{|\text{Ground}-\text{truth visits}|} \tag{38}$$

Normalized Discounted Cumulative Gain at K (NDCG@K): NDCG@K evaluates not only whether relevant points of interest are included in the recommendation list but also their ranking positions. Higher-ranked relevant items receive higher scores. It is defined as:

$$\text{NDCG@K}=\frac{\text{DCG@K}}{\text{IDCG@K}} \tag{39}$$

$$\text{DCG@K}=\sum_{i=1}^{K}\frac{2^{rel_i}-1}{\log_2(i+1)} \tag{40}$$

where, rel_i denotes the relevance of the i-th recommendation result (set to 1 if it is relevant, and 0 otherwise), and IDCG@K represents the discounted cumulative gain value under the ideal ranking. Results were reported for K=5, K=10, and K=20. All metrics were computed on the test set and averaged over all users. It should be noted that the evaluation settings differ between the two tasks considered. In the general top-K point-of-interest recommendation task, P_u contains multiple ground-truth items from the test set, and ranking is performed over the full set of candidate points of interest. In contrast, in the next point-of-interest recommendation task, each user has only a single ground-truth next point of interest (i.e., $|P_u|=1$), and evaluation is conducted over a candidate set consisting of one positive sample and 99 randomly sampled negative samples.

5.1.3 Baseline methods

For the two recommendation paradigms considered, corresponding baseline methods were selected for comparative evaluation.

(1) Baselines for the SF-LGCN model

To evaluate the effectiveness of the proposed information enhancement strategy, the following representative methods were adopted for comparison:

New recommendation approach by exploiting geographical correlations, social correlations and categorical correlations among users and points of interest (GeoSoCa) [45]: A probabilistic modeling-based point-of-interest recommendation method that integrates geographical, social, and categorical information. This approach is used to evaluate the effectiveness of multi-source contextual modeling.

Ranking-based geographical factorization model [10]: A weighted matrix factorization method that incorporates geographical factors into latent representation learning. It serves as a representative approach among traditional methods.

Local geographical-based logistic matrix factorization [57]: A model that combines local geographical information with logistic matrix factorization. Prediction probabilities are adjusted using geographical weights, making it suitable for evaluating the capability of geographical information modeling.

Light graph convolutional network [43]: A lightweight graph collaborative filtering model that retains only the neighborhood aggregation operation. It is employed as the direct baseline of the proposed model.

(2) Baselines for the DisenPOI++ model

To evaluate the effectiveness of the proposed multi-view disentangled representation learning and adaptive fusion mechanisms, the following representative methods were selected for comparison:

Ranking-based geographical factorization model [10]: A matrix factorization-based method, included as a representative traditional approach for comparison with deep learning models.

Long- and short-term preference modeling [17]: A recurrent neural network-based spatiotemporal modeling method that integrates long-term and short-term preferences, serving as a representative approach for sequential modeling.

Light graph convolutional network [26]: A representative graph collaborative filtering method, used to evaluate the capability of modeling user–point-of-interest interactions based on graph structures.

Spatio-temporal attention network [18]: A spatiotemporal attention-based method that dynamically models the importance of historical behaviors.

DisenPOI [38]: A disentangled representation learning method that decomposes user preferences into geographical and sequential components, serving as the most directly comparable baseline.

Disentangled contrastive hypergraph learning [33]: A hypergraph-based recommendation method that captures high-order relationships through multiple hypergraph structures, representing recent advances in graph-based modeling.

5.1.4 Experimental platform and parameter settings

The hardware and software environments used for the experiments are summarized in Table 3.

For all baseline methods, publicly available source code implementations were used whenever possible, and experiments were conducted following the parameter settings recommended in the original studies. For the two proposed models, optimal hyperparameters were determined through grid search on the validation set.

5.2 Experimental results and analysis of the proposed medel

5.2.1 Comparison with baseline methods

The performance comparison between SF-LGCN and baseline methods on the Foursquare and Gowalla datasets is presented in Tables 4 and 5, respectively. Based on the experimental results reported in the two tables, the following observations and analyses can be derived.

Table 3. Experimental environment configuration

Configuration Item	Specification
Graphics Processing Unit	NVIDIA RTX 3090 (24 GB)
Central Processing Unit	14 virtual central processing units, Intel Xeon Gold 6330 @ 2.00 GHz
Memory	90 GB
Operating System	Ubuntu 20.04
Deep Learning Framework	PyTorch 2.0.0
Compute Unified Device Architecture Version	11.8
Python Version	3.8

Table 4. Performance comparison on the Foursquare dataset

Method	Recall@5	Recall@10	Recall@20	NDCG@5	NDCG@10	NDCG@20
GeoSoCa	0.018	0.033	0.058	0.019	0.028	0.039
RankGeoFM	0.03831	0.06290	0.099	0.056	0.061	0.075
LGLMF	0.026	0.042	0.067	0.033	0.028	0.024
LightGCN	0.046	0.069	0.107	0.071	0.072	0.085
SF-LightGCN	**0.048**	**0.071**	**0.110**	**0.073**	**0.075**	**0.089**

GeoSoCa = Geographical, Social and Categorical correlations, RankGeoFM = Ranking-based geographical factorization model, LGLMF = Local geographical-based logistic matrix factorization, LightGCN = Light graph convolutional network, SF-LightGCN = Spatial Frequency–aware Light Graph Convolutional Network

Table 5. Performance comparison on the Gowalla dataset

Method	Recall@5	Recall@10	Recall@20	NDCG@5	NDCG@10	NDCG@20
GeoSoCa	0.020	0.033	0.051	0.025	0.035	0.044
RankGeoFM	0.032	0.055	0.088	0.067	0.069	0.078
LGLMF	0.021	0.036	0.060	0.044	0.040	0.035
LightGCN	0.036	0.058	0.090	0.078	0.077	0.085
SF-LightGCN	**0.036**	**0.062**	**0.096**	**0.080**	**0.081**	**0.090**

GeoSoCa = Geographical, Social and Categorical correlations, RankGeoFM = Ranking-based geographical factorization model, LGLMF = Local geographical-based logistic matrix factorization, LightGCN = Light graph convolutional network, SF-LightGCN = Spatial Frequency–aware Light Graph Convolutional Network

(1) Comparison between SF-LGCN and the light graph convolutional network

The experimental results indicate that SF-LGCN achieves improvements of 2.8% and 4.1% over the light graph convolutional network in terms of Recall@10 and NDCG@10, respectively. Consistent performance gains are observed across both datasets, thereby validating the effectiveness of the proposed information enhancement strategy. These improvements can be attributed to two primary factors. First, check-in frequency is incorporated as edge weights, enabling differentiated modeling of user preference intensity. Second, geographical coordinates of points of interest are integrated into the initial embeddings, providing spatial semantic information that enhances representation expressiveness.

(2) Comparison with other baseline methods

GeoSoCa exhibits relatively weak performance on both datasets, suggesting that traditional probabilistic models are limited in handling complex point-of-interest recommendation scenarios. In contrast, SF-LGCN significantly outperforms the ranking-based geographical factorization model and the local geographical-based logistic matrix factorization, demonstrating the superiority of graph neural networks in modeling high-order interaction relationships. Although the ranking-based geographical factorization model and the local geographical-based logistic matrix factorization incorporate geographical factors, their performance is constrained by the limitations of linear modeling, which restricts their ability to capture multi-factor coupling relationships. By leveraging graph-based propagation, SF-LGCN is able to more effectively model complex behavioral patterns, resulting in superior recommendation performance.

5.2.2 Ablation study

To evaluate the independent contributions of each component in SF-LGCN, several ablation variants were designed as follows:

SF-LGCN-F: Only frequency-weighted edges are retained, while geographical information is not incorporated;

SF-LGCN-S: Only geographical information is introduced (with a fixed weight of 0.3), while an unweighted graph structure is adopted;

SF-LGCN-S-learn: Geographical information is incorporated with a learnable weight, while the graph structure remains unweighted.

The experimental results of these variants are presented in Table 6.

The following conclusions can be drawn from Table 6:

Superior performance of the full model. The SF-LGCN model consistently outperforms all ablation variants across all evaluation metrics, indicating that frequency weighting and geographical fusion exhibit clear complementarity. Their joint modeling enables a more comprehensive characterization of user preferences.

Table 6. Ablation results of the Spatial Frequency–aware Light Graph Convolutional Network (SF-LightGCN) model

Variant	Foursquare Recall@20	Foursquare NDCG@20	Gowalla Recall@20	Gowalla NDCG@20
SF-LGCN-F	0.107	0.084	0.095	0.088
SF-LGCN-S	0.109	0.087	0.093	0.089
SF-LGCN-S-learn	0.109	0.088	0.093	0.088
SF-LGCN (Full)	0.110	0.089	0.096	0.090

More significant contribution of geographical fusion. The variant incorporating only geographical information (SF-LGCN-S) achieves performance levels close to those of the full model (e.g., Recall@20 on Foursquare: 0.109 vs. 0.110). This observation suggests that latitude–longitude information provides an effective spatial semantic prior and serves as the primary source of performance improvement.

Auxiliary role of frequency weighting. The performance of SF-LGCN-F is slightly lower than that of the geographical fusion variants but remains superior to several baseline methods. This indicates that check-in frequency contributes additional information regarding preference intensity and plays a complementary role in the modeling process.

Limited impact of learnable weighting. Comparable results are observed between SF-LGCN-S and SF-LGCN-S-learn, suggesting that a fixed weighting coefficient (0.3) is sufficient to balance geographical information and the original embedding. Consequently, the additional benefit of learning this parameter is marginal.

The ablation study demonstrates that geographical fusion constitutes the primary source of performance gains, while frequency weighting provides complementary benefits. The integration of both components yields optimal recommendation performance.

5.2.3 Parameter sensitivity analysis

To further investigate the contributions of different components in SF-LGCN, a series of hyperparameter experiments were conducted. By varying key parameters, including the number of graph convolution layers, embedding dimensionality, and learning rate, the sensitivity of model performance—measured by recall and Normalized Discounted Cumulative Gain (NDCG)—was systematically analyzed. This analysis provides insights into the model's behavior under different configurations and offers guidance for appropriate parameter selection and subsequent optimization.

(a) Effect of the number of graph convolution layers (L). The number of graph convolution layers determines the extent of message propagation in the graph. As the number of layers increases, each node is able to aggregate information from more distant neighbors; however, excessive depth may lead to the over-smoothing problem. As illustrated in Figure 11, model performance on both datasets improves as the number of layers increases. Nevertheless, after reaching three layers, although marginal improvements are still observed, the performance gains become significantly diminished. Therefore, setting the number of graph convolution layers to three is generally sufficient, as further increases yield limited benefits.

(b) Effect of the embedding dimension (d). The embedding dimensionality determines the representational capacity of the model. Higher-dimensional embeddings are capable of encoding richer information; however, they also introduce an increased risk of overfitting and higher computational cost. As illustrated in Figure 12, model performance continues to improve as the embedding dimension increases, although the magnitude of improvement gradually diminishes. This trend is consistent with the observations regarding the number of graph convolution layers. Therefore, to balance model expressiveness and computational efficiency, an embedding dimension of 64 is considered an appropriate choice in this study, as it achieves competitive performance without imposing excessive computational overhead.

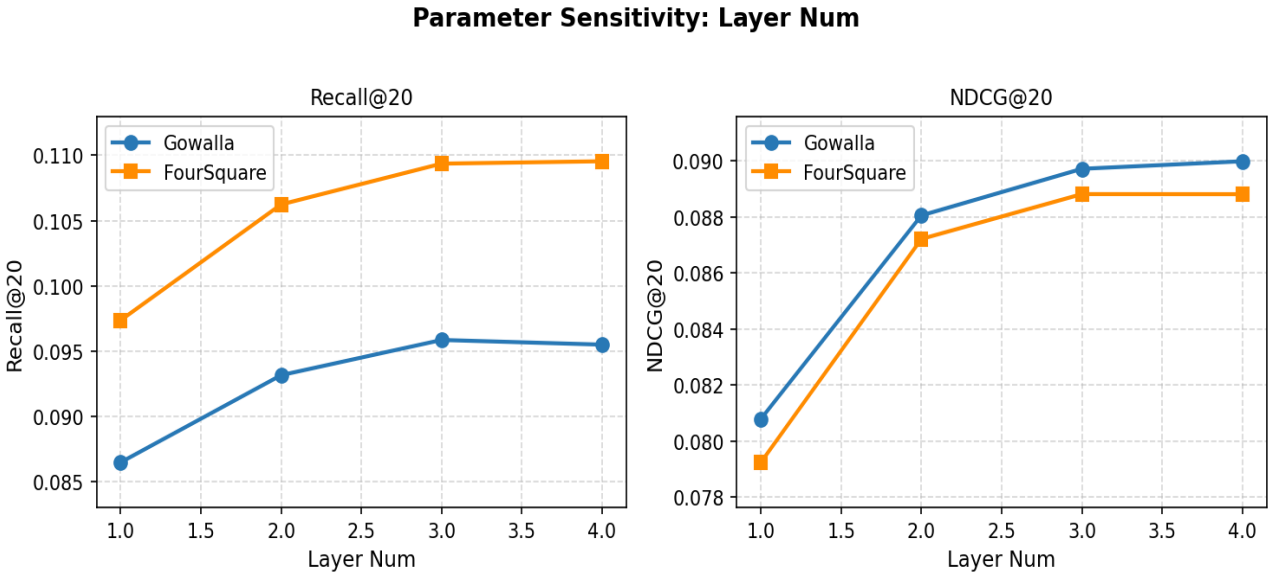


Figure 11. Effect of the number of graph convolution layers on the performance of the Spatial Frequency–aware Light Graph Convolutional Network (SF-LightGCN) model

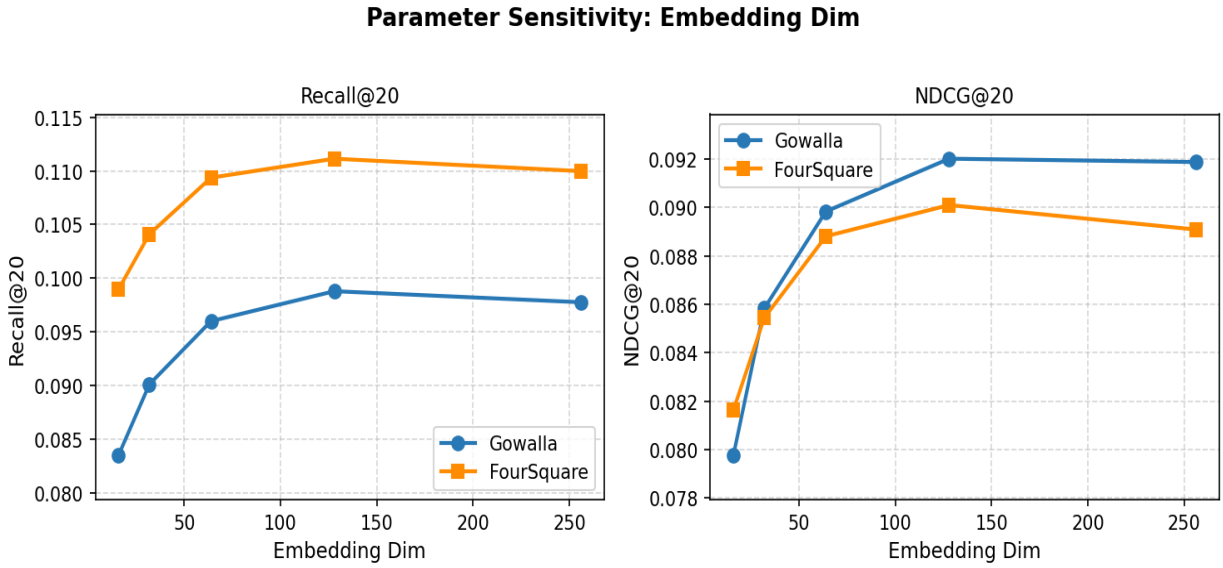


Figure 12. Effect of embedding dimension on the performance of the Spatial Frequency–aware Light Graph Convolutional Network (SF-LightGCN) model

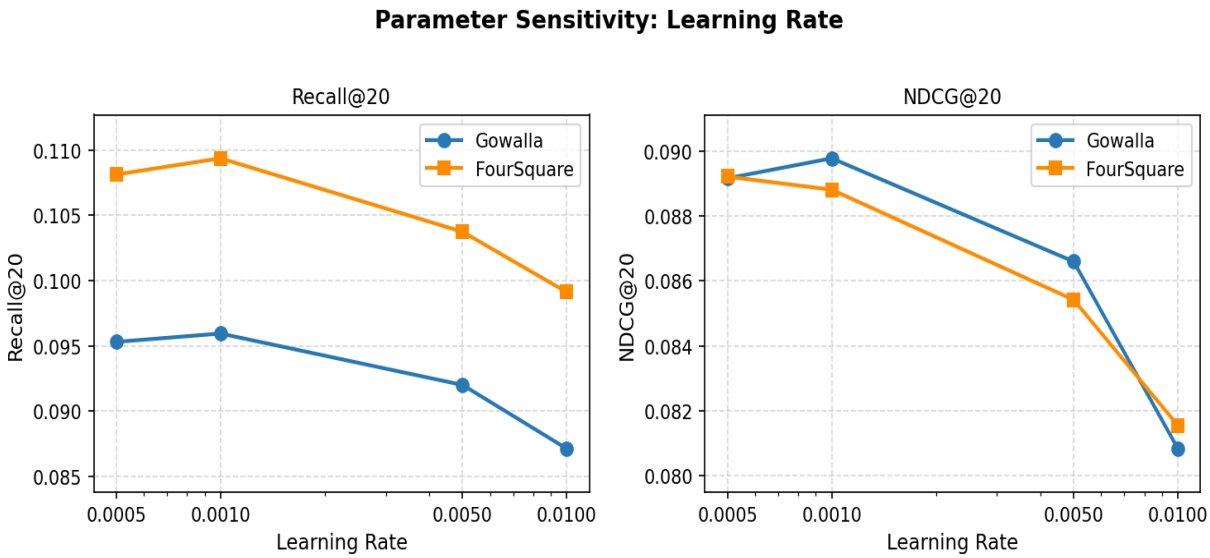


Figure 13. Effect of learning rate on the performance of the Spatial Frequency–aware Light Graph Convolutional Network (SF-LightGCN) model

(c) Effect of the learning rate. The learning rate determines the step size of parameter updates during model training. A higher learning rate can accelerate convergence but may lead to training instability or convergence to suboptimal solutions, whereas a lower learning rate results in slower convergence and increased training time. As shown in Figure 13, when the learning rate is set to 0.001, the model achieves relatively strong performance on both datasets. As the learning rate increases beyond this value, performance begins to deteriorate. Therefore, setting the learning rate to 0.001 is empirically validated as an appropriate configuration in this study.

5.3 Experimental results and analysis of the disentangled dual-graph framework for point-of-interest recommendation (DisenPOI)++

5.3.1 Comparison with baseline methods

The performance comparisons between DisenPOI++ and baseline methods on the two datasets are presented in Tables 7 and 8. Based on the experimental results from both datasets, several observations can be made. As shown in Tables 7 and 8, DisenPOI++ consistently outperforms the original DisenPOI across all evaluation metrics, thereby directly validating the effectiveness of the proposed collaborative graph enhancement and adaptive gating mechanisms. Specifically, on the Foursquare dataset, DisenPOI++ achieves a 3.7% improvement in Recall@5 and an 8.4% improvement in NDCG@5 compared to DisenPOI. On the Gowalla dataset, improvements of 7.5% in Recall@5 and 10.8% in NDCG@10 are observed. These results further indicate that the proposed model exhibits stronger performance on the Gowalla dataset, suggesting enhanced robustness under different data distributions.

When compared with the most recent baseline method, disentangled contrastive hypergraph learning, DisenPOI++ does not achieve superior performance across all metrics. This discrepancy can be attributed to the fact that the disentangled contrastive hypergraph learning leverages hypergraph

structures to capture higher-order and more complex relationships within the data, thereby enabling a richer representation of user–point-of-interest interactions. Such modeling capability is not explicitly incorporated into DisenPOI++. Nevertheless, DisenPOI++ demonstrates higher performance than the disentangled contrastive hypergraph learning in terms of Recall@20 on both datasets. This observation suggests that, as the candidate set expands, the proposed model remains effective in identifying relevant target points of interest. Consequently, an advantage in long-tail recommendation scenarios is implied, where capturing less frequent but relevant items becomes critical. From a broader perspective, consistent trends can be observed across different categories of methods. Graph-based approaches (e.g., the light graph convolutional network, DisenPOI, the disentangled contrastive hypergraph learning, and DisenPOI++) generally outperform traditional matrix factorization methods (e.g., the ranking-based geographical factorization model), highlighting the superiority of graph neural networks in modeling user–point-of-interest interaction structures. Furthermore, methods incorporating disentangled representation learning (e.g., DisenPOI, DisenPOI++, and the disentangled contrastive hypergraph learning) consistently outperform graph-based methods without disentanglement, thereby validating the effectiveness of multi-factor disentangled modeling in enhancing recommendation performance.

Table 7. Performance comparison of the disentangled dual-graph framework for point-of-interest recommendation (DisenPOI)++ on the Foursquare dataset

Method	Recall@5	Recall@10	Recall@20	NDCG@5	NDCG@10	NDCG@20
Ranking-based geographical factorization model	0.641	0.711	0.785	0.519	0.532	0.559
Light graph convolutional network	0.654	0.728	0.803	0.526	0.544	0.571
Long- and short-term preference modeling	0.600	0.674	0.757	0.522	0.546	0.567
Spatio-temporal attention network	0.693	0.817	0.898	0.524	0.569	0.587
DisenPOI	0.708	0.828	0.911	0.535	0.574	0.595
Disentangled contrastive hypergraph learning	0.794	0.856	0.903	0.671	0.691	0.703
DisenPOI++	**0.734**	**0.849**	**0.922**	**0.580**	**0.616**	**0.634**

Table 8. Performance comparison of the disentangled dual-graph framework for point-of-interest recommendation (DisenPOI)++ on the Gowalla dataset

Method	Recall@5	Recall@10	Recall@20	NDCG@5	NDCG@10	NDCG@20
Ranking-based geographical factorization model	0.658	0.768	0.829	0.484	0.511	0.527
Light graph convolutional network	0.671	0.785	0.862	0.499	0.529	0.542
Long- and short-term preference modeling	0.684	0.776	0.843	0.495	0.521	0.538
Spatio-temporal attention network	0.657	0.798	0.911	0.489	0.510	0.552
DisenPOI	0.668	0.804	0.918	0.501	0.535	0.564
Disentangled contrastive hypergraph learning	0.783	0.865	0.921	0.652	0.678	0.693
DisenPOI++	**0.718**	**0.847**	**0.934**	**0.555**	**0.597**	**0.619**

5.3.2 Ablation study

Ablation experiments were conducted to systematically evaluate the contributions of individual components in DisenPOI++. Four model variants were designed as follows:

(a) M1 (Original DisenPOI)

The original DisenPOI model is adopted, which employs a dual-view disentangled structure consisting of a geographical graph and a sequential graph. A fixed weighting coefficient (λ) is used for fusion. This variant serves as the baseline model.

(b) M2 (+ Collaborative Graph)

A collaborative graph is introduced on top of DisenPOI, extending the model to a three-view disentangled structure. The fusion strategy remains based on fixed weights. This variant is used to evaluate the independent contribution of the collaborative graph.

(c) M3 (+ Collaborative Graph + Adaptive Weighting)

Building upon the three-view disentangled structure, user-specific adaptive gating weights are further incorporated. This variant is designed to assess the combined effect of multi-view disentanglement and adaptive gating mechanisms.

(d) M4 (Full DisenPOI++ Model)

The complete DisenPOI++ model is employed, integrating all proposed components. This variant is used to evaluate the overall recommendation performance of the final model.

Table 9. Ablation results of the disentangled dual-graph framework for point-of-interest recommendation (DisenPOI)++ on the Foursquare dataset

Variant	Foursquare Recall@10	Foursquare Recall@20	Foursquare NDCG@10	Foursquare NDCG@20
M1	0.828	0.911	0.574	0.595
M2	0.830	0.916	0.586	0.599
M3	0.839	0.919	0.597	0.620
M4	0.849	0.922	0.616	0.634

Table 10. Ablation results of the disentangled dual-graph framework for point-of-interest recommendation (DisenPOI)++ on the Gowalla dataset

Variant	Gowalla Recall@10	Gowalla Recall@20	Gowalla NDCG@10	Gowalla NDCG@20
M1	0.804	0.918	0.535	0.564
M2	0.811	0.920	0.557	0.579
M3	0.829	0.924	0.580	0.596
M4	0.847	0.934	0.597	0.619

To validate the effectiveness of each core component in DisenPOI++, ablation experiments were conducted on both the Foursquare and Gowalla datasets. The corresponding results are presented in Tables 9 and 10. Several observations and insights can be derived from these results.

Across both datasets, model performance exhibits a monotonic improvement as the collaborative graph, adaptive weighting, and attention mechanism are progressively introduced. No performance degradation is observed in any intermediate variant, indicating that the proposed components are well-designed, mutually compatible, and free from redundancy or conflict. Notably, differences in performance gains are observed across the two datasets. On the Gowalla dataset, the improvements introduced by each component are more pronounced, whereas on the Foursquare dataset, the gains are relatively moderate. This discrepancy can be explained by the intrinsic characteristics of the datasets. The Foursquare dataset is collected from densely populated metropolitan areas such as New York and Tokyo, where points of interest are highly concentrated and user check-in behaviors exhibit strong spatial locality. Under such conditions, the original DisenPOI model—based on geographical–sequential disentanglement—is already capable of effectively capturing user preferences, leaving limited room for further improvement. In contrast, the Gowalla dataset covers a broader geographical range with a sparser point-of-interest distribution and more diverse user behavior patterns. In such scenarios, collaborative information and personalized fusion mechanisms become more beneficial, thereby yielding larger performance gains. This difference further demonstrates that the proposed enhancements provide stable and consistent improvements across datasets with varying characteristics.

A more detailed quantitative analysis can be derived from the results. On the Gowalla dataset, the introduction of the collaborative graph leads to a 4.1% improvement in NDCG@10, indicating a substantial enhancement in ranking quality due to collaborative signals. The incorporation of adaptive weighting further increases Recall@10 from 0.811 to 0.829, highlighting the positive impact of personalized fusion on hit rate. Additionally, the attention mechanism contributes to a 2.9% improvement in NDCG@20, validating its effectiveness in optimizing long-sequence ranking performance. On the Foursquare dataset, although the magnitude of improvement is smaller than that observed on Gowalla, consistent performance gains are still achieved. This result provides further evidence of the general applicability of the proposed approach and demonstrates the robustness and compatibility of the designed modules.

5.3.3 Training convergence analysis

Figure 14 illustrates the variation of Recall@20 on the validation set with respect to training epochs for DisenPOI++ on the Gowalla dataset. Overall, as the model evolves progressively from M1 to M4, both the convergence speed during the early training stage and the overall performance are improved. This observation indicates that the introduced modules effectively enhance the model's representation learning capability and optimization efficiency. From the perspective of the overall convergence curves, M1 exhibits the lowest initial performance and relatively slow convergence. After partial enhancements are introduced, M2 demonstrates a noticeably faster convergence rate and achieves a stable improvement in validation performance. With further optimization, M3 not only maintains a higher Recall@20 level during the middle and later training stages but also presents a smoother convergence curve, reflecting improved training stability. The complete model, M4, achieves the fastest performance growth during the initial epochs and reaches its performance peak at an earlier stage, indicating superior early convergence capability and higher peak performance.

At the same time, it can be observed that M4 exhibits slight fluctuations after reaching its peak, whereas M3 demonstrates greater stability during the later training stage. This phenomenon suggests that, as the model's expressive capacity increases, the optimization process becomes more sensitive to parameter updates. However, the magnitude of these fluctuations remains limited, and no sustained performance degradation is observed, indicating that severe overfitting does not occur. Overall, the convergence behavior from M1 to M4 provides strong empirical evidence for the effectiveness of the proposed components. The full model achieves advantages in both convergence speed and peak performance, while the optimized variants exhibit satisfactory training stability throughout the learning process.

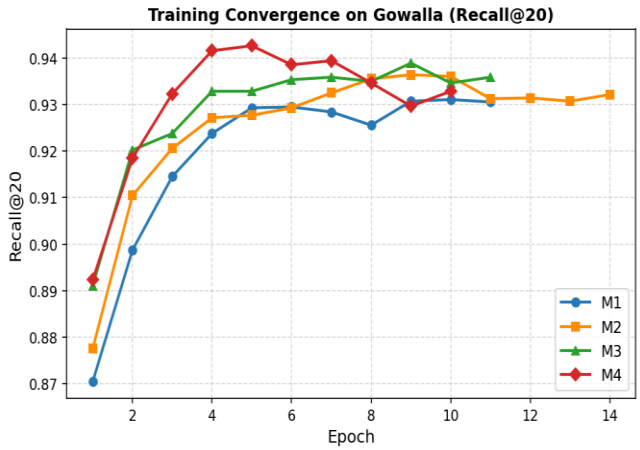


Figure 14. Convergence curves of the disentangled dual-graph framework for point-of-interest recommendation (DisenPOI)++ variants on the Gowalla dataset (Recall@20)

5.3.4 Visualization analysis of adaptive weights

To provide an intuitive validation that the adaptive gating mechanism effectively learns user-specific weight distributions, a visualization analysis of the gate values was conducted. The Gowalla dataset was taken as a representative example, and the distribution of adaptive gate values was examined to verify that differentiated user preferences had indeed been captured. Figure 15 presents the histogram of gate values for all users in the test set.

From the distribution results, it can be observed that the gate values corresponding to the three views do not collapse to a fixed constant. Instead, distinct and dispersed distributions are exhibited, indicating that the gating network does not perform static weighting but rather adaptively adjusts the contributions of multiple views according to individual user preference characteristics. Specifically, the average weights of the sequential view and the geographical view are approximately 0.509 and 0.440, respectively, with relatively wide distribution ranges. This suggests that these two sources of information serve as the primary drivers for next-point-of-interest recommendation. In contrast, the weight of the collaborative view is concentrated within a relatively small range, with an average value of approximately 0.051, indicating that collaborative signals primarily function as a complementary component under the current dataset and modeling configuration. Overall, the visualization of gate value distributions provides empirical evidence that the proposed user-level adaptive weighting mechanism is effective. It enables dynamic allocation of importance across the three semantic views based on heterogeneous user preferences.

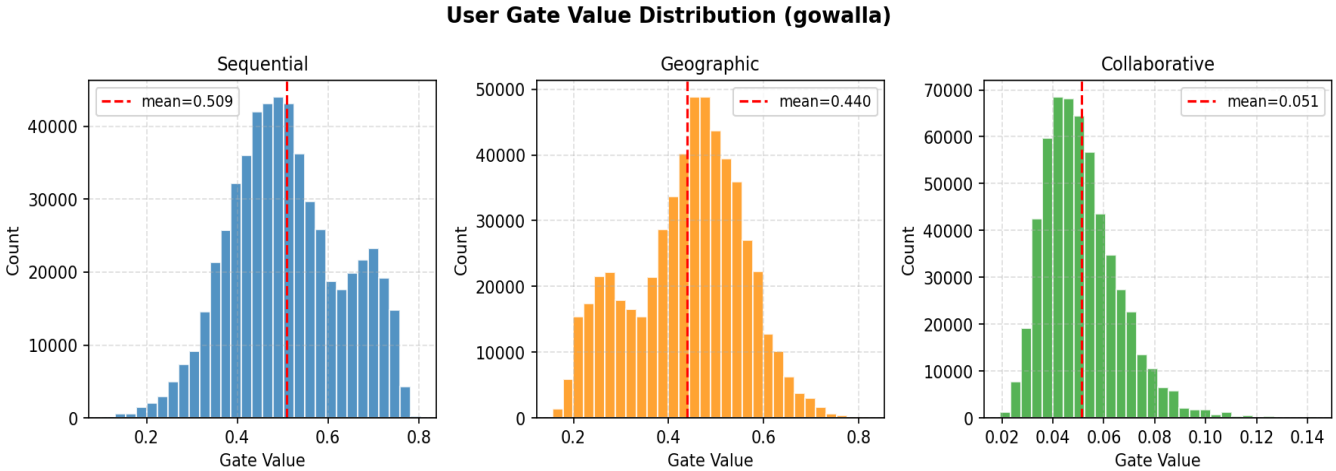


Figure 15. Histogram of user gate value distributions

5.3.5 Hyperparameter sensitivity analysis

To investigate the impact of key hyperparameters on the performance of DisenPOI++, a series of sensitivity experiments was conducted.

(a) Effect of embedding dimension

As illustrated in Figure 16, when the embedding dimension increases from 32 to 256, both Recall@10 and NDCG@10 on the two datasets exhibit an overall upward trend. This observation indicates that higher-dimensional representation spaces facilitate the modeling of richer sequential, geographical, and collaborative features. Although performance improvements continue as the dimensionality increases, the magnitude of improvement on the Gowalla dataset gradually diminishes, revealing a clear pattern of diminishing marginal returns. Overall, while higher embedding dimensions can yield optimal performance, dimensions of 64 and above are sufficient to provide stable and effective representations. Therefore, in practical applications, an appropriate embedding dimension should be selected by balancing performance gains against computational cost.

(b) Effect of the number of geographical graph convolution layers

The impact of the number of geographical graph convolution layers on model performance is illustrated in Figure 17. Overall, increasing the depth of the geographical graph does not consistently lead to performance improvement. For the Foursquare dataset, optimal performance across all evaluation metrics is achieved when the number of layers is set to two. Further increasing the depth to three or four layers results in performance degradation. This phenomenon indicates that moderate geographical neighborhood aggregation is sufficient to capture spatial proximity relationships, whereas excessive propagation introduces redundant spatial information and may lead to over-smoothing. For the Gowalla dataset, performance fluctuations are relatively minor. Specifically, NDCG@10 reaches its peak at two layers, while Recall@10 exhibits a slight improvement at four layers; however, the overall differences remain marginal. By jointly considering the results from both datasets, a two-layer geographical graph convolution structure is observed to provide an effective balance between performance and stability.

(c) Effect of the number of collaborative graph convolution layers

The experimental results for varying the number of collaborative graph convolution layers are presented in Figure 18. It can be observed that, on both datasets, optimal or near-optimal performance in terms of Recall@10 and NDCG@10 is achieved when a single layer is employed. As the number of layers increases to two and three, an overall decline in performance is observed, with a more pronounced degradation on the Foursquare dataset. This phenomenon suggests that the effectiveness of the collaborative view primarily originates from first-order user–point-of-interest collaborative relationships. A single propagation layer is sufficient to incorporate shared preference signals across users. When the propagation depth increases, noise and irrelevant higher-order collaborative relations are progressively introduced, thereby weakening the contribution of meaningful signals. In conjunction with the preceding analysis, this finding further indicates that the collaborative view is more suitable as a lightweight complement to the original DisenPOI framework rather than as a deeply stacked component. Consequently, the number of collaborative graph convolution layers is set to one in subsequent experiments.

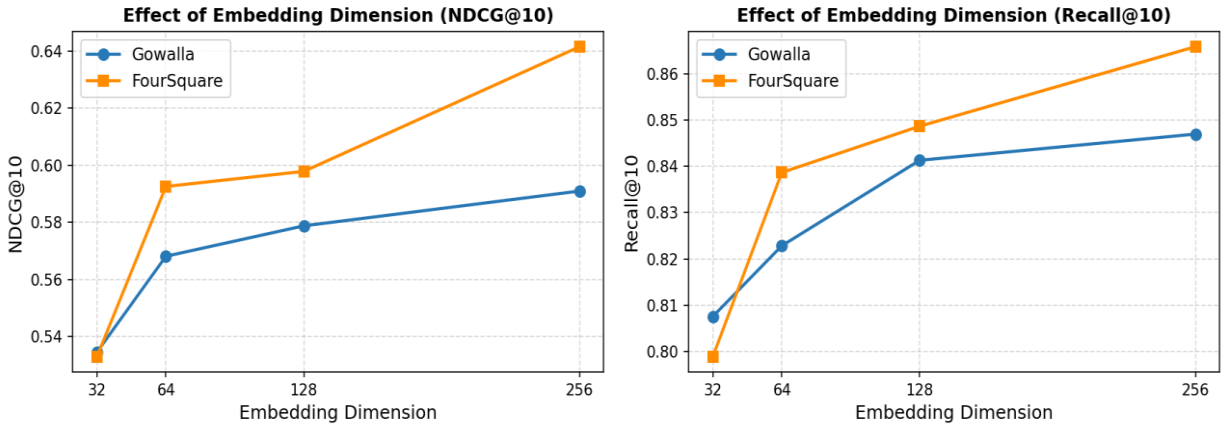


Figure 16. Impact of embedding dimension on the performance of the disentangled dual-graph framework for point-of-interest recommendation (DisenPOI)++

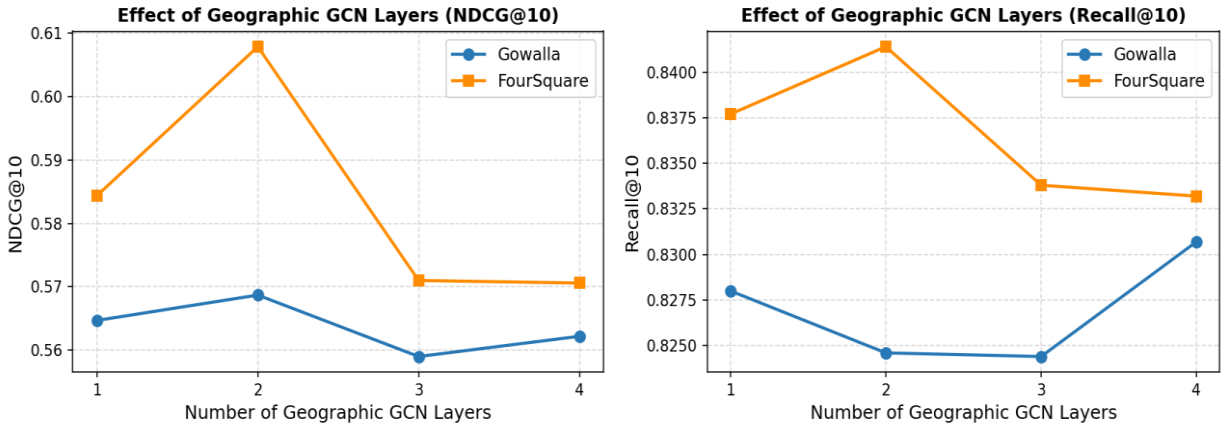


Figure 17. Impact of the number of geographical graph convolution layers on the performance of the disentangled dual-graph framework for point-of-interest recommendation (DisenPOI)++

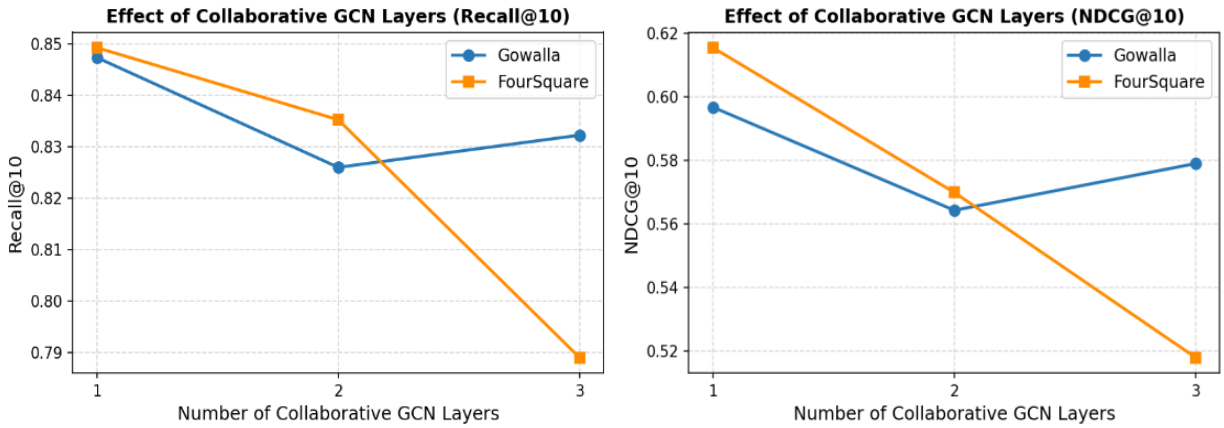


Figure 18. Impact of the number of collaborative graph convolution layers on the performance of the disentangled dual-graph framework for point-of-interest recommendation (DisenPOI)++

A comprehensive experimental evaluation of the SF-LGCN and DisenPOI++ models was conducted. The results demonstrate that both approaches achieve consistent performance improvements in point-of-interest recommendation tasks. For SF-LGCN, performance gains are obtained through the incorporation of check-in frequency weighting and geographical information fusion. The model consistently outperforms the light graph convolutional network and other baseline methods on both the Foursquare dataset and Gowalla dataset. Ablation studies further indicate that frequency weighting and geographical fusion exhibit clear complementary effects. For DisenPOI++, enhanced modeling capability is achieved by introducing a collaborative view and an adaptive fusion mechanism. Superior performance is observed in the next-point-of-interest prediction task. Ablation analysis confirms that each component contributes stable performance gains, with more pronounced improvements observed on the Gowalla dataset. Overall, the experimental findings validate the effectiveness of the proposed methods in terms of performance, stability, and generalization capability.

6. CONCLUSION AND FUTURE WORK

This study systematically investigated spatial behavior modeling and point-of-interest recommendation, addressing the limitations of conventional methods in terms of insufficient utilization of spatial information and the entanglement of multi-source semantics. First, check-in frequency and geographical information were incorporated into the base model, and a weighted graph structure was constructed to enhance the representation of spatial distribution characteristics. Building upon this foundation, the DisenPOI++ model was further proposed. Through multi-view modeling, user behavior was decomposed into distinct semantic spaces, including geographical, sequential, and collaborative components. By integrating disentangled representation learning with an adaptive fusion mechanism, effective integration of multi-source information and personalized modeling were achieved. Experimental results demonstrated consistent improvements across multiple evaluation metrics, thereby validating the effectiveness of the proposed multi-view disentanglement and dynamic fusion strategies in spatial recommendation tasks. Compared with single-representation approaches, the proposed model enables a more fine-grained characterization of user preference structures and exhibit enhanced robustness and generalization capability in complex scenarios.

Despite the progress achieved in spatial modeling, several directions for further improvement remain. First, richer contextual information, such as temporal dynamics, social relationships, and environmental factors, could be incorporated to further refine behavior modeling. Second, more efficient lightweight architectures could be explored to reduce model complexity and improve scalability in large-scale applications. Finally, multimodal data, including images and textual information, could be integrated for joint modeling, thereby extending the applicability of the approach to heterogeneous data environments. In summary, an effective framework for multi-view spatial behavior modeling has been established, providing a solid foundation for future research on point-of-interest recommendation in complex settings.

REFERENCES

[1] Ye, M., Yin, P., Lee, W.C., Lee, D.L. (2011). Exploiting geographical influence for collaborative point-of-interest recommendation. In Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval, Beijing, China, pp. 325-334. https://doi.org/10.1145/2009916.2009962

[2] Liu, Y., Wei, W., Sun, A., Miao, C. (2014). Exploiting geographical neighborhood characteristics for location recommendation. In Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management, Shanghai, China, pp. 739-748. https://doi.org/10.1145/2661829.2662002

[3] Guo, D., Zhang, M., Jia, N., Wang, Y. (2020) Survey of point-of-interest recommendation research fused with deep learning. Geomatics & Information Science of Wuhan University, 45(12): 1890-1902. https://doi.org/10.13203/j.whugis20200334

[4] Jiang, T., Xu, S., Li, X., Zhang, Z., Wang, Y., Luo, A., He, X. (2024). POI recommendation of spatiotemporal sequence embedding in gated dilation residual network. Geomatics and Information Science of Wuhan University, 49(9): 1683-1692. https://doi.org/10.13203/j.whugis20220658

[5] Chen, J., Zhang, W. (2022). Review of point of interest recommendation systems in location-based social networks. Journal of Frontiers of Computer Science and Technology, 16(7): 1462-1478. https://doi.org/10.3778/j.issn.1673-9418.2112037

[6] Wang, Y., Zhang, X. (2024). Multi-coding next point of interest recommendation model based on GT model. Application Research of Computers, 41(11): 3382-3388. https://doi.org/10.19734/j.issn.1001-3695.2024.03.0092

[7] Ren, R., Li, Y., Yang, Y., Song, P. (2025). Review of POI recommendation algorithms. Computer Engineering and Applications, 61(13): 62-77. https://doi.org/10.3778/j.issn.1002-8331.2410-0453

[8] Zhu, H., Wu, H. (2024). Popularity bias mitigation in recommender systems via contrastive decoupling learning. Journal of Yunnan University(Natural Sciences Edition), 4(2): 224-232. https://doi.org/10.7540/j.ynu.20240045

[9] Bao, J., Zheng, Y., Mokbel, M.F. (2012). Location-based and preference-aware recommendation using sparse geo-social networking data. In Proceedings of the 20th International Conference on Advances in Geographic Information Systems, Redondo Beach, California, pp. 199-208. https://doi.org/10.1145/2424321.2424348

[10] Lian, D., Zhao, C., Xie, X., Sun, G., Chen, E., Rui, Y. (2014). GeoMF: Joint geographical modeling and matrix factorization for point-of-interest recommendation. In Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, USA, pp. 831-840. https://doi.org/10.1145/2623330.2623638

[11] Cheng, C., Yang, H., Lyu, M.R., King, I. (2013). Where you like to go next: Successive point-of-interest recommendation. In Proceedings of the Twenty-Third international Joint Conference on Artificial Intelligence, Beijing, China, pp. 2605-2611.

[12] Feng, S., Cong, G., An, B., Chee, Y.M. (2017). Poi2vec: Geographical latent representation for predicting future visitors. In Proceedings of the AAAI Conference on Artificial Intelligence, San Francisco, California, USA, pp. 102-108.

[13] Zhao, S., Zhao, T., Yang, H., Lyu, M., King, I. (2016). STELLAR: Spatial-temporal latent ranking for successive point-of-interest recommendation. In Proceedings of the AAAI Conference on Artificial Intelligence, Phoenix, Arizona, pp. 315-321.

[14] Zhao, P., Luo, A., Liu, Y., Xu, J., et al. (2020). Where to go next: A spatio-temporal gated network for next poi recommendation. IEEE Transactions on Knowledge and Data Engineering, 34(5): 2512-2524. https://doi.org/10.1109/TKDE.2020.3007194

[15] Hidasi, B., Karatzoglou, A., Baltrunas, L., Tikk, D. (2015). Session-based recommendations with recurrent neural networks. arXiv preprint arXiv:1511.06939. https://doi.org/10.48550/arXiv.1511.06939

[16] Liu, Q., Wu, S., Wang, L., Tan, T. (2016). Predicting the next location: A recurrent model with spatial and temporal contexts. In Proceedings of the AAAI Conference on Artificial Intelligence, Phoenix, Arizona, pp. 194-200.

[17] 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. Proceedings of the AAAI Conference on Artificial Intelligence, 34(1): 214-221. https://doi.org/10.1609/aaai.v34i01.5353

[18] Luo, Y., Liu, Q., Liu, Z. (2021). Stan: Spatio-temporal attention network for next location recommendation. In Proceedings of the Web Conference 2021, New York, USA, pp. 2177-2185. https://doi.org/10.1145/3442381.3449998

[19] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., et al. (2017). Attention is all you need. arXiv preprint arXiv:1706.03762. https://doi.org/10.48550/arXiv.1706.03762

[20] 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, Madrid, Spain, pp. 1144-1153. https://doi.org/10.1145/3477495.3531983

[21] Kang, W.C., McAuley, J. (2018). Self-attentive sequential recommendation. In 2018 IEEE International Conference on Data Mining (ICDM), Singapore, pp. 197-206. https://doi.org/10.1109/ICDM.2018.00035

[22] Huang, L., Jiang, B., Lv, S., Liu, Y., Li, D. (2018). Survey on deep learning based recommender systems. Chinese Journal of Computers.

[23] Xing, X., Liu, J., Wang, T., Wang, H., Jia, Z. (2023). A survey of multi-scenario applications of graph neural network in recommendation system. Journal of Bohai University (Natural Science Edition), 44(4): 368-375. https://doi.org/10.3969/j.issn.1673-0569.2023.04.012

[24] Chen, J., Meng, X., Ji, W., Zhang, Y. (2020). POI recommendation based on multidimensional context-aware graph embedding model. Journal of Software, 31(12): 3700-3715. https://doi.org/10.3969/j.issn.1000-9825.2020.12.003

[25] Kipf, T.N., Welling, M. (2016). Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907. https://doi.org/10.48550/arXiv.1609.02907

[26] He, X., Deng, K., Wang, X., Li, Y., Zhang, Y., Wang, M. (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, Virtual Event, China, pp. 639-648. https://doi.org/10.1145/3397271.3401063

[27] Rao, X., Chen, L., Liu, Y., Shang, S., Yao, B., Han, P. (2022). Graph-flashback network for next location recommendation. In Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Washington DC, USA, pp. 1463-1471. https://doi.org/10.1145/3534678.3539383

[28] Lim, N., Hooi, B., Ng, S.K., Goh, Y.L., Weng, R., Tan, R. (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, Madrid, Spain, pp. 1133-1143. https://doi.org/10.1145/3477495.3531989

[29] Li, X., Cong, G., Li, X.L., Pham, T.A.N., Krishnaswamy, S. (2015). Rank-geofm: A ranking based geographical factorization method for point of interest recommendation. In Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, Santiago, Chile, pp. 433-442. https://doi.org/10.1145/2766462.2767722

[30] Wu, S., Tang, Y., Zhu, Y., Wang, L., Xie, X., Tan, T. (2019). Session-based recommendation with graph neural networks. Proceedings of the AAAI Conference on Artificial Intelligence, 33(1): 346-353. https://doi.org/10.1609/aaai.v33i01.3301346

[31] Li, Y., Tarlow, D., Brockschmidt, M., Zemel, R. (2015). Gated graph sequence neural networks. arXiv preprint arXiv:1511.05493. https://doi.org/10.48550/arXiv.1511.05493

[32] Li, W., Zhang, C., Tang, W., Zeng, Z., Li, H. (2025). Multi-semantic hypergraph learning with fine-grained spatio-temporal information for next POI recommendation. Application Research of Computers, 42(2): 398-405. https://doi.org/10.19734/j.issn.1001-3695.2024.07.0288

[33] Lai, Y., Su, Y., Wei, L., He, T., et al. (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, Washington DC, USA, pp. 1452-1462. https://doi.org/10.1145/3626772.3657726

[34] Cao, Z., Zhu, Z. (2025). Gait recognition using disentangled representation learning based on information entropy. Computer Applications and Software, 42(4): 150-155,222. https://doi.org/10.3969/j.issn.1000-386x.2025.04.023

[35] Bengio, Y., Courville, A., Vincent, P. (2013). Representation learning: A review and new perspectives. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(8): 1798-1828. https://doi.org/10.1109/TPAMI.2013.50

[36] Ma, J., Zhou, C., Cui, P., Yang, H., Zhu, W. (2019). Learning disentangled representations for recommendation. arXiv preprint arXiv:1910.14238. https://doi.org/10.48550/arXiv.1910.14238

[37] Wang, X., Jin, H., Zhang, A., He, X., Xu, T., Chua, T.S. (2020). Disentangled graph collaborative filtering. In Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, Virtual Event, China, pp. 1001-1010. https://doi.org/10.1145/3397271.3401137

[38] Qin, Y., Wang, Y., Sun, F., Ju, W., et al. (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, Singapore, pp. 508-516. https://doi.org/10.1145/3539597.3570408

[39] Chen, T., Kornblith, S., Norouzi, M., Hinton, G. (2020). A simple framework for contrastive learning of visual representations. arXiv preprint arXiv:2002.05709. https://doi.org/10.48550/arXiv.2002.05709

[40] Zhang, R., Bian, Z. (2025). Overview of multimodal generation for recommender systems. Journal of Frontiers of Computer Science and Technology, 19(12): 3224-3242. https://doi.org/10.3778/j.issn.1673-9418.2511039

[41] Wu, J., Wang, X., Feng, F., He, X., Chen, L., Lian, J., Xie, X. (2021). Self-supervised graph learning for

recommendation. In Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, Virtual Event, Canada, pp. 726-735. https://doi.org/10.1145/3404835.3462862

[42] Yu, J., Xia, X., Chen, T., Cui, L., Hung, N.Q.V., Yin, H. (2023). XSimGCL: Towards extremely simple graph contrastive learning for recommendation. IEEE Transactions on Knowledge and Data Engineering, 36(2): 913-926. https://doi.org/10.1109/TKDE.2023.3288135

[43] Cho, E., Myers, S.A., Leskovec, J. (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, San Diego, California, USA, pp. 1082-1090. https://doi.org/10.1145/2020408.2020579

[44] Yuan, Q., Cong, G., Ma, Z., Sun, A., Thalmann, N.M. (2013). Time-aware point-of-interest recommendation. In Proceedings of the 36th International ACM SIGIR Conference on Research and development in Information Retrieval, Dublin, Ireland, pp. 363-372. https://doi.org/10.1145/2484028.2484030

[45] Zhang, J.D., Chow, C.Y. (2015). Geosoca: Exploiting geographical, social and categorical correlations for point-of-interest recommendations. In Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, Santiago, Chile, pp. 443-452. https://doi.org/10.1145/2766462.2767711

[46] Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., Yu, P.S. (2020). A comprehensive survey on graph neural networks. IEEE Transactions on Neural Networks and Learning Systems, 32(1): 4-24. https://doi.org/10.1109/TNNLS.2020.2978386

[47] Wang, X., He, X., Wang, M., Feng, F., Chua, T. S. (2019). Neural graph collaborative filtering. In Proceedings of the 42nd international ACM SIGIR Conference on Research and Development in Information Retrieval, Paris, France, pp. 165-174. https://doi.org/10.1145/3331184.3331267

[48] Hamilton, W.L., Ying, R., Leskovec, J. (2017). Inductive representation learning on large graphs. arXiv preprint arXiv:1706.02216. https://doi.org/10.48550/arXiv.1706.02216

[49] Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y. (2017). Graph attention networks. arXiv preprint arXiv:1710.10903. https://doi.org/10.48550/arXiv.1710.10903

[50] Wang, X., Wang, B., Wang, D., Wang, Z., Sun, G., Zhao, B., Chen, M. (2025). MultiPerG: Multiple Periodic Geography convolution for next POI recommendation. Vicinagearth, 2(1): 3. https://doi.org/10.1007/s44336-025-00012-1

[51] Wang, Y., Jiang, Q. (2025). Enhancing next POI recommendation via multi-graph modeling and multi-granularity contrastive learning. Journal of King Saud University Computer and Information Sciences, 37(10): 343. https://doi.org/10.1007/s44443-025-00325-7

[52] Rao, X., Shang, S., Chen, L., Jiang, R., Han, P. (2025). Disentangled and personalized representation learning for next point-of-interest recommendation. In Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence, Montreal, Canada, pp. 7697-7705. https://doi.org/10.24963/ijcai.2025/856

[53] Oord, A.V.D., Li, Y., Vinyals, O. (2018). Representation learning with contrastive predictive coding. arXiv preprint arXiv:1807.03748. https://doi.org/10.48550/arXiv.1807.03748

[54] Lei, Y., Shen, L., Sun, Z., He, T., Feng, S., Liu, G. (2025). COSTA: Contrastive Spatial and Temporal Debiasing framework for next POI recommendation. Neural Networks, 185: 107212. https://doi.org/10.1016/j.neunet.2025.107212

[55] Rendle, S., Freudenthaler, C., Gantner, Z., Schmidt-Thieme, L. (2012). BPR: Bayesian personalized ranking from implicit feedback. arXiv preprint arXiv:1205.2618. https://doi.org/10.48550/arXiv.1205.2618

[56] Kingma, D.P., Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980. https://doi.org/10.48550/arXiv.1412.6980

[57] Rahmani, H.A., Aliannejadi, M., Ahmadian, S., Baratchi, M., Afsharchi, M., Crestani, F. (2019). LGLMF: local geographical based logistic matrix factorization model for POI recommendation. In Information Retrieval Technology: 15th Asia Information Retrieval Societies Conference, AIRS 2019, Hong Kong, China, pp. 66-78. https://doi.org/10.1007/978-3-030-42835-8_7