Entropy-oriented Multi-view Representation Fusion With Wavelet Graph Diffusion For Next POI Recommendation

Rong Wang[1], Yingcun Wang[2*], Jinfeng Wang[2], and Jiajia Ma[2]

[1] *School of International Business, Weifang Vocational College, Weifang 262737, China*
[2] *School of Information Engineering, Weifang Vocational College, Weifang, 261031, China*
[*] *Corresponding author. E-mail: wangyingcun1979@163.com*

Received: Jun. 05, 2026; Accepted: Jul. 13, 2026

Next point-of-interest (POI) recommendation is a critical task in location-based social networks, offering significant value for personalized services and smart city governance. However, existing methods still struggle with the extreme scale of candidate search spaces and the inherent sparsity of check-in data. In this paper, we propose EMRF-WGD, an Entropy-oriented Multi-view Representation Fusion framework with Wavelet Graph Diffusion. To capture heterogeneous mobility dependencies, we first design a multi-view feature extraction module that jointly encodes POI transitions, user preferences, geographical regions, and temporal periods. Then, to effectively aggregate information, we introduce an entropy-aware view fusion strategy grounded in information theory, which maximizes mutual information while suppressing view-specific noise through variational approximation. Furthermore, a hybrid decoding framework is implemented to jointly predict the next POI, time, and region, supplemented by a spatiotemporal heuristic refinement stage that leverages historical relational tensors to narrow the search space. Extensive experiments on the NYC dataset demonstrate that EMRF-WGD consistently outperforms state-of-the-art baselines across all evaluation metrics, including Acc@k and MRR.

Keywords: Spatio-temporal partition; wavelet graph neural network; entropy fusion

© The Author('s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are cited.

http://dx.doi.org/10.6180/jase.202612_35.039

1. Introduction

Against the backdrop of the rapid development of contemporary mobile internet, location-based social networks have become an essential component of the digital social infrastructure [1, 2]. Applications such as Foursquare and Yelp have not only accumulated massive amounts of user check-in data but also recorded human behavioral trajectories and social interactions in physical space [3, 4]. Supported by these rich multidimensional datasets, next point-of-interest recommendation has evolved from a mere technical endeavor into a research topic of profound impact and significant challenge [5–8]. From the perspective of user experience, accurate POI prediction can markedly alleviate information overload [9]. By learning users' historical preferences and current spatio-temporal context, it enables highly personalized services, such as precise dining, entertainment, or travel suggestions. However, its significance extends far beyond commercial recommendation. At the broader level of smart city governance, by predicting population movement intentions, decision-makers can dynamically anticipate urban traffic flows, optimize the allocation of public resources, and even achieve more scientifically grounded risk prevention and control planning during public health events [10]. Therefore, extracting deep movement features from complex spatio-temporal sequence data and transforming them into accurate predictions of future behavior has become a core task in the fields of computer science and spatio-temporal data mining. Although this area has attracted substantial research attention, achieving

highly reliable predictions in dynamic urban environments remains a critical scientific challenge that demands urgent resolution [11].

Although next POI recommendation has made considerable progress driven by deep learning algorithms, existing models still face two major bottlenecks in practical deployment. The primary challenge lies in the extreme scale and dispersion of the search space. In a typical modern metropolis, the number of candidate POIs often reaches tens or even hundreds of thousands [12–14]. If a recommendation system merely searches through such an enormous candidate pool based solely on a user's historical sequence, the signal-to-noise ratio of the predictions becomes extremely low [15, 16]. Due to the large cardinality of the candidate set, even if the model captures certain semantic regularities, they are easily overwhelmed by a vast number of negative samples, resulting in a bottleneck in prediction accuracy. This greatly limits the usability of recommendation systems in real-world scenarios. In addition to the spatial dimension challenge, data sparsity is also a persistent problem. Although human mobility patterns exhibit regularity, they show pronounced individual-specific spatio-temporal clustering effects [17]. Specifically, users tend to move frequently between geographically close locations or exhibit high visit concentration within specific time windows (e.g., mainly appearing in parks and office areas in the early morning). However, individual users' personal check-in records are usually very limited and fragmented. Relying solely on these fragmented historical trajectories, it is difficult for models to accurately capture such deep spatio-temporal clustering characteristics. Without effectively modeling this clustering effect and narrowing the search scope, recommendation systems will continue to search inefficiently "in the vast ocean," making truly intelligent decision-making unattainable [18].

In this paper, we propose an entropy-oriented multi-view representation fusion framework with wavelet graph diffusion for next POI recommendation. To comprehensively capture the complex, heterogeneous dependencies in user mobility, we first design a multi-view feature extraction module that jointly encodes POI transitions, user preferences, geographical regions, and temporal periods. Specifically, a global POI transition graph is processed by a graph wavelet neural network to capture multi-scale spatial diffusion patterns, while user representations are dynamically constructed via relational attention over a user-POI bipartite graph. Region and time-period representations are similarly derived through clustering and embedding techniques. To effectively aggregate complementary information from these diverse views, we introduce an entropy-aware view fusion strategy that minimizes the conditional entropy between the fused representation and each view-specific embedding, thereby preserving maximal mutual information. The resulting fused sequence is then fed into a hybrid decoding module consisting of a shared Transformer block followed by three task-specific MLP heads, which jointly predict the next POI, the next time period, and the next geographical region. Finally, a refinement stage leverages historical user-period-POI and user-region-POI relational tensors to filter and adjust the prediction, mitigating spatial noise and error propagation. Through this integrated design, our model achieves robust and accurate next-POI recommendation under real-world sparsity and large-scale candidate space constraints.

The main contributions of this paper are threefold:

- We propose a multi-view feature extraction module that jointly models POI transitions, user preferences, geographical regions, and temporal periods. In particular, a graph wavelet neural network is employed on the global POI transition graph to capture multi-scale spatial diffusion patterns, while user representations are dynamically constructed via a relational attention mechanism over a user-POI bipartite graph. This design effectively captures the heterogeneous and complex dependencies inherent in location-based social networks.

- To integrate complementary information from different views, we introduce an entropy-aware view fusion strategy grounded in information theory. By minimizing the conditional entropy between the fused representation and each view-specific representation, the proposed method maximizes the mutual information preserved from individual views. A variational approximation with a Gaussian parameterization is

  adopted to make the optimization tractable, leading to a compact and semantically robust fused representation of user trajectories.

- We design a hybrid decoding framework that consists of a shared Transformer block followed by three task-specific MLP heads, enabling joint prediction of the next POI, the next time period, and the next geographical region. Furthermore, a refinement stage leverages historical user-period-POI and user-region-POI relational tensors to filter and adjust the initial predictions. This spatiotemporal heuristic mechanism mitigates the issues of excessive spatial noise and error propagation, substantially improving prediction accuracy under large-scale candidate spaces and data sparsity.

2. Method

2.1. Problem Formulation

Let $\mathrm{V} = \{v_1, v_2, \ldots, v_N\}$ denote a set of N distinct Points-of-Interest (POIs), and $\mathrm{U} = \{u_1, u_2, \ldots, u_M\}$ represent a set of M unique users. The core components of the next POI recommendation task are formalized as follows:

- Check-in event: A check-in instance is defined as a triplet $e = (u, v, \tau)$, signifying that user $u \in \mathrm{U}$ visited location $v \in \mathrm{V}$ at a specific timestamp τ.
- User visit sequence: The historical trajectory of a user u_i is represented as a chronological sequence of check-in events, denoted by $\mathrm{S}_{u_i} = \{e_1, e_2, \ldots, e_k\}$, where k is the total number of historical visits in the sequence.
- Next POI recommendation: Given a prefix trajectory S_{u_i} of user u_i up to step k, the objective is to predict the most probable POI v_{k+1} that the user will visit at the subsequent timestamp τ_{k+1}. Formally, the task aims to generate a ranked list of candidate POIs from V based on the conditional probability $P(v_{k+1} \mid \mathrm{S}_{u_i})$.

To achieve the above goal, an entropy-oriented multi-view representation fusion with wavelet graph diffusion is proposed for next POI recommendation, which contains a multi-view feature extraction, an entropy-aware view fusion, and a representation hybrid decoding, as shown in Fig. 1.

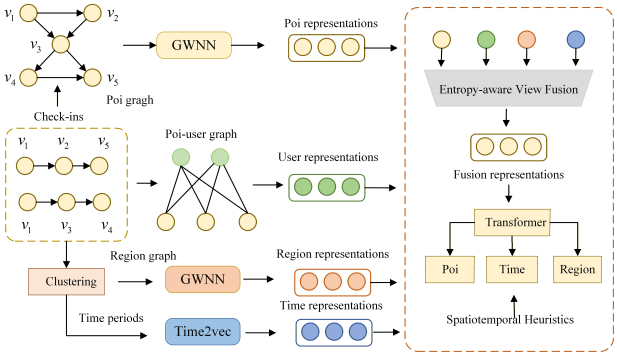


Fig. 1. The illustration of the proposed method

2.2. Multi-view Feature Extraction

To comprehensively capture the complex and heterogeneous dependencies in location-based social networks, we propose a multi-view feature extraction module. This module systematically encodes check-in dynamics from four interconnected views: POI representations, user representations, region representations, and time period representations.

POI representations: We represent the collective movement patterns as a global POI transition graph $\mathrm{G}_{poi} = \left(\mathrm{V}, \mathrm{E}_{poi}\right)$. Each node $v \in \mathrm{V}$ corresponds to a unique POI. A directed edge $h^{ij}_{poi} \in \mathrm{E}_{poi}$ is established if any user sequentially visits v_i and v_j within a predefined session window. To surpass the localized receptive field of standard graph convolutional networks, we utilize a graph wavelet neural network. The spectral wavelet transform of the i-th layer for a POI representations Z_v^{l-1} is formulated as:

$$Z_v^l = \psi_s \left(\sum_{k=0}^{K} \theta_k (s\Lambda)^k \right) \psi_s^{-1} Z_v^{l-1} \tag{1}$$

where $\psi_s = \mathbf{U} e^{s\Lambda} \mathbf{U}^T$ denotes the graph wavelet basis at scale s. Here, U is the eigenvector matrix derived from the graph Laplacian $\mathbf{L} = \mathbf{D}_{poi} - \mathrm{E}_{poi}\mathbf{D}_{poi}$ is the diagonal matrix. $ Lambda$ is the diagonal matrix of eigenvalues. s is a scale parameter that controls the diffusion of the wavelet. The term θ_k represents the learnable filter coefficients of the K-order Chebyshev polynomial, which defines the receptive field. The output of graph wavelet neural network is denoted as Z_v.

User representations: User preferences are modeled via a user-POI bipartite graph $\mathrm{G}_u = (\mathrm{U} \cup \mathrm{V}, \mathrm{E}_u)$. An undirected edge $h^{mn}_u \in \mathrm{E}_u$ exists if user $u_m \in \mathrm{U}$ has interacted with POI $v_n \in \mathrm{V}$. Instead of static ID-based representations, we dynamically construct user representations Z_u by aggregating the semantic features of their historical footprint using a relational attention mechanism:

$$Z_{u_m} = \sigma \left(\mathbf{W}_u \cdot \left[Z^{init}_{u_m} \square \sum_{j \in \mathrm{N}(u_m)} \alpha_{mj} Z_{v_j} \right] \right) \tag{2}$$

where σ is the activation function. $\mathbf{W}_u$ is a weight matrix for feature transformation. $Z^{\mathrm{init}}_{u_m}$ is the initial trainable ID embedding of user u_m $\cdot$ $[\cdot]$ denotes the concatenation function. The weight α_{mj} is computed via a compatibility function between the users and the POIs, ensuring that the user embedding adaptively reflects the most relevant historical interests. $\mathrm{N}(u_m)$ denotes the set of POIs visited by the user.

Region representations: To capture the spatial patterns of human mobility, we categorize POIs into distinct regions, i.e., geographical zones containing clusters of nearby locations. This categorization relies on the premise that proximal POIs are more likely to be visited in succession. Specifically, we employ the k-means algorithm on coordinate data to partition POIs into a set of regions G. By constructing a region-level graph and applying the graph wavelet neural network, we derive region representations Z_r.

Time period representations. To model temporal dynamics, we discretize a day into 20-minute intervals (totaling 72 slices), each populated with check-in records from global trajectories. These intervals are grouped into a set of periods T via ordered sample clustering. For vector representation, we map these periods into an embedding space, denoted by Z_t, leveraging the time2vec approach.

2.3. Entropy-aware View Fusion

After gaining the representations for POI, user, region, and time period from the check-in events, we introduce a entropy-aware view fusion strategy to full aggregation complementary information scattered in various views with the help of the information theory, which learns a semantic-robust fusion representations of history trajectories for prediction next POI.

Specifically, given the representations for POI, user, region, and time period, Z_v, Z_u, Z_r, and Z_t, we obtain fusion representations via the concatenation function, i.e., $Z = [Z_v, Z_u, Z_r, Z_t]$. Thus, each input trajectory $\{e_1, e_2, \ldots, e_k\}$ can be represented by $\{Z_1, Z_2, \ldots, Z_k\}$. And then, we minimize the conditional entropy between fusion representations and view-specific representations to ensure effective information aggregation between views, as follows:

$$H\left(Z_* \mid Z\right) = -\mathrm{E}_{\mathrm{P}_{\hat{Z},Z}}\left[\log \mathrm{P}\left(Z_* \mid Z\right)\right] \tag{3}$$

where $* \in \{v, u, r, t\}$. $H(\cdot)$ is the entropy function. Minimizing the conditional entropy ensures that fusion representations retain maximal mutual information from each individual view. However, direct computation of $P\left(Z_* \mid Z\right)$ is often intractable in high-dimensional latent spaces. Following the principles of variational inference, we introduce a variational distribution $Q\left(Z_* \mid Z\right)$ to approximate the true posterior. This transforms the problem into maximizing the lower bound of the log-likelihood:

$$\max \mathrm{E}_{\mathrm{P}_{*,Z}}\left[\log \mathrm{P}\left(Z_* \mid Z\right)\right] \Leftrightarrow \mathrm{E}_{\mathrm{P}_{*,Z}}\left[\log Q\left(Z_* \mid Z\right)\right] \tag{4}$$

Specifically, we parameterize Q as a Gaussian distribution $\mathrm{N}\left(Z_* \mid G_*(Z), \sigma\mathbf{I}\right)$, where $G_*(\cdot)$ denotes a cross-view mapping function designed to recover view-specific features from the fusion state, and $\sigma\mathbf{I}$ represents the isotropic variance. Under this Gaussian assumption, the objective becomes:

$$\max\ \mathbb{E}_{P_{Z_*,Z}}\left[\log Q(Z_* \mid Z)\right] = \mathbb{E}_{P_{Z_*,Z}}\left[\log\left(\frac{1}{\sqrt{2\pi\sigma\mathbf{I}}} \times e^{-\frac{(Z_* - G_*(Z))^2}{2\sigma\mathbf{I}}}\right)\right]. \tag{5}$$

By discarding the constant log-partition function and the scaling factor 2σ, the maximization problem is equivalent to minimizing the expected reconstruction error:

$$\max -\mathrm{E}_{\mathrm{P}_{Z_*,Z}}\left\|\left(Z_* - G_*(Z)\right)\right\|_2^2. \tag{6}$$

Hence, the entropy-aware view fusion loss is accordingly defined as:

$$L_e = \sum_{*}^{\{v,u,r,t\}} \left\|Z_* - G_*(Z)\right\|_2^2. \tag{7}$$

2.4 Representation Hybrid Decoding

Representation hybrid decoding contains a sharing Transformer block and specific MLP block for achieving prediction if the next POI, the next time period, and the next region. For an input trajectory of length k, we represent the sequence by stacking check-in fusion representations with added positional encodings, forming the initial tensor X, and then input the Transformer block that consists of a multi-head self-attention (MHSA) module and a feed-forward network (FFN):

$$X^t = \mathrm{FFN}(\mathrm{MHSA}(X)) \tag{8}$$

where X^t is the output of the Transformer block.

Afterward, we split X^t into three specialized branches. Each branch utilizes an MLP to map the high-dimensional latent space to the respective label space. To predict the specific next Poi (Y_v), the MLP maps X^t to a probability distribution over the set of all POIs V.

$$Y_v = \mathrm{Softmax}\left(\mathbf{W}_v X^t + \mathbf{b}_v\right) \tag{9}$$

where $\mathbf{W}_v$ and $\mathbf{b}_v$ are learnable parameters. To predict the categorical time period Y_t of the next check-in:

$$Y_t = \mathrm{Softmax}\left(\mathbf{W}_t X^t + \mathbf{b}_t\right) \tag{10}$$

where $\mathbf{W}_t$ and $\mathbf{b}_t$ are learnable parameters. To predict the geographical region Y_r the user will enter:

$$Y_r = \mathrm{Softmax}\left(\mathbf{W}_r X^t + \mathbf{b}_r\right) \tag{11}$$

where $\mathbf{W}_r$ and $\mathbf{b}_r$ are learnable parameters. The optimization of our model is guided by a composite objective. We employ the cross-entropy criterion to calculate the individual losses for three forecasting tasks: next-POI prediction $\left(L_{\text{poi}}\right)$, temporal period estimation $\left(L_{\text{time}}\right)$, and geographical region identification $\left(L_{\text{region}}\right)$. The overall objective, L_{total}, is formulated as follows:

$$L_{\text{total}} = L_{\text{poi}} + \alpha L_{\text{time}} + \beta L_{\text{region}} + \lambda L_e \tag{12}$$

where α, β, and λ are trade-off parameters. Refinement via Spatiotemporal Heuristics. Although the model generates predictions for regions and time intervals, directly masking POIs based on these outputs often yields suboptimal results due to excessive spatial noise and the risk of error propagation. To address this, we introduce a secondary adjustment phase that leverages historical user behavior. Specifically, we construct two relational tensors, i.e., user-period-POI and user-region-POI, to quantify long-term preferences. By applying a filtering threshold, e.g., selecting the top 40 regions and top 5 time periods, we extract relevant preference vectors. These historical signals are then integrated with the initial MLP-based POI probabilities through element-wise operations, producing a final prediction that aligns with both current context and established user habits.

3. Experiment

3.1. Setup

Dataset: NYC is a widely used check-in dataset collected from New York City, spanning the period from April 2012 to February 2013. NYC comprises 106919 check-in records generated by 1083 unique users across 5132 POIs. After preprocessing and sequence segmentation, a total of 32541 trajectories are extracted. In the experiments, to ensure data quality, we filter the dataset by removing POIs with fewer than 10 check-ins and users with fewer than 20 records. For each user, chronological sequences are partitioned into training, validation, and test sets with an 80/10/10 ratio. Trajectories are segmented based on a 24-hour window, with each day's activity defined as a single sequence. To mitigate the impact of excessively long sequences, we truncate trajectories at a maximum length of 50 , retaining only the most recent check-ins.

Metric: To evaluate the performance of next POI recommendation, we employ two standard metrics: Accuracy@k (Acc@k) and Mean Reciprocal Rank (MRR). Acc@k measures the proportion of instances where the ground-truth POI is successfully captured within the top-k recommendation slots. MRR accounts for the specific ranking position of the correct prediction. It computes the average of the reciprocal ranks for the target POIs across multiple test samples.

Comparison Methods: To demonstrate the superiority of the proposed model, we conduct comparative experiments against several state-of-the-art baselines, categorized as follows: LDPTRec [14], MF [15], FPMC [16], GETNext [17], AGRAN [18], and GUGEN [19].

Implementation details. All experiments are conducted on a single NVIDIA Geforce RTX 3090 GPU. In the experiments, the dimension numbers of POI representations, user representations, region representations, and time period representations are set as 128, 128, 64 and 64, respectively. In the Transformer, the dimensions of the feed forward network and the number of attention heads are set as 1024 and 2, respectively. The overall network is training via the Adam optimizer with the learning rate of 0.001, the batch size of 64, and the epoch number of 100.

3.2. Comparisons with State-of-the-arts

The performance of our proposed EMRF-WGD is evaluated against several competitive baselines. As illustrated in Table 1, EMRF-WGD consistently achieves the superior performance across all evaluation metrics, including accuracy at various thresholds and MRR. The consistent lead of EMRF-WGD can be attributed to three critical technical innovations: First, the multi-view feature extraction with wavelet graph diffusion captures multi-scale spatial dependencies that conventional graph convolutional networks often miss. By operating on the global POI transition graph with a spectral wavelet basis, our model effectively encodes both local and long-range mobility patterns, leading to more discriminative POI and region representations. This provides a clear advantage over baselines that rely solely on sequential or single-scale graph encodings. Second, the entropy-aware view fusion strategy minimizes conditional entropy between the fused trajectory representation and each view-specific embedding. Rooted in information theory, this approach maximizes the mutual information preserved from the POI, user, region, and time period views while suppressing view-specific noise. As a result, the fused representation becomes semantically robust and more predictive of future check-ins, which is reflected in consistently higher accuracy and MRR compared to methods that use simple concatenation or heuristic weighting. Third, the hybrid decoding with spatiotemporal heuristic refinement jointly predicts the next POI, time period, and geographical region, and then refines the initial POI distribution using historical user-period-POI and user-region-POI relational tensors. This two-stage mechanism directly addresses the two major bottlenecks discussed in the introduction: it narrows the enormous candidate search space by leveraging predicted regions and time periods, and it mitigates data sparsity by incorporating long-term preference heuristics. Consequently, EMRF-WGD achieves superior performance across all evaluation metrics, demonstrating that the integration of multi-scale spatial modeling, information-theoretic fusion, and spatiotemporal refinement yields a synergistic effect that pushes the state of the art in next POI recommendation.

Table 1. Comparison results (Mean ± Std 1%) with different models

Method	Acc@1	Acc@5	Acc@10	Acc@20	MRR
LDPTRec	24.12 ± 0.16	52.45 ± 0.31	62.55 ± 0.29	72.12 ± 0.37	40.12 ± 0.19
MF	17.45 ± 1.12	31.20 ± 1.25	38.55 ± 1.27	45.33 ± 1.34	22.15 ± 1.12
FPMC	18.62 ± 0.65	35.88 ± 0.71	42.12 ± 0.78	49.55 ± 0.85	25.33 ± 0.52
GETNext	22.33 ± 0.52	42.15 ± 0.60	51.45 ± 0.83	60.12 ± 0.72	32.45 ± 0.68
AGRAN	22.45 ± 0.20	49.33 ± 0.35	59.12 ± 0.30	68.45 ± 0.40	38.66 ± 0.21
GUGEN	23.15 ± 2.18	45.62 ± 2.28	55.33 ± 2.25	64.22 ± 2.38	35.12 ± 2.15
EMRF-WGD	26.12 ± 0.18	55.95 ± 0.32	65.84 ± 0.27	75.94 ± 0.41	42.01 ± 0.22

3.3. Ablation Analysis

To verify the effectiveness of the individual components and the loss functions within EMRF-WGD, we conduct a series of ablation experiments on the NYC dataset.

Impact of model components: We first investigate the contribution of the core architectural modules by evaluating three variants: EMRF-WGD w/o WGD replaces the graph wavelet neural network with a standard graph convolutional network to test the necessity of multi-scale spatial diffusion patterns. EMRF-WGD w/o MV removes the multi-view feature extraction, relying solely on the POI transition graph. EMRF-WGD w/o SH removes the spatiotemporal heuristic refinement stage, using only the raw outputs from the MLP heads for prediction The results in Table 2 demonstrate that multi-view extraction is the most critical component, as its removal leads to the sharpest decline in performance. This confirms that capturing heterogeneous dependencies (user preferences, regions, and time) is essential for overcoming data sparsity. Furthermore, the WGD module's ability to capture long-range mobility patterns proves superior to localized graph convolutional networks.

Impact of loss functions: To evaluate the contribution of each objective function to the overall performance of EMRF-WGD, we conduct a detailed ablation study by systematically removing specific loss components. We compare the full EMRF-WGD model against four variants, each excluding a primary loss term, i.e., EMRF-WGD w/o L_{poi} , EMRF-WGD w/o L_{time} , EMRF-WGD w/o L_{region} , and EMRF-WGD w/o L_e. As illustrated in Table 2, the full EMRF-WGD model achieves the superior performance across all evaluation metrics, demonstrating that the integration of these four objectives yields a synergistic effect. The most drastic performance decline is observed upon the removal of L_{poi}, as this loss provides the fundamental supervisory signal required to map trajectories to specific POI identities. The degradation resulting from the exclusion of L_{time} and L_{region} highlights that joint prediction of auxiliary spatio-temporal labels helps the model capture deep clustering effects and narrows the search scope within large-scale candidate spaces. Furthermore, the significant drop in metrics for the variant without L_e confirms that the entropy-aware strategy is essential for maximizing the mutual information preserved from diverse views while suppressing view-specific noise. Collectively, these observations suggest that while L_{poi} ensures task-specific alignment, the auxiliary and fusion losses are critical for enhancing representation robustness and mitigating the challenges of data sparsity.

3.4. Discussion and Limitations

Although the proposed EMRF-WGD framework achieves encouraging performance in next POI recommendation, several limitations should be acknowledged. First, the current experimental validation is mainly conducted on the NYC check-in dataset. While this dataset is widely used and can reflect typical urban mobility patterns, experiments on a single city may not fully demonstrate the generalization ability of the proposed model across different geographical regions, population densities, and mobility behaviors. In future work, we will further evaluate EMRF-WGD on additional public check-in datasets, such as Tokyo or Gowalla, to verify its robustness under cross-city and cross-dataset settings. Second, the spatio-temporal refinement strategy introduces prior knowledge based on user-period-POI and user-region-POI tensors. This design can effectively narrow the candidate search space and alleviate data sparsity, but it also relies on empirically selected thresholds, such as the number of candidate regions and active time periods. Although these thresholds are effective in our experiments, they may introduce heuristic bias when user mobility patterns vary significantly across datasets. Therefore, an adaptive threshold selection mechanism or a learnable refinement strategy could be explored to further improve the flexibility of the model. Finally, the graph wavelet diffusion mechanism enhances high-order structural representation learning, but it may introduce additional computational cost when the user-POI graph becomes very large. Although the current implementation remains feasible on the adopted dataset, improving the scalability of EMRF-WGD for large-scale industrial recom-

Table 2. Ablation Analysis of the loss functions and individual components

Variant	Acc@1	Acc@10	MRR
EMRF-WGD w/o WGD	24.55	62.10	39.42
EMRF-WGD w/o MV	22.18	58.45	39.42
EMRF-WGD w/o SH	23.40	60.22	36.80
EMRF-WGD w/o L_{poi}	8.55	35.74	20.36
EMRF-WGD w/o L_{time}	25.96	64.35	41.84
EMRF-WGD w/o L_{region}	24.22	62.73	40.02
EMRF-WGD w/o L_e	22.77	61.37	38.23
EMRF-WGD	26.12	65.84	42.01

mendation scenarios is still an important future direction. Overall, these limitations do not undermine the effectiveness of the proposed framework, but they provide valuable directions for further improving its generalization, adaptability, and practical applicability.

4. Conclusions

In this paper, we proposed EMRF-WGD, an entropy-oriented multi-view representation fusion framework with wavelet graph diffusion for next POI recommendation. To address the challenges of extreme search space and data sparsity, we designed a multi-view feature extraction module that jointly encodes POI transitions, user preferences, geographical regions, and temporal periods. By employing a graph wavelet neural network, the model effectively captures multi-scale spatial diffusion patterns that conventional graph convolutional networks often miss.Furthermore, we introduced an entropy-aware view fusion strategy grounded in information theory to aggregate complementary information from diverse views while suppressing noise. A hybrid decoding framework, supplemented by a spatiotemporal heuristic refinement stage, was implemented to narrow the candidate search space and mitigate error propagation. Experimental results on the NYC dataset demonstrate that EMRF-WGD consistently outperforms state-of-the-art baselines across all evaluation metrics.

References

[1] P. Zhao, A. Luo, Y. Liu, J. Xu, Z. Li, F. Zhuang, V. S. Sheng, and X. Zhou, (2022) *"Where to Go Next: A Spatio-Temporal Gated Network for Next POI Recommendation"* **IEEE Transactions on Knowledge and Data Engineering** ***34***(5): 2512–2524. DOI: 10.1109/TKDE.2020.3007194.

[2] H. Li and C. Zhao, (2024) *"Fusion Cycle GAN: a multiple feature fusion based cycle-consistent generative adversarial network for person re-identification"* **IFS/ACM Transactions on Machine Learning** ***1***(1): 7–12. DOI: 10.70891/JSE.2024.100006.

[3] J. Gao, M. Liu, P. Li, A. A. Laghari, A. R. Javed, N. Victor, and T. R. Gadekallu, (2024) *"Deep Incomplete Multiview Clustering via Information Bottleneck for Pattern Mining of Data in Extreme-Environment IoT"* **IEEE Internet of Things Journal** ***11***(16): 26700–26712. DOI: 10.1109/JIOT.2023.3325272.

[4] J. Gao, M. Liu, P. Li, J. Zhang, and Z. Chen, (2024) *"Deep Multiview Adaptive Clustering With Semantic Invariance"* **IEEE Transactions on Neural Networks and Learning Systems** ***35***(9): 12965–12978. DOI: 10.1109/TNNLS.2023.3265699.

[5] S. Yin, L. Wang, T. Chen, H. Huang, J. Gao, J. Zhang, M. Liu, P. Li, and C. Xu, (2026) *"LKAFormer: A lightweight kolmogorov-arnold transformer model for image semantic segmentation"* **ACM Transactions on Intelligent Systems and Technology** ***17***(3): 1–24. DOI: 10.1145/3759254.

[6] Z. Zhan, X. Mao, H. Liu, and S. Yu, (2025) *"STGL: Self-Supervised Spatio-Temporal Graph Learning for Traffic Forecasting"* **Journal of Artificial Intelligence Research** ***2***(1): 1–8. DOI: 10.70891/JAIR.2025.040001.

[7] W. Liu, (2024) *"Channel Reorganization for Few-Shot Segmentation"* **Journal of Artificial Intelligence Research** ***1***(1): 36–40. DOI: 10.70891/JAIR.2024.100025.

[8] W. Zhang and J. Wang, (2024) *"English text sentiment analysis network based on CNN and U-Net"* **IFS/ACM Transactions on Machine Learning** ***1***(1): 13–18. DOI: 10.70891/JSE.2024.100009.

[9] F. Yin, Y. Liu, Z. Shen, L. Chen, S. Shang, and P. Han. "Next POI recommendation with dynamic graph and explicit dependency". In: *Proceedings of the AAAI conference on artificial intelligence*. **37**. 4. 2023, 4827–4834. DOI: 10.1609/aaai.v37i4.25608.

[10] Z. Wang, Y. Zhu, Q. Zhang, H. Liu, C. Wang, and T. Liu, (2022) *"Graph-enhanced spatial-temporal network for next POI recommendation"* **ACM Transactions on Knowledge Discovery from Data (TKDD)** ***16***(6): 1–21. DOI: 10.1145/3513092.

[11] W. Chen, H. Wan, S. Guo, H. Huang, S. Zheng, J. Li, S. Lin, and Y. Lin, (2022) *"Building and exploiting spatial–temporal knowledge graph for next POI recommendation"* **Knowledge-Based Systems** ***258***: 109951. DOI: 10.1016/j.knosys.2022.109951.

[12] X. Sun, Z. Gao, D. Yu, and B. Huang, (2025) *"Joint Modeling of Multimodal Information Based on Dynamic and Static Knowledge Graphs for Next POI Recommendation"* **IEEE Transactions on Computational Social Systems** ***12***(6): 3985–3999. DOI: 10.1109/TCSS.2025.3567512.

[13] Y. Xie, L. Wang, W. Nie, C. Shi, X. Dong, S. Chen, H. Sun, W. Rao, and A.-A. Liu, (2026) *"Next Chain Prediction: A Generative Recommendation Model With Sequence-Chain Attention"* **IEEE Transactions on Multimedia** ***28***: 321–332. DOI: 10.1109/TMM.2025.3623521.

[14] Y. Huang, X. Liu, Y. Miao, J. Zhang, and D. Cheng, (2025) *"LDPTRec: A Differential Privacy Based Transformer Framework for Next POI Recommendation"* **IEEE Transactions on Dependable and Secure Computing** ***22***(6): 7043–7059. DOI: 10.1109/TDSC.2025.3594398.

[15] Y. Koren, R. Bell, and C. Volinsky, (2009) *"Matrix Factorization Techniques for Recommender Systems"* **Computer** ***42***(8): 30–37. DOI: 10.1109/MC.2009.263.

[16] S. Rendle, C. Freudenthaler, and L. Schmidt-Thieme. "Factorizing personalized markov chains for next-basket recommendation". In: *Proceedings of the 19th international conference on World wide web*. 2010, 811–820. DOI: 10.1145/1772690.1772773.

[17] S. Yang, J. Liu, and K. Zhao. "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*. New York, NY, USA: Association for Computing Machinery, 2022, 1144–1153. DOI: 10.1145/3477495.3531983.

[18] Z. Wang, Y. Zhu, C. Wang, W. Ma, B. Li, and J. Yu. "Adaptive Graph Representation Learning for Next POI Recommendation". In: *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*. New York, NY, USA: Association for Computing Machinery, 2023, 393–402. DOI: 10.1145/3539618.3591634.

[19] C. Zuo, X. Zhang, L. Yan, and Z. Zhang, (2024) *"GUGEN: Global User Graph Enhanced Network for Next POI Recommendation"* **IEEE Transactions on Mobile Computing** ***23***(12): 14975–14986. DOI: 10.1109/TMC.2024.3455107.