SemPOI-RL: Aligning LLM Semantic Reasoning for Interpretable Out-of-Town POI Sequential Generation

Anonymous ACL submission

Abstract

Large language models (LLMs) exhibit strong semantic reasoning and open-ended generation abilities, but aligning these abilities with structured sequential generation remains challenging. This challenge is particularly evident in out-of-town (OOT) POI sequence generation, where a model must infer transferable travel intent from a user's hometown behaviors, adapt to cross-city interest drift, and generate a coherent destination trajectory under structural constraints. Existing approaches either rely on latent ID-based transfer with limited interpretability or directly use LLMs for sequence generation without explicitly grounding inferred semantics into position-aware predictions. To address this gap, we propose SemPOI-RL, a framework that aligns LLM semantic reasoning with structured sequence generation for interpretable OOT recommendation. Specifically, we first fine-tune an LLM to infer destination-oriented travel styles from users' hometown trajectories, using natural language as an interpretable semantic intermediate. We then introduce a Semantic POI Alignment Module (SPAM) to ground these inferred styles into a style-conditioned masked autoencoder for position-aware trajectory generation. Finally, we apply reinforcement learning with recommendation-oriented rewards to align LLM-generated styles with downstream sequence quality. Experiments on two real-world datasets show that SemPOI-RL consistently outperforms both traditional recommenders and direct LLM baselines, while providing interpretable style attribution across different phases of a trip. The code is available at https://anonymous.4open.science/r/SemPOI-RL-0495.

1 Introduction

Large language models (LLMs) have shown impressive capabilities in semantic abstraction, contextual reasoning, and open-ended generation, making them a promising foundation for structured generation tasks (Xie et al., 2024; Singh et al., 2024; Tang et al., 2024). However, a key challenge remains unresolved: how to align high-level semantic reasoning in natural language with downstream sequence generation that must satisfy domain-specific structural constraints (Wang et al., 2024; Ning et al., 2025). This challenge is especially salient in out-of-town (OOT) POI sequence generation, where a model must infer a user's transferable travel intent from hometown behaviors and then generate a coherent destination trajectory in a new city. OOT POI sequence generation differs fundamentally from conventional next-POI recommendation. In standard intra-city settings, models often rely on repeated user-POI interactions, transition regularities, or region-specific spatio-temporal patterns. In contrast, the OOT setting requires the model to generalize from a user's hometown trajectory to an unfamiliar destination city, where direct POI overlap is limited and user interests may shift substantially across regions (Feng et al., 2024; Wang et al., 2025; Yin et al., 2016). Moreover, the goal is not merely to predict the next location, but to generate an ordered sequence of destination POIs that is both semantically suitable and structurally coherent. This makes OOT recommendation a particularly difficult form of structured sequence generation under preference transfer (Liu et al., 2022; Li et al., 2024; Shu et al., 2024).

Existing sequential recommendation and trajectory forecasting methods predominantly rely on latent ID-based representations or spatio-temporal graph encoders (Yang et al., 2022; Yan et al., 2023; Zhang et al., 2025). While effective for intra-city next-POI prediction, these approaches treat sequences as flat lists of discrete tokens, offering limited interpretability and struggling to model cross-domain interest drift when users travel from familiar hometowns to novel destinations (Xin et al., 2022; Liu et al., 2025). Recent efforts have begun integrating LLMs into mobility prediction by lever-

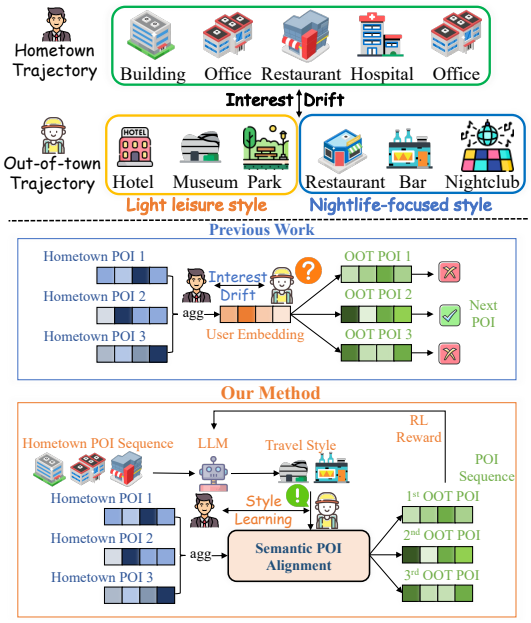

Figure 1: Comparison with Previous Work: Out-of-town POI Sequence Prediction is inherently challenging as the model must capture complex user preferences while mitigating the issue of *interest drift*. Prior approaches typically represent each POI and user independently for next POI recommendation. Our method introduces reinforcement learning–based LLMs, which allows the model to capture the travel styles between different regions more precisely for POI sequence prediction.

aging their semantic reasoning to infer preferences or generate explanations (Feng et al., 2025; Cheng et al., 2025; Liu et al., 2024a). Yet, most rely on supervised fine-tuning (SFT) with fixed style templates or direct sequence generation, which often overfits to frequent patterns, collapses output diversity, and lacks explicit alignment between inferred semantics and downstream structural accuracy (Li et al., 2024, 2025). Crucially, the alignment between high-level textual style inference and fine-grained positional sequence prediction remains an open problem in LLM-based structured generation.

To bridge these gaps, we propose **SemPOI-RL**, a reinforcement learning framework that aligns LLM semantic inference with structured sequence generation for interpretable cross-city itinerary planning. Unlike prior approaches that treat LLMs as direct sequence generators or rely on opaque latent mappings, our method explicitly decouples semantic style extraction from structural prediction, using natural language as an interpretable intermediary. Specifically, we first adapt an LLM via supervised fine-tuning to infer destination-oriented travel styles from users' hometown textual trajectories, effectively modeling cross-domain preference shifts through semantic reasoning (Xin et al., 2022; Feng et al., 2024). We then introduce a *Semantic POI Alignment Module* (SPAM) that injects these LLM-derived textual styles into a masked autoencoder framework. By learning style-weighted attention distributions across sequence positions, the model dynamically associates different linguistic styles with specific temporal phases, enabling fine-grained semantic control over sequence generation (Li et al., 2022). Finally, to align the LLM's textual output with downstream structural accuracy, we employ Group Relative Policy Optimization (GRPO) guided by a composite reward function (Shao et al., 2024; Yu et al., 2025). This RL stage encourages the LLM to explore diverse, semantically rich style descriptions that are explicitly optimized for positional hit rate, recall, category consistency, and diversity, thereby bridging the gap between plausible text generation and structurally accurate sequence prediction (Li et al., 2025; Schulman et al., 2017).

We evaluate SemPOI-RL on two large-scale real-world datasets, demonstrating that our framework significantly outperforms both traditional sequential modeling baselines and direct LLM generation methods. Beyond quantitative gains, our approach offers substantial interpretability benefits: the learned attention maps reveal how different LLM-inferred styles dominate specific sequence positions, providing human-readable attributions for generated trajectories. This work makes three primary contributions:

- We propose **SemPOI-RL**, a novel reinforcement learning framework with semantic alignment for cross-city POI recommendation. By explicitly modeling *interest drift*, our framework leverages LLMs to infer destination-oriented travel styles from hometown trajectories, enabling interpretable and style-aware sequence generation without relying on latent ID mappings.

- We design SPAM that decomposes global travel styles into temporally varying prototypical styles, and integrate it with a masked autoencoder to dynamically weight style influences across sequence positions. Furthermore, we introduce a multi-objective reinforcement learning strategy guided by hit rate, recall, category consistency, and diversity rewards, which refines LLM-generated styles to bridge

the gap between semantic plausibility and structural prediction accuracy.

- We conduct comprehensive experiments on two large-scale real-world datasets, demonstrating that SemPOI-RL consistently outperforms strong baselines in both accuracy and trajectory coherence. Detailed ablation studies and interpretability analyses further validate the necessity of each component, revealing how explicit modeling of interest drift and style-aware alignment leads to more faithful and human-understandable out-of-town recommendations.

2 Related Work

2.1 POI Recommendation

Early studies primarily framed POI sequence prediction as a Markovian process based on transition probabilities(Cheng et al., 2013). To address sequential dependency, models such as ST-LSTM (Wang et al., 2017; Zhao et al., 2019; Zhang et al., 2024) and transformer-based models (Kang and McAuley, 2018; Shu et al., 2024) incorporated spatio-temporal gating mechanisms to account for time intervals and geographic distances between consecutive check-ins. Meanwhile, GNNs have also gained prominence in POI recommendation due to their strong capability in modeling graph structures. For instance, GETNext(Yang et al., 2022) and STHGCN(Yan et al., 2023) leverage GCN over trajectory flow graphs to capture POI-to-POI transition probabilities. To tackle challenges of data sparsity and interest drift, researchers have proposed approaches based on transfer learning. (Xin et al., 2022) first studys the pre-travel out-of-town POI recommendation problem. KDDC considers the distributions of conformity and interest(Liu et al., 2024b). SPOT-Trip learns the dual static-dynamic user preferences to enrich the semantic modeling(Liu et al., 2025). Traditional transfer learning approaches typically rely on learning mapping functions within a low-dimensional latent space to achieve preference transfer. In contrast, LLMs abstract a user's core interests into high-level, interpretable concepts, effectively bypassing reliance on specific POI IDs.

2.2 LLMs for Recommendation

In recent years, LLMs have emerged as a promising paradigm in recommender systems, owing to their strong natural language understanding, zero-shot and few-shot adaptability, and capacity to generate interpretable explanations(Xie et al., 2024; Singh et al., 2024). Specifically, LLMs not only generate recommendations based on users' historical behaviors but also provide intuitive, human-readable rationales(Tang et al., 2024). Prior work includes USER-LLM(Ning et al., 2025), which treats the LLM as a user preference encoder; and LLMob(Wang et al., 2024), which infers users' occupations and travel motivations via LLMs to generate user trajectories. (Wang et al., 2025) maps POI identifiers into semantically enriched representations. Moreover, LLMs have achieved strong performance on basic temporal prediction tasks(Jin et al., 2024), inspiring extensions to next-point prediction via POI feature extraction using LLMs(Feng et al., 2025; Wang et al., 2023; Liu et al., 2024a; Feng et al., 2024). In addition, several recent studies have explored training LLMs for POI recommendation through either supervised fine-tuning(Li et al., 2024) or reinforcement learning(Li et al., 2025), enabling them to directly generate recommendation POIs. However, these methods exhibit two major limitations. First, they struggle to generalize to cross-city POI prediction. Second, when the candidate set of POIs becomes large, the limited context window of LLMs constrains both prediction accuracy and computational efficiency.

3 Preliminary

We follow prior studies(Xin et al., 2022; Liu et al., 2025) to formally define the out-of-town trip recommendation problem.

Definition 1 (POI). A POI is a spatial item related to a geographical location. We use v to represent a POI, r_v to denote its located region, l to denote its location as longitude and latitude coordinates and g to denote its category.

Definition 2 (Check-in). If a user checks in a POI, it indicates that s/he has a direct interaction with the POI in the real world. Formally, we denote a user u's check-in activity on POI v at time t with a three-tuple $c = (u, t, v)$.

Definition 3 (User's Hometown). Given a user u, we denote a region r_h as the user's hometown where the user lives or works for a period of time. The algorithm to discover a user's hometown remains the same as (Xin et al., 2022).

Definition 4 (Out-of-town Travel Behavior). Given a user u and her/his hometown r_h in which

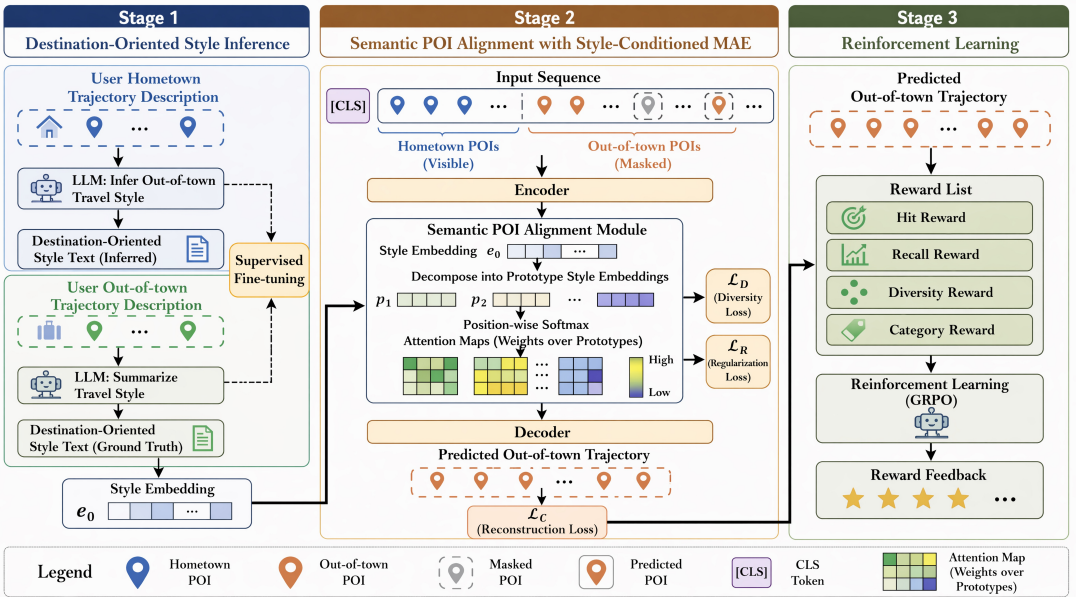

Figure 2: The overall framework of SemPOI-RL.

s/he leaves check-in records $\vec{c}_h$, when s/he travels to the out-of-town region r_o and also leaves check-in records $\vec{c}_o$ in r_o, we call such behavior an out-of-town travel behavior, denoted as a five-tuple $\xi = (u, \vec{c}_h, \vec{c}_o, r_h, r_o)$.

Problem Statement: (Out-of-town Trip Recommendation). Given a set of users $\mathcal{U} = \{u_i\}_{i=1}^{|\mathcal{U}|}$, POIs $\mathcal{V} = \{v_i\}_{i=1}^{|\mathcal{V}|}$, regions $\mathcal{R} = \{r_i\}_{i=1}^{|\mathcal{R}|}$, and out-of-town trip records $\mathcal{O} = \{\xi_i\}_{i=1}^{|\mathcal{O}|}$, our objective is to learn a recommender function $\mathcal{F}$ based on historical records $\mathcal{O}$. For a new user $u^* \notin \mathcal{U}$ at r_h^* with hometown check-ins $\vec{c}_h$ and out-of-town origin-destination trip query $Q_o^* = \{v_s^o, v_e^o, N\}$ in region r_o^*, where v_s^o, v_e^o, and N denote the start, end points and the number of POIs, the learned $\mathcal{F}$ generates a sequence of POIs $\tau = \{v_1^o, v_2^o, \ldots, v_N^o\}$, where $v_1^o = v_s^o$ and $v_N^o = v_e^o$ for u^*. The recommended POIs in τ are in r_o^*.

4 Methodology

We propose **SemPOI-RL**, a three-stage framework that aligns LLM-based semantic inference with structured POI sequence generation for cross-city recommendation. As discussed in the Introduction, the main challenge of out-of-town recommendation is not only to infer what a user may prefer in a new city, but also to map such high-level semantics into an ordered destination trajectory under dynamic interest drift. To address this semantic-to-structure gap, we decouple the problem into: (i) inferring destination-oriented travel styles from hometown trajectories, (ii) grounding these styles into position-aware trajectory generation, and (iii) refining style generation with recommendation-oriented rewards. Figure 2 illustrates the overall pipeline.

4.1 Stage 1: Destination-Oriented Style Inference

A key difficulty in cross-city recommendation is that hometown behaviors and destination behaviors are not directly aligned at the POI-ID level. Rather than transferring preferences through latent IDs only, we use an LLM to infer an interpretable *destination-oriented travel style* from the user's hometown trajectory.

Given a user u, hometown trajectory $\vec{c}_h$ in region r_h, and destination trajectory $\vec{c}_o$ in region r_o, we construct two prompts for an initial LLM $\mathrm{LLM}_i(\cdot)$:

1. **Inference from hometown trajectory**

$$\mathbf{y}^{(h)} = \mathrm{LLM}_i(\mathrm{Prompt}_h(u, \vec{c}_h, r_h)), \quad (1)$$

where $\mathbf{y}^{(h)}$ is the inferred destination style text.

2. **Style summarization from destination trajectory**

$$\mathbf{y}^{(o)} = \mathrm{LLM}_i(\mathrm{Prompt}_o(u, \vec{c}_o, r_o)), \quad (2)$$

where $\mathbf{y}^{(o)}$ is the style summary directly obtained from the observed destination trajectory.

We treat $\mathbf{y}^{(o)}$ as a supervision signal and fine-tune the LLM to predict it from the hometown prompt. Let $p_\theta(y_t \mid y_{<t}, \mathrm{Prompt})$ denote the token probability. The supervised objective is:

$$\mathcal{L}_{\mathrm{CE}}(\theta) = -\sum_{t=1}^{T_o} \log p_\theta\Big(y_t^{(o)} \mid y_{<t}^{(o)}, \mathrm{Prompt}_h(u, \vec{c}_h, r_h)\Big). \tag{3}$$

After fine-tuning, the LLM generates a destination-oriented style summary:

$$\begin{aligned} s_o &= \mathrm{LLM}_t(\mathrm{Prompt}_h(u, \vec{c}_h, r_h)), \\ e_o &= \mathrm{text_embedding}(s_o), \end{aligned} \tag{4}$$

where s_o represents the user's destination travel style, e_o is the embedding of the generated style text. A global style vector captures coarse semantic preference, but it cannot model the temporal variation emphasized in the Introduction. In practice, users often exhibit different styles at different phases of a trip (e.g., daytime leisure versus nighttime entertainment). Therefore, we further decompose the global style into multiple prototypical semantic components.

4.2 Stage 2: Semantic POI Alignment with Style-Conditioned MAE

To bridge textual style inference and structured sequence prediction, we introduce the **Semantic POI Alignment Module (SPAM)** and integrate it with a masked autoencoder (MAE).

4.2.1 Semantic POI Alignment Module

Given the style embedding e_o, SPAM decomposes it into M prototypical style embeddings:

$$F_p^{(i)} = e_o \circ \mathrm{sigmoid}\Big(W_{\mathrm{c2}}^{(i)} \tanh\Big(W_{\mathrm{c1}}^{(i)} e_o\Big)\Big), \tag{5}$$

where $i = 1, \ldots, M$,, $\circ$ denotes the Hadamard product, and $W_{\mathrm{c1}}^{(i)}$, $W_{\mathrm{c2}}^{(i)}$ are learnable parameters.

This decomposition provides a set of latent style prototypes that can be selectively activated at different trajectory positions, instead of forcing the entire trip to be explained by a single semantic vector.

To avoid redundant prototypes, we impose a diversity loss:

$$\mathcal{L}_D = \frac{2}{M(M-1)} \sum_{1 \le i < j \le M} \Big(\cos(\hat{F}_p^{(i)}, \hat{F}_p^{(j)})\Big)^2, \tag{6}$$

where $\hat{F}_p^{(i)} = F_p^{(i)} / \|F_p^{(i)}\|_2$. This loss encourages the prototypes to capture distinct semantic factors.

4.2.2 Style-Conditioned MAE

We use the semantic prototypes to guide a masked autoencoder for destination trajectory reconstruction. For each check-in, location l_i and timestamp t_i are embedded as $x_i = \phi(l_i, t_i) + \mathrm{pos}(i)$, where $\phi(\cdot)$ is the input embedding function and $\mathrm{pos}(\cdot)$ denotes positional encoding.

During training, we keep all hometown points visible and mask only a proportion ρ of destination points, yielding the partially observed sequence $\tilde{x}$. This masking strategy is consistent with inference: the hometown trajectory is given, while the destination trajectory must be predicted.

The encoder produces contextual representations $h = E(\tilde{x})$, where h_i is the hidden state at position i. To align each position with the most relevant semantic prototype, we compute a prototype distribution

$$\alpha_i = \mathrm{SoftMax}\big(\{\langle h_i, F_p^{(k)} \rangle\}_{k=1}^{M}\big),$$

where $\alpha_i = [\alpha_{i,1}, \ldots, \alpha_{i,M}]$. This position-wise style assignment is the key step that grounds high-level style semantics into local trajectory generation.

We then form a style-aware representation $\tilde{h}_i = \sum_{k=1}^{M} \alpha_{i,k} W_v F_p^{(k)}$, and feed it to the decoder together with the global `[CLS]` representation, i.e., $z = D(h_{\mathrm{cls}}, \tilde{h})$, where z denotes the logits over destination POIs.

The model is trained to reconstruct the masked destination positions with cross-entropy:

$$\mathcal{L}_C = -\frac{1}{N} \sum_{n=1}^{N} \log \frac{\exp(z_{n,v_n^o})}{\sum_{i \in r_o} \exp(z_{n,i})}, \tag{7}$$

where v_n^o is the ground-truth POI at the n-th masked position. To avoid collapsing all positions to a single prototype, we further regularize the style assignments with entropy:

$$\mathcal{L}_R = -\frac{1}{|\mathcal{I}|} \sum_{i \in \mathcal{I}} \sum_{k=1}^{M} \alpha_{i,k} \log \alpha_{i,k}, \tag{8}$$

where $\mathcal{I}$ indexes the sequence positions.

The overall training objective of this section is:

$$\mathcal{L} = \mathcal{L}_C + \lambda_D \mathcal{L}_D + \lambda_R \mathcal{L}_R. \tag{9}$$

4.3 Stage 3: Reinforcement Learning for LLM Style Generation

The supervised fine-tuning stage equips the LLM with a basic ability to infer destination-oriented

travel styles from hometown trajectories, but it is still insufficient for our final objective. A key reason is that Stage I optimizes the model only at the *text* level: the LLM is trained to mimic destination-style descriptions, rather than being directly supervised by the *ground-truth destination trajectory*. As a result, the generated style may be linguistically plausible yet suboptimal for downstream POI sequence prediction. We therefore introduce a reinforcement learning stage to further align style generation with the actual recommendation target. Furthermore, it relaxes the rigidity of supervised learning by allowing the LLM to explore multiple semantically reasonable style descriptions, and then retain those that lead to better structural prediction outcomes. In this way, reinforcement learning connects high-level semantic generation with downstream recommendation quality. The total reward R is defined as:

$$R = \lambda_{HR} \cdot R_{HR} + \lambda_{RR} \cdot R_{RR} + \lambda_{CM} \cdot R_{CM} + \lambda_{DIV} \cdot R_{DIV}, \tag{10}$$

where R_{HR}, R_{RR}, R_{CM}, and R_{DIV} denote the hit rate reward, recall reward, category-matching reward, and diversity reward, respectively. R_{DIV} is the proportion of non-repetitive predictions of POIs. The detailed computation of other rewards is provided in Section 5.1. λ_{HR}, λ_{RR}, λ_{CM}, λ_{DIV} are weighting hyperparameters.

For policy optimization, we adopt GRPO with a DAPO-style objective for stable training. This reward design directly reflects the desiderata of cross-city trajectory recommendation. Hit rate and recall encourage the generated style to support accurate POI recovery; category consistency encourages semantic faithfulness at the trajectory level; and diversity discourages degenerate or repetitive outputs. More importantly, all these rewards are computed from the predicted destination trajectory against the ground truth, thereby providing the missing supervision signal that is absent in Stage 1.

4.4 Joint Style-Conditioned POI Sequence Prediction

At test time, given a new user's hometown trajectory, we first use the RL-tuned LLM to infer a destination-oriented style. We then compute the prototype style embeddings with SPAM and feed them into the trained style-conditioned MAE to generate the destination trajectory.

5 Experiment

5.1 Experiment Setup

Datasets Our experiments are conducted on two widely used real-world datasets, **Foursquare** and **Yelp**. Detailed dataset statistics and preprocessing procedures are provided in Appendix A.

Metrics We evaluate all methods using six metrics: **Hit Rate (HR)**, **Recall Rate (RR)**, **Edit Distance (ED)**, **Dynamic Time Warping Distance (DTW)**, **Category Match (CM)**, and **Region Match (RM)**. Their formal definitions are given in Appendix B.

Settings Our task is more challenging than conventional next-POI prediction, because we aim to generate the *entire* out-of-town POI sequence rather than predict only the next location. Following recent out-of-town trip recommendation settings such as SPOT-Trip, we assume that the start and end POIs of the destination trajectory are given, and the model predicts the intermediate POIs between them. This setting is motivated by a realistic travel-planning scenario: travelers usually know where they start and where they finally want to arrive, while the intermediate stops remain flexible. In addition, removing the start/end constraints makes the search space extremely large and causes prediction quality to deteriorate substantially, making stable training difficult. More implementation details are provided in Appendix C.

Baselines We compare SemPOI-RL with 10 baselines, including 7 traditional trip recommendation methods: LSTM (Hochreiter and Schmidhuber, 1997), GETNext (Yang et al., 2022), STHGCN (Yan et al., 2023), MatTrip (Zhang et al., 2024), AR-Trip (Shu et al., 2024), KDDC (Liu et al., 2024b), and SPOT-Trip (Liu et al., 2025); and 3 LLM-related methods: LLMMove (Feng et al., 2024), LLM4POI (Li et al., 2024), and RefinePOI (Li et al., 2025). More details are provided in Appendix D.

5.2 Overall Result

In this section, we compare **SemPOI-RL** with all baselines on the two benchmark datasets. The results in Table 1 show that **SemPOI-RL** consistently achieves the best overall performance across both datasets. In particular, our method improves Hit Rate by about 15% on Foursquare and 10% on Yelp over the strongest non-LLM baselines. Com-

Table 1: The overall comparison between SemPOI-RL and baselines

Model	Foursquare						Yelp					
	HR(↑)	RR(↑)	ED(↓)	DTW(↓)	CM(↑)	RM(↑)	HR(↑)	RR(↑)	ED(↓)	DTW(↓)	CM(↑)	RM(↑)
LSTM	0.0156	0.0187	0.9844	0.4920	0.0336	0.3504	0.0031	0.0032	0.9969	0.4985	0.0059	0.1255
GETNext	0.0069	0.0359	0.9925	0.4966	0.0322	0.1732	0.0052	0.0108	0.9934	0.4956	0.0516	0.1379
STHGCN	0.0084	0.0371	0.9910	0.4944	0.0367	0.2002	0.0066	0.0126	0.9926	0.4931	0.0587	0.1502
MatTrip	0.0108	0.0358	0.9884	0.4942	0.0294	0.1881	0.0095	0.0301	0.9905	0.4953	0.0402	0.1760
KDDC	0.0130	0.0365	0.9906	0.4931	0.0152	0.1820	0.0083	0.0329	0.9880	0.5186	0.0610	0.1045
AR-Trip	0.0083	0.0308	0.9917	0.4958	0.0282	**1.0000**	0.0238	0.0307	0.9762	0.4880	0.0618	0.8607
SPOT-Trip	0.0190	0.0410	0.9806	0.4903	0.0360	**1.0000**	0.0248	0.0389	0.9726	0.4863	0.0664	0.9443
LLMMove	0.0028	0.0355	0.9929	0.5344	0.0388	0.9314	0.0132	0.0256	0.9862	0.5387	0.0617	0.8843
LLM4POI	0.0057	0.0083	0.9956	0.4980	0.0223	0.2036	0.0148	0.0413	0.9836	0.4918	0.0625	0.3022
Refine-POI	0.0060	0.0112	0.9940	0.4970	0.0349	0.3159	0.0065	0.0178	0.9924	0.4962	0.0484	0.1517
SemPOI-RL	**0.0219**	**0.0437**	**0.9778**	**0.4876**	**0.0572**	**1.0000**	**0.0273**	**0.0419**	**0.9701**	**0.4843**	**0.0703**	**0.9449**

pared with traditional POI recommendation methods, the gains mainly come from explicitly modeling *interest drift* between hometown and destination. Instead of relying only on latent user embeddings, SemPOI-RL uses the semantic reasoning ability of LLMs to infer destination-oriented travel styles from hometown trajectories, which provides stronger guidance for sequence generation. Compared with existing LLM-based baselines, our method also performs better under all settings, showing that simple prompting or direct fine-tuning is insufficient for this task without semantic alignment and recommendation-oriented optimization.

Among the LLM-based baselines, LLMMove shows relatively competitive Recall Rate because we manually sample 100 negative POIs from a much larger candidate space for the untrained LLM to rank. Nevertheless, all three LLM-based baselines perform poorly on Foursquare, likely because the proportion of out-of-town records is smaller, making destination-style extraction more difficult. In addition, these methods generally underperform on Region Match, suggesting that they struggle to preserve regional coherence in generated trajectories. By contrast, **SemPOI-RL** achieves stronger performance on **CM**, showing that it better captures category-level semantics. It also obtains lower **ED** and **DTW**, indicating that the generated trajectories are structurally closer to the ground-truth routes and require fewer edits to match real user behavior. These improvements are practically important for travel planning, where both semantic suitability and visit order matter.

5.3 Ablation Study

To validate the effectiveness of each component, we design the following ablation settings:

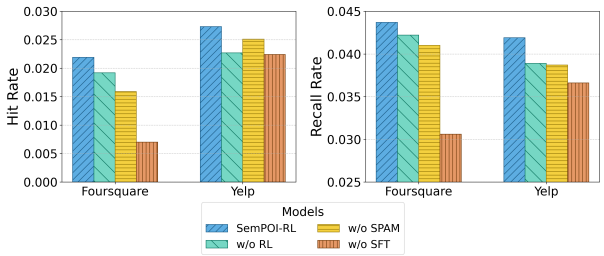

Figure 3: Ablation study of the proposed SemPOI-RL.

- **w/o SPAM**: remove the Semantic POI Alignment Module and directly fuse MAE representations with text embeddings for prediction;
- **w/o SFT**: skip the supervised fine-tuning stage, requiring the LLM to infer destination travel styles directly from hometown trajectories without prior adaptation;
- **w/o RL**: remove the reinforcement learning stage and use the style representations learned before RL for downstream POI prediction.

Figure 3 and the quantitative results below show that all ablated variants underperform the full model across most metrics, confirming the necessity of each module.

First, removing **SFT** causes the largest performance drop, especially on Foursquare. This indicates that supervised fine-tuning is critical for teaching the LLM how to map hometown behavior patterns to destination-oriented travel styles. Second, removing **SPAM** degrades performance on both trajectory-level and semantic metrics. SPAM enables the model to decompose overall user preference into temporally varying prototype styles, which helps the generator capture different trip phases more accurately. Third, removing **RL** also leads to consistent degradation. This shows that

recommendation-oriented reward optimization further refines the LLM-inferred styles and improves their usefulness for downstream sequence generation. We add more hyper-parameter analysis in Appendix E.

5.4 Case Study and Interpretability Analysis

To provide an intuitive understanding of the interpretability of **SemPOI-RL**, we present a case study in Table 2. Based on the hometown trajectory, the LLM infers that the user prefers dynamic urban experiences involving recreation, exploration, and interactive activities.

As shown in Table 2, the predicted destination categories reveal a clear phase transition. The first part of the generated sequence mainly contains venues correspond to leisure exploration and entertainment-oriented activities. In contrast, the latter part shifts to rest and trip closure. This pattern is broadly consistent with the target trajectory, whose early positions are dominated by social and nightlife categories and whose later positions mainly contain accommodation-related POIs. The prototype assignments also support this interpretation. In this example, the early positions, while prototype style 1 appears in the latter positions, suggesting that the model has learned to use different latent styles for different trip stages. Although these prototypes are latent embeddings rather than directly human-readable labels, their semantic meanings can be approximately characterized through the category distributions of the POIs they activate. Across the full dataset, we find that the same prototype tends to correspond to the similar trip phases, for example, the main categories of attractions with prototype styles 3 are entertainment and city exploration.

Table 2: A case study for user 2842 on the Foursquare dataset. Purple highlights key semantic cues in the inferred travel style. Blue and red indicate two dominant trip phases.

Hometown Trajectory	[Bike Shop, Track, University, Boutique, Tech Startup]
Travel Style Inference	User's travel trajectory in New York indicates a pattern of engaging with urban and recreational spaces, suggesting a travel style that leans toward cultural and recreational exploration. The frequent visits to a bike shop and a track suggest an interest in physical activity and outdoor movement. The visit to a university points to an intellectual or academic curiosity, while the boutique visit implies an appreciation for unique shopping experiences. The engagement with a tech startup further highlights an interest in innovation and modern urban environments. These behaviors collectively suggest a preference for dynamic, interactive, and culturally rich experiences. The user appears to seek environments that offer both activity and discovery, favoring destinations that provide opportunities for exploration, learning, and engagement with local culture and innovation. This profile aligns with a travel style characterized by curiosity, adaptability, and a desire for meaningful, experiential interactions.
Predicted IDs	[159, 767, 611, 611, 186, 188, 108, 237, 6]
Predicted POI Category	[BBQ Joint, Lounge, Music Venue, Music Venue, Rock Club, Hotel Bar, Baseball Stadium, Hotel, Airport]
Prototype Style	[3, 3, 3, 3, 3, 3, 1, 1, 1]
Target POI IDs	[159, 188, 767, 186, 239, 188, 67, 774, 1112]
Target POI Category	[BBQ Joint, Rock Club, Lounge, Rock Club, Hotel Bar, Rock Club, Hotel, Hotel, Hotel]

6 Conclusion

We presented **SemPOI-RL**, a semantic-aligned framework for cross-city POI sequence recommendation. Our method addresses a core challenge of out-of-town recommendation: bridging the gap between high-level user preference semantics and structured destination trajectory generation. To this end, we first use an LLM to infer destination travel styles from users' hometown trajectories, providing an interpretable semantic abstraction of cross-region interest drift. We then introduce a Semantic POI Alignment Module to decompose global style signals into temporally varying prototypes and inject them into a style-conditioned masked autoencoder for position-aware POI sequence prediction. Finally, we further refine LLM style generation with multi-objective reinforcement learning for downstream recommendation accuracy. Experiments on two real-world datasets show that SemPOI-RL consistently outperforms both traditional sequential recommendation models and recent LLM-based baselines. In addition to improving trajectory accuracy and structural coherence, our framework offers interpretable style attribution across different phases of a trip, providing a more transparent view of how semantic reasoning can guide sequential recommendation. We hope this work can encourage future research on aligning LLM-based semantic inference with structured decision-making tasks beyond POI recommendation.

Limitations

Our work has several limitations. First, although SemPOI-RL improves interpretability by using natural-language style summaries and prototype-level alignment, the learned prototype styles are still latent vectors rather than fully human-readable concepts. As a result, the semantic meaning of each prototype often needs to be inferred indirectly from the POI categories it activates. Second, the framework depends on the quality of LLM-generated style descriptions. While the reinforcement learning stage partially alleviates this issue through trajectory-level rewards, it does not fully eliminate the risk of unstable or inconsistent style generation. Third, our experiments follow a constrained out-of-town recommendation setting in which the start POI, end POI, and trajectory length are given in advance. This setting is standard in prior work and allows fair comparison, but it is less general than open-ended itinerary generation, where users may not specify route boundaries or trip length beforehand. Fourth, our multi-stage design improves performance, but it also increases computational cost and implementation overhead compared with simpler end-to-end recommenders. Our experiments are conducted on an open-source backbone to enable supervised fine-tuning and reinforcement learning instead of GPT-style API models, mainly because their parameter access and training interfaces are restricted. Therefore, it remains unclear whether stronger closed-source models could reduce or amplify the benefit of semantic alignment in OOT recommendation.

Ethical Considerations

User privacy and location sensitivity. POI trajectories are highly sensitive because they may reveal users' daily routines, lifestyle patterns, and potentially private attributes such as workplace, residence area, health-related visits, or religious activities. To mitigate this risk, our study uses only anonymized datasets and does not attempt to recover user identities or infer protected personal attributes. Nevertheless, we acknowledge that trajectory modeling remains inherently privacy-sensitive, and any real-world deployment should incorporate stronger safeguards such as data minimization, access control, aggregation, and privacy-preserving learning mechanisms.

Bias and representational harms. The proposed framework relies on both historical check-in data and LLM-generated style summaries. As a result, it may inherit biases from two sources. First, the underlying datasets may over-represent active users, popular urban regions, or frequently visited POI categories, while under-representing less active users and minority travel patterns. Second, the LLM may produce stereotypical or overly generic style descriptions based on skewed data distributions. These biases could lead to less accurate or less fair recommendations for certain user groups, regions, or travel behaviors.

References

Chen Cheng, Haiqin Yang, Michael R. Lyu, and Irwin King. 2013. Where you like to go next: successive point-of-interest recommendation. In *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence*, IJCAI '13, page 2605–2611. AAAI Press.

Jiawei Cheng, Jingyuan Wang, Yichuan Zhang, Jiahao Ji, Yuanshao Zhu, Zhibo Zhang, and Xiangyu Zhao. 2025. Poi-enhancer: An llm-based semantic enhancement framework for POI representation learning. In *AAAI-25, Sponsored by the Association for the Advancement of Artificial Intelligence, February 25 - March 4, 2025, Philadelphia, PA, USA*, pages 11509–11517. AAAI Press.

Jie Feng, Yuwei Du, Jie Zhao, and Yong Li. 2025. Agentmove: A large language model based agentic framework for zero-shot next location prediction. In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL 2025 - Volume 1: Long Papers, Albuquerque, New Mexico, USA, April 29 - May 4, 2025*, pages 1322–1338. Association for Computational Linguistics.

Shanshan Feng, Haoming Lyu, Caishun Chen, and Yew-Soon Ong. 2024. Where to move next: Zero-shot generalization of llms for next POI recommendation. *CoRR*, abs/2404.01855.

Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural Computation*, 9(8):1735–1780.

Ming Jin, Shiyu Wang, Lintao Ma, Zhixuan Chu, James Y. Zhang, Xiaoming Shi, Pin-Yu Chen, Yuxuan Liang, Yuan-Fang Li, Shirui Pan, and Qingsong Wen. 2024. Time-llm: Time series forecasting by reprogramming large language models. In *The Twelfth International Conference on Learning Representations, ICLR 2024, Vienna, Austria, May 7-11, 2024*. OpenReview.net.

Wang-Cheng Kang and Julian J. McAuley. 2018. Self-attentive sequential recommendation. In *IEEE International Conference on Data Mining, ICDM 2018, Singapore, November 17-20, 2018*, pages 197–206. IEEE Computer Society.

Gang Li, Heliang Zheng, Daqing Liu, Chaoyue Wang, Bing Su, and Changwen Zheng. 2022. Semmae: Semantic-guided masking for learning masked autoencoders. In *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022*.

Peibo Li, Shuang Ao, Hao Xue, Yang Song, Maarten de Rijke, Johan Barthélemy, Tomasz Bednarz, and Flora D. Salim. 2025. Refine-poi: Reinforcement fine-tuned large language models for next point-of-interest recommendation. *CoRR*, abs/2506.21599.

Peibo Li, Maarten de Rijke, Hao Xue, Shuang Ao, Yang Song, and Flora D. Salim. 2024. Large language models for next point-of-interest recommendation. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2024, Washington DC, USA, July 14-18, 2024*, pages 1463–1472. ACM.

Shuai Liu, Ning Cao, Yile Chen, Yue Jiang, and Gao Cong. 2024a. nextlocllm: next location prediction using llms. *CoRR*, abs/2410.09129.

Xin Liu, Yongjian Yang, Yuanbo Xu, Funing Yang, Qiuyang Huang, and Hong Wang. 2022. Real-time POI recommendation via modeling long- and short-term user preferences. *Neurocomputing*, 467:454–464.

Yinghui Liu, Hao Miao, Guojiang Shen, Yan Zhao, Xiangjie Kong, and Ivan Lee. 2025. Spot-trip: Dual-preference driven out-of-town trip recommendation. *CoRR*, abs/2506.01705.

Yinghui Liu, Guojiang Shen, Chengyong Cui, Zhenzhen Zhao, Xiao Han, Jiaxin Du, Xiangyu Zhao, and Xiangjie Kong. 2024b. KDDC: knowledge-driven disentangled causal metric learning for pre-travel out-of-town recommendation. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence, IJCAI 2024, Jeju, South Korea, August 3-9, 2024*, pages 2207–2215. ijcai.org.

Lin Ning, Luyang Liu, Jiaxing Wu, Neo Wu, Devora Berlowitz, Sushant Prakash, Bradley Green, Shawn O'Banion, and Jun Xie. 2025. User-llm: Efficient LLM contextualization with user embeddings. In *Companion Proceedings of the ACM on Web Conference 2025, WWW 2025, Sydney, NSW, Australia, 28 April 2025 - 2 May 2025*, pages 1219–1223. ACM.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. 2017. Proximal policy optimization algorithms. *CoRR*, abs/1707.06347.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Mingchuan Zhang, Y. K. Li, Y. Wu, and Daya Guo. 2024. Deepseekmath: Pushing the limits of mathematical reasoning in open language models. *CoRR*, abs/2402.03300.

Wenzheng Shu, Kangqi Xu, Wenxin Tai, Ting Zhong, Yong Wang, and Fan Zhou. 2024. Analyzing and mitigating repetitions in trip recommendation. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2024, Washington DC, USA, July 14-18, 2024*, pages 2276–2280. ACM.

Harmanpreet Singh, Nikhil Verma, Yixiao Wang, Manasa Bharadwaj, Homa Fashandi, Kevin Ferreira, and Chul Lee. 2024. Personal large language model agents: A case study on tailored travel planning. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: EMNLP 2024 - Industry Track, Miami, Florida, USA, November 12-16, 2024*, pages 486–514. Association for Computational Linguistics.

Yihong Tang, Zhaokai Wang, Ao Qu, Yihao Yan, Zhaofeng Wu, Dingyi Zhuang, Jushi Kai, Kebing Hou, Xiaotong Guo, Jinhua Zhao, Zhan Zhao, and Wei Ma. 2024. Itinera: Integrating spatial optimization with large language models for open-domain urban itinerary planning. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: EMNLP 2024 - Industry Track, Miami, Florida, USA, November 12-16, 2024*, pages 1413–1432. Association for Computational Linguistics.

Dongsheng Wang, Yuxi Huang, Shen Gao, Yifan Wang, Chengrui Huang, and Shuo Shang. 2025. Generative next POI recommendation with semantic ID. *CoRR*, abs/2506.01375.

Jiawei Wang, Renhe Jiang, Chuang Yang, Zengqing Wu, Makoto Onizuka, Ryosuke Shibasaki, Noboru Koshizuka, and Chuan Xiao. 2024. Large language models as urban residents: An LLM agent framework for personal mobility generation. In *Advances in Neural Information Processing Systems 38: Annual Conference on Neural Information Processing Systems 2024, NeurIPS 2024, Vancouver, BC, Canada, December 10 - 15, 2024*.

Xinglei Wang, Meng Fang, Zichao Zeng, and Tao Cheng. 2023. Where would I go next? large language models as human mobility predictors. *CoRR*, abs/2308.15197.

Yunbo Wang, Mingsheng Long, Jianmin Wang, Zhifeng Gao, and Philip S. Yu. 2017. Predrnn: Recurrent neural networks for predictive learning using spatiotemporal lstms. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pages 879–888.

Jian Xie, Kai Zhang, Jiangjie Chen, Tinghui Zhu, Renze Lou, Yuandong Tian, Yanghua Xiao, and Yu Su. 2024.

Travelplanner: A benchmark for real-world planning with language agents. In *Forty-first International Conference on Machine Learning, ICML 2024, Vienna, Austria, July 21-27, 2024*. OpenReview.net.

Haoran Xin, Xinjiang Lu, Nengjun Zhu, Tong Xu, Dejing Dou, and Hui Xiong. 2022. CAPTOR: A crowd-aware pre-travel recommender system for out-of-town users. In *SIGIR '22: The 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, Madrid, Spain, July 11 - 15, 2022*, pages 1174–1184. ACM.

Xiaodong Yan, Tengwei Song, Yifeng Jiao, Jianshan He, Jiaotuan Wang, Ruopeng Li, and Wei Chu. 2023. Spatio-temporal hypergraph learning for next POI recommendation. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2023, Taipei, Taiwan, July 23-27, 2023*, pages 403–412. ACM.

An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, Chujie Zheng, Dayiheng Liu, Fan Zhou, Fei Huang, Feng Hu, Hao Ge, Haoran Wei, Huan Lin, Jialong Tang, and 40 others. 2025. Qwen3 technical report. *CoRR*, abs/2505.09388.

Song Yang, Jiamou Liu, and Kaiqi Zhao. 2022. Getnext: Trajectory flow map enhanced transformer for next POI recommendation. In *SIGIR '22: The 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, Madrid, Spain, July 11 - 15, 2022*, pages 1144–1153. ACM.

Hongzhi Yin, Xiaofang Zhou, Bin Cui, Hao Wang, Kai Zheng, and Nguyen Quoc Viet Hung. 2016. Adapting to user interest drift for POI recommendation. *IEEE Trans. Knowl. Data Eng.*, 28(10):2566–2581.

Qiying Yu, Zheng Zhang, Ruofei Zhu, Yufeng Yuan, Xiaochen Zuo, Yu Yue, Tiantian Fan, Gaohong Liu, Lingjun Liu, Xin Liu, Haibin Lin, Zhiqi Lin, Bole Ma, Guangming Sheng, Yuxuan Tong, Chi Zhang, Mofan Zhang, Wang Zhang, Hang Zhu, and 16 others. 2025. DAPO: an open-source LLM reinforcement learning system at scale. *CoRR*, abs/2503.14476.

Jiale Zhang, Mingqian Ma, Xiaofeng Gao, and Guihai Chen. 2024. Encoder-decoder based route generation model for flexible travel recommendation. *IEEE Trans. Serv. Comput.*, 17(3):905–920.

Qianru Zhang, Peng Yang, Junliang Yu, Haixin Wang, Xingwei He, Siu-Ming Yiu, and Hongzhi Yin. 2025. A survey on point-of-interest recommendation: Models, architectures, and security. *IEEE Trans. Knowl. Data Eng.*, 37(6):3153–3172.

Pengpeng Zhao, Haifeng Zhu, Yanchi Liu, Jiajie Xu, Zhixu Li, Fuzhen Zhuang, Victor S. Sheng, and Xiaofang Zhou. 2019. Where to go next: A spatio-temporal gated network for next POI recommendation. In *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019*, pages 5877–5884. AAAI Press.

A Datasets

Our experiments are conducted on two widely used travel behavior datasets: **Foursquare** and **Yelp**. Following prior work(Xin et al., 2022; Liu et al., 2025), we identify users who have check-in activity both in their hometown and in out-of-town locations, and reorganize their check-in records to form cross-region travel trajectories. To improve data reliability, we filter out users who have fewer than three check-ins in the destination region, or whose travel itineraries span less than 1 hour or more than 30 days. Furthermore, to ensure dataset quality, we remove POIs visited fewer than twice, and discard users whose travel records do not satisfy all of the following conditions:$|\vec{c}_h| \geq 4$, $|\vec{c}_o| \geq 3$, $\text{freq}(r_h, r_o) \geq 10$. Finally, we split users into training / validation / test sets in an 8:1:1 ratio. Table 3 presents the statistical summary of two datasets.

B Metrics

Let $\hat{V} = [\hat{v}_1, \hat{v}_2, \cdots, \hat{v}_N]$ denote the predicted POI sequence and $V = [v_1, v_2, \cdots, v_N]$ denote the ground-truth POI sequence. Here, r_{v_i} refers to the region of POI v_i, and g_{v_i} refers to its category. Both of these sequences exclude the first and last POIs we input into the model. To evaluate the performance of POI sequence prediction, we adopt the following metrics:

1. Hit Rate (HR). Measures position-wise accuracy by checking whether each predicted POI exactly matches the ground-truth POI at the same index.

$$HR = \frac{1}{N} \sum_{i=1}^{N} I\big(\hat{v}_i = v_i\big). \tag{11}$$

2. Recall Rate (RR). Evaluates the fraction of unique ground-truth POIs covered by the predictions, regardless of their positions.

$$RR = \frac{\left|\hat{V} \cap V\right|}{N}. \tag{12}$$

3. Edit Distance (ED). Computes the normalized Levenshtein edit distance between predicted and ground-truth POI sequences.

Table 3: Statistics of the two datasets.

Dataset	Users	Regions	POIs	Hometown check-ins	Out-of-town heck-ins
Foursquare	3,007	21	23,884	109,225	16,994
Yelp	4,417	214	29,930	58,403	20,479

4. Dynamic Time Warping Distance (DTW). Measures sequence alignment cost based on $0/1$ matching, where equality costs 0 and inequality costs 1.

$$DTW = \min_{\pi} \sum_{(i,j)\in\pi} I\big(\hat{v}_i \neq v_j\big). \tag{13}$$

where π denotes a warping path aligning $\hat{V}$ and V.

5. Category Match (CM). Measures the consistency in POI categories in the position.

$$CM = \frac{1}{N}\sum_{i=1}^{N} I\big(g_{\hat{v}_i} = g_{v_i}\big). \tag{14}$$

6. Region Match (RM). Measures the consistency in POI regions in the position.

$$RM = \frac{1}{N}\sum_{i=1}^{N} I\big(r_{\hat{v}_i} = r_{v_i}\big). \tag{15}$$

C Implementation Details

We set the mask proportion ρ of MAE to 0.75, the number of prototypical style embeddings to 8, and both λ_D and λ_R to 0.1. The reward weights are set as $\lambda_{HR} = 2$, $\lambda_{RR} = 0.5$, $\lambda_{CM} = 0.5$, and $\lambda_{DIV} = 1$. We train our model using the Adam optimizer with a learning rate of 1×10^{-3}. The hidden layer dimension is set to 256, batch size is 4, and we apply early stopping if the validation metric does not improve for 2 epochs. For both our proposed model and all LLM-based baselines, we employ Qwen3-8B(Yang et al., 2025) as the backbone model, with experiments conducted on four NVIDIA GeForce RTX 3090 GPUs. We fine-tune the LLM using LoRA with the following configuration: we set the LoRA rank parameter $r = 16$ and the scaling factor $\alpha = 32$ with a batch size of 1 per GPU, gradient accumulation steps set to 8, and a sequence length of 2, 048 tokens. Each input consists of the most recent 50 check-in records of the user. During the SFT stage, the learning rate is set to 2×10^{-5}, while in the RL stage, it is increased to 5×10^{-5}.

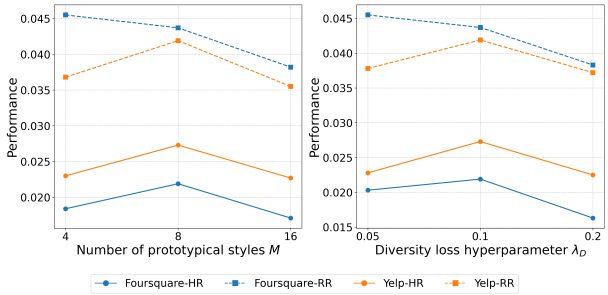

Figure 4: Effects of the number of prototype style embeddings M and diversity regularization coefficient λ_D on two datasets.

D Baselines

We compare our proposed framework against several representative baselines:

- LSTM(Hochreiter and Schmidhuber, 1997): LSTMs are capable of capturing both short-term and long-term dependencies in sequential patterns, making them more effective for a range of sequential data tasks.
- GETNext(Yang et al., 2022): GETNext is a transformer-based approach that leverages a global trajectory flow map constructed in a user-agnostic manner to enhance next POI prediction. It also introduces a GCN module to generate effective POI embeddings. However, GETNext assumes that all users appearing in the test set are also present in the training set. In contrast, our task focuses on predicting POI sequences for unseen users in out-of-town scenarios, where the test users do not appear in the training data. To adapt GETNext to our setting, we select the user in the training set with the highest Jaccard similarity in POI sequence as the proxy ID for each test user. Moreover, since GETNext is not a sequential generation model, we construct the final predicted POI sequence by selecting the top-N POIs with the highest confidence scores in order.
- STHGCN(Yan et al., 2023): STHGCN constructs a hypergraph to capture both inter and intra-user relations, and proposes a hypergraph transformer

Table 4: Hyper-parameter sensitivity analysis on M and λ_D.

Config	Foursquare						Yelp					
	HR	RR	ED	DTW	CM	RM	HR	RR	ED	DTW	CM	RM
$M = 16, \lambda_D = 0.1$	0.0171	0.0382	0.9839	0.4920	0.0322	**1.0000**	0.0227	0.0355	0.9791	0.4896	0.0554	0.9385
$M = 16, \lambda_D = 0.2$	0.0125	0.0370	0.9874	0.4937	0.0356	**1.0000**	0.0226	0.0325	0.9784	0.4892	0.0629	**0.9578**
$M = 4, \lambda_D = 0.05$	0.0187	0.0367	0.9808	0.4904	0.0314	**1.0000**	0.0225	0.0357	0.9780	0.4890	**0.0706**	0.9433
$M = 4, \lambda_D = 0.1$	0.0184	**0.0455**	0.9846	0.4923	0.0364	**1.0000**	0.0230	0.0368	0.9789	0.4893	0.0625	0.9385
$M = 8, \lambda_D = 0.05$	0.0203	**0.0455**	0.9798	0.4889	0.0377	**1.0000**	0.0228	0.0378	0.9792	0.4896	0.0632	0.9522
$M = 8, \lambda_D = 0.1$	**0.0219**	0.0437	**0.9778**	**0.4876**	**0.0572**	**1.0000**	**0.0273**	**0.0419**	**0.9701**	**0.4843**	0.0703	0.9449
$M = 8, \lambda_D = 0.2$	0.0163	0.0383	0.9870	0.4935	0.0349	**1.0000**	0.0225	0.0372	0.9795	0.4898	0.0617	0.9344

Table 5: Ablation results on all evaluation metrics.

Model	Foursquare						Yelp					
	HR	RR	ED	DTW	CM	RM	HR	RR	ED	DTW	CM	RM
w/o SPAM	0.0159	0.0410	0.9838	0.4919	0.0275	**1.0000**	0.0251	0.0387	0.9726	0.4862	0.0580	0.9346
w/o RL	0.0192	0.0422	0.9803	0.4901	0.0552	**1.0000**	0.0227	0.0389	0.9773	0.4887	0.0593	0.9379
w/o SFT	0.0070	0.0306	0.9930	0.4965	0.0282	**1.0000**	0.0224	0.0366	0.9776	0.4888	0.0569	0.9373
SemPOI-RL	**0.0219**	**0.0437**	**0.9778**	**0.4876**	**0.0572**	1.0000	**0.0273**	**0.0419**	**0.9701**	**0.4843**	**0.0703**	**0.9449**

Table 6: Effect of different reward settings in RL. The tuple denotes the weights of (HR, RR, DIV, CAT).

Setting	Dataset	HR	RR	ED	DTW	CM	RM
Default $(2 : 0.5 : 0.5 : 1)$	Foursquare	0.0219	0.0437	**0.9778**	**0.4876**	0.0572	**1.0000**
Default $(2 : 0.5 : 0.5 : 1)$	Yelp	0.0273	0.0419	**0.9701**	**0.4843**	0.0703	**0.9449**
w/o DIV $(2 : 0.5 : 0.5 : 0)$	Foursquare	**0.0225**	0.0431	0.9783	0.4880	**0.0574**	**1.0000**
w/o DIV $(2 : 0.5 : 0.5 : 0)$	Yelp	**0.0279**	0.0413	0.9706	0.4847	**0.0705**	0.9445
HR:RR= $1 : 1$ $(1 : 1 : 0.5 : 1)$	Foursquare	0.0213	**0.0443**	0.9781	0.4879	0.0569	**1.0000**
HR:RR= $1 : 1$ $(1 : 1 : 0.5 : 1)$	Yelp	0.0267	**0.0425**	0.9704	0.4846	0.0700	0.9442

to address the cold-start problem. As its input–output structure is consistent with GETNext, we apply exactly the same experimental adaptation settings when transferring STHGCN to our proposed task.

- MatTrip(Zhang et al., 2024): This method adopts dual LSTM-based encoders to jointly capture users' category-level preferences and the geographical proximity of POIs. An optimized search strategy is further applied to generate user-preferred trip sequences.

- AR-Trip(Shu et al., 2024): Built upon a Transformer encoder-only architecture, AR-Trip incorporates prior positional information as auxiliary input to effectively reduce repetitive POIs and recommend more diverse trajectories.

- KDDC(Liu et al., 2024b): This approach leverages a knowledge-graph-based representation to strengthen semantic interactions. By modeling the distributions of conformity and individual interests, it aligns the recommended POIs more closely with target user preferences. For fair comparison, we select the top-probability POIs (excluding start and end points) as the final predictions.

- SPOT-Trip(Liu et al., 2025): SPOT-Trip introduces an ordinary differential equation (ODE) formulation to model the continuous evolution of latent dynamic user preferences. It further incorporates a temporal point process to characterize the instantaneous probability of preference transitions.

- LLMMove(Feng et al., 2024): LLMMove employs LLMs to extract user preferences and geographical distances via carefully designed prompts. Since our task focuses on out-of-town prediction, we extend the original design by sampling 100 negative candidates in the prompt, allowing the LLM to select the most plausible destination POIs.

- LLM4POI(Li et al., 2024): This SFT-based LLM recommender reformulates the next POI predic-

tion task as a question-answering problem. It introduces prompt-based trajectory similarity to jointly leverage historical trajectories and cross-user trajectories. For our setting, we provide both the user's hometown trajectory and the start–end POIs in the destination city, and directly predict the full trajectory in the destination.

- Refine-POI(Li et al., 2025): A reinforcement learning-based LLM recommender designed for next POI recommendation. It incorporates some recommendation-driven reward functions to enable the LLM to generate effective top-k recommendation lists.

E Hyper-parameter Analysis

We further investigate the sensitivity of **SemPOI-RL** to key hyper-parameters, including the number of prototype style embeddings M, the diversity regularization coefficient λ_D in SPAM, and the weight allocation in the RL reward. Figure 4 and Table 4 summarize the results. We observe that the model is moderately sensitive to M and λ_D, but the default setting $M = 8$ and $\lambda_D = 0.1$ consistently provides the best overall balance across datasets.

When the number of prototype styles is too small, the model cannot adequately represent different temporal phases within a trip. When M is too large, the style space becomes overly fragmented, which harms generalization and weakens alignment quality. This trend is particularly evident on Foursquare, where excessively fine-grained prototypes reduce both HR and RR.

The coefficient λ_D controls the diversity among prototype style embeddings. A very small λ_D leads to homogeneous prototypes, limiting the model's ability to capture stage-specific travel semantics. In contrast, an overly large λ_D pushes prototypes to be too dissimilar, which hurts recommendation accuracy. The best trade-off is achieved at $\lambda_D = 0.1$.

For the RL reward, our default weight setting is $(2, 0.5, 0.5, 1)$ for *Hit*, *Recall*, *Diversity*, and *Category*, respectively. This design intentionally prioritizes correctly positioned POI hits over merely recovering relevant POIs, because ordered trajectory generation requires stronger positional precision. The sensitivity results in Table 6 show that the model is reasonably robust to moderate changes in reward weights, although different trade-offs may slightly favor HR or RR.

F Prompt Templates for Travel Style Generation

In this appendix, we provide the two prompt templates used in Stage 1. The first prompt asks the LLM to summarize a user's destination-oriented travel style from the observed destination trajectory. The second prompt asks the LLM to infer the user's likely destination travel style from the hometown trajectory. In both prompts, the region name is represented by `{region_name}` and the trajectory text is represented by `{trajectory_text}`.

F.1 Prompt for Summarizing Destination Travel Style

```
You are an expert in user travel behavior
analysis.

A user has visited the following points of
interest in {region_name}.

Destination trajectory:
{trajectory_text}

Please summarize the user's travel style in this
 region based on the trajectory within 200 words.

Your summary should describe the user's overall
preferences, such as activity patterns,
lifestyle tendencies, and possible interests
reflected by the visited places.

Requirements:
1. Write one coherent paragraph in natural
language.
2. Focus on high-level travel style rather than
listing POIs one by one.
3. Capture semantic preferences such as leisure,
 nightlife, food exploration, cultural interest,
 business activity, outdoor preference, or
shopping tendency.
4. Do not mention that you are an AI model.
5. Do not output bullet points or JSON.
6. Keep the summary concise but informative.

Travel style summary:
```

F.2 Prompt for Inferring Destination Travel Style from Hometown Trajectory

```
You are an expert in user travel behavior
analysis.

A user has the following historical trajectory
in their hometown region {region_name}.

Hometown trajectory:
{trajectory_text}

Based on this hometown trajectory, infer what
travel style the user is likely to exhibit when
visiting another city within 200 words. Your
goal is not to describe the hometown itself, but
 to infer the user's destination-oriented travel
 preferences in an out-of-town scenario.
```

```
Requirements:
1. Write one coherent paragraph in natural
language.
2. Infer high-level travel style and likely
preferences in a new city.
3. Focus on transferable interests, such as
whether the user may prefer restaurants,
nightlife, cultural attractions, shopping areas,
 parks, business-related places, or mixed urban
exploration.
4. Do not simply restate the visited POIs;
provide an abstract style-level inference.
5. Do not mention that you are an AI model.
6. Do not output bullet points or JSON.
7. Keep the summary concise but informative.

Inferred destination travel style:
```