KP4POI: Efficient POI Recommendation on Large-scale Datasets via Knowledge Prompting of Venues and Users

Yuuki Tachioka[1]

[1]Denso IT Laboratory, 13F Shintora Yasuda Bldg., 4-3-1 Shimbashi, Minato-ku, Tokyo, Japan

Abstract

Accurate prediction of the next point of interest (POI) that a user will visit is important to location-based services for personalized and context-aware recommendation. Traditional approaches, particularly those based on collaborative filtering, struggle to fully utilize the semantic, temporal, and spatial information embedded in user behavior data. Although the emergence of large language models (LLMs) can capture semantic information, large-scale LLMs are infeasible for large-scale data due to computational constraints, as our experiment shows. In this paper, we propose an efficient and lightweight POI recommendation framework, KP4POI, that leverages small-scale LLM enhanced with knowledge prompting, which encodes knowledge graphs of venue and user into natural language. By transforming user and venue knowledge graphs into natural language prompts, our method enables LLMs to incorporate semantic, geographic, and social information effectively. Experiments on real-world datasets demonstrate that our approach improves recommendation quality while maintaining scalability.

Keywords

Point of Interest Recommendation, Large Language Model, Knowledge Prompting, Sequential Recommendation, Privacy-aware Recommendation, Location-based Services

1. Introduction

The proliferation of location-based services has precipitated an increasing demand for systems capable of recommending the next Point of Interest (POI) that is relevant and useful to the user, conditioned on their spatio-temporal check-in history. Such predictive capabilities are pivotal not only for improving user-centric experiences, but also for optimizing marketing strategies through context-aware personalization [1]. Traditional POI recommendation methods are mainly based on collaborative filtering (CF), which models user-POI interactions using unique identifiers [2]. Although effective to some extent in small computational load, these methods are limited in modeling semantic similarity among POIs and fail to consider sequential and spatial dependencies in check-in behaviors.

Subsequently, many hybrid methods have been proposed that incorporate categorical, temporal, or contextual information [3, 4, 5, 6]. Among them, recent deep learning approaches have employed graph neural networks to encode high-order dependencies in heterogeneous spatial graphs [7, 8], and hybrid models that combine matrix factorization with deep interactions have shown promise in mitigating data sparsity and semantic heterogeneity [9, 10, 11]. Furthermore, sequential recommendation techniques have been adapted to POI scenarios, using temporally localized user trajectories to model higher-order transitions and intent-aware behavior [12, 13].

With the rise of large language models (LLMs), new avenues for context-aware recommendation have emerged [14, 15, 16]. Two main types of LLM-based POI recommendation methods have gained attention: prompt-based methods and embedding-based methods, as shown in Figs. 1 (a) and (b). Prompt-based methods convert recommendation tasks into natural language format and use compact LLMs to generate candidates [17]. These methods are computationally efficient but often lack the ability to integrate structured knowledge such as venue categories or spatial relationships. Embedding-based methods generate semantic user embeddings from user check-in history and POI embeddings from POI attributes by using large-scale LLMs [18, 19, 20, 21], and calculate the similarity between them at

Workshop on Recommenders in Tourism (RecTour 2025), September 22, 2025, co-located with the 19th ACM Conference on Recommender Systems, Prague, Czech Republic.
tachioka.yuki@core.d-itlab.co.jp (Y. Tachioka)
0009-0002-0587-2943 (Y. Tachioka)

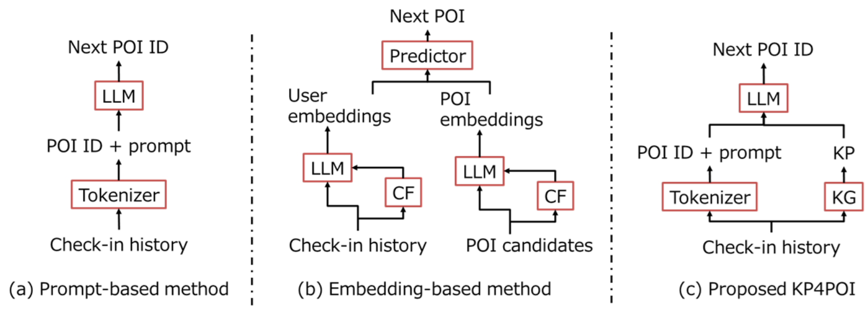

Figure 1: Overview of LLM-empowered POI recommendation methods: (a) Prompt-based, (b) Embedding-based, and (c) Our proposed KP-enhanced prompt-based method, KP4POI, using venue and user KGs.

the cost of heavy fine-tuning and inference. Although expressive, these methods are computationally demanding and impractical for large-scale datasets [21, 22].

To address this challenge, we designed a framework, KP4POI, that enriches prompt-based methods with structured knowledge through *knowledge prompting* (KP) [23, 24], as shown in Fig. 1 (c). Specifically, we transform venue and user knowledge graphs (KGs) into natural language prompts, allowing compact LLMs to effectively utilize not only venue attributes (e.g., category, location) but also user context (e.g., social connections) during prediction. In contrast to embedding-based approaches, our method achieves efficiency by leveraging compact LLMs, while the integration of KGs can help mitigate hallucinations in LLM predictions and improve accuracy [25]. Moreover, because our method uses structured knowledge, such as venue categories, addresses, or user communities, to substitute for raw location coordinates, it partially supports privacy-aware settings where fine-grained location data may not be available [26]. In such scenarios, user-level knowledge, such as social groupings, can further complement missing venue information, enabling robust recommendations even under limited or anonymized metadata.

In this paper, we compare our KP4POI with embedding-based methods in terms of fine-tuning efficiency on large-scale POI datasets. We examine the contribution of different types of knowledge—specifically, categorical affinity and geographic proximity—and assess the utility of user KP derived from social graphs [27]. Our findings demonstrate that compact LLMs augmented with structured knowledge offer a practical and scalable solution for personalized POI recommendation.

2. LLM-empowered POI recommendation

As discussed in Section 1, LLM-empowered POI recommendation methods can be categorized into prompt-based and embedding-based approaches. One of the most important prompt-based methods is P5. The P5 uses a prompt to reformulate the recommendation task as a natural language task by encoding the IDs of the recommendation items [17] with the masked personalized prompt (MPP) [17, 23]. Fig. 2 shows an example of MPP templates, where $\mathcal{T}$ denotes a set of prompt templates to encode historical check-ins as natural language. The MPP $X_{mpp}(u, H_u^m; \mathcal{T})$ is constructed for the user u based on the IDs of previously visited m venues $H_u^m = \{v_1, \dots, v_m\}$. The next POI target is masked using multiple templates $\mathcal{T}$, where the wider variety of templates is important [28]. Based on this prompt, an LLM can predict the next POI ID, denoted v_{m+1}, that the user may be interested in. Combining the venue knowledge prompt with the standard MPP in the following final input:

$$X_p(u, H_u^m) = [SP]X_{mpp}(u, H_u^m; \mathcal{T})[SP]. \tag{1}$$

In Eq. (1), the special token [SP] serves as a delimiter separating the user's check-in history from the appended knowledge prompts. To minimize vocabulary size and improve generalization, efficient tokenization represents IDs split into two-digit segments, e.g., 1298 as <12><98>. Since the output in P5 is ID, high generative capacity is not required and performance and speed are balanced using a compact LLM such as T5 [29]. Advanced recommendation methods such as Prompt Distillation for Efficient LLM-based Recommendation (POD) [28] and KP [23] are based on P5. POD distills discrete

$X_{mpp}(u, H_u^m; \mathcal{T}_1)$: User $\{u\}$ has previously visited $\{H_u^m\}$ and is now going to {masked target POI}.

$X_{mpp}(u, H_u^m; \mathcal{T}_2)$: User $\{u\}$'s travel history includes $\{H_u^m\}$, but now they will call {masked target POI}.

$X_{mpp}(u, H_u^m; \mathcal{T}_3)$: User $\{u\}$'s past travel history involves $\{H_u^m\}$, but {masked target POI} is the upcoming place they will be interested in.

Figure 2: Examples of a masked personalized prompt X_{mpp} using different types of templates $\mathscr{T}$.

KG: ({poi},category,{category})
→ X_{kp}^v: The category of {poi} is {category}.

KG: ({poi},address,{address})
→ X_{kp}^v: The address of {poi} is {address}.

KG: ({poi},latlon,{latlon})
→ X_{kp}^v: The latitude and longitude of {poi} is {latlon}.

Figure 3: Examples of a venue knowledge prompt X_{kp}^v.

prompt templates into task-specific continuous prompt vectors, thus improving the training efficiency of LLM-based recommendation systems. Functionally, POD and P5 differ only in that POD uses multiple prompt templates [28].

Instead of prompting, embedding-based methods directly use LLM embeddings to make recommendations [21]. However, embedding-based methods that use large-scale LLMs demand a high computational load because the entire check-in history must be inputted to LLM for inference as a natural language prompt. Both fine-tuning and inference require high computational load. Section 4.2.1 compares the fine-tuning time of two types of methods.

3. KP4POI: prompt-based POI recommendation with knowledge prompting

Building on the motivation outlined in Section 1, we present KP4POI, a prompt-based POI recommendation framework enhanced by structured KP. KP4POI leverages knowledge about the venue and user in a natural language format to guide LLMs for effective recommendation. This section details the architecture of KP4POI and describes how KPs are generated and integrated into the LLM pipeline.

3.1. Venue knowledge prompting

The knowledge about the venues can be represented as KGs. Public datasets used in the experiment may provide venue information, such as category or latitude / longitude (lat/lon), which can be converted to addresses. For POI recommendation, categorical affinity and geographical proximity play an important role [30]. Categorical affinity [31, 32] refers to co-occurrence of categories (e.g., visiting a train station before an office), while geographical proximity [33, 34] involves nearby locations (e.g., dining at a restaurant close to the office). This knowledge can be concatenated into MPPs in the form of a KP. To measure geometrical proximity, to address privacy concerns, reverse-geocoded address features are aggregated at the street level (NYC) or chome (block) (Tokyo), ensuring that no fine-grained and personally identifiable location information is used. At inference time, these features are retrieved from venue metadata and do not require real-time or private user data.

KGs consist of triples (head / relation / tail) represented as (h, r, t). To incorporate KG knowledge into LLM, a KP transforms a KG triple using a relation-specific template $\mathscr{T}_r$, as illustrated in Fig. 3. A KG $\kappa = (v, \text{category}, \textit{category})$ is converted using the relation-related template $\mathscr{T}_r(\kappa)$ = "The category of v is *category*" when r = category. Let $\mathscr{K}_v$ be the set of KGs related to the venue v. A venue KP X_{kp}^v can be generated from the KGs of the venues visited H_u^m as follows: $X_{kp}^v(H_u^m, \mathscr{K}_v) = \bigcup_{v \in H_u^m} \bigcup_{\kappa \in \mathscr{K}_v} \mathscr{T}_r(\kappa)$, where $\bigcup$ denotes the concatenation of prompts. The final prompt combining venue

KG: ({user},community,{community})
→ X^u_{kp}: User_{user}'s community is {community}.

Figure 4: Example of a user knowledge prompt X^u_{kp}.

knowledge is given by:

$$X_p(u, H^m_u, \mathcal{K}_v) = X_p(u, H^m_u) X^v_{kp}(H^m_u, \mathcal{K}_v)[SP]. \tag{2}$$

For example, for a user who visited v_1 ($\kappa_1 = (v_1, \text{category}, A)$) then v_2 ($\kappa_2 = (v_2, \text{category}, B)$)), the final prompt becomes: "$[SP]X_{mpp}[SP]$ The category of v_1 is A. The category of v_2 is B. $[SP]$".

3.2. User knowledge prompting

User-specific information has been shown to improve the performance of recommendation [27]. Attributes such as gender and age can offer valuable personalization signals. However, such personal data are rarely available in POI recommendation due to heightened privacy concerns, especially because high-resolution location data can inadvertently expose sensitive details such as a user's residence or workplace. In contrast, location-based social networks like Foursquare often include user social graphs, which are less sensitive and are more readily available. These social graphs can implicitly capture behavioral traits and allow for unsupervised grouping of users with similar characteristics. When a social graph is available, user clustering can be effective for POI recommendation [35].

To take advantage of this, we apply graph partitioning methods [36] to identify user communities in the social graph. Each user is assigned a community that represents interpretable group-level characteristics. Given that the POI recommendation benefits from approaches using graph neural networks [7, 8] or matrix factorization [9, 10, 11], we derive community labels using node embeddings [37] using node2vec, and cluster them [38] by k-means++ clustering [39]. Each cluster ID serves as a community-level pseudo-attribute label that represents the user's social context, where each user is assigned to exactly one cluster. These community labels can be encoded as KG triples ($u, community, label$). We can convert them into natural language prompts using a relation-specific template as in Fig. 4, and incorporate them into the model via user KP. This allows the compact LLM to benefit from social information even in privacy-aware settings where fine-grained personal metadata are not accessible[1].

Let $\mathcal{K}_u$ be the set of KG triples related to user u. Then, a user KP X^u_{kp} is $X^u_{kp}(u, \mathcal{K}_u) = \bigcup_{\kappa \in \mathcal{K}_u} \mathcal{T}_r(\kappa)$ and the final prompt combining user knowledge is given by:

$$X_p(u, H^m_u, \mathcal{K}_v, \mathcal{K}_u) = X_p(u, H^m_u, \mathcal{K}_v) X^u_{kp}(u, \mathcal{K}_u)[SP]. \tag{3}$$

In KP4POI, the loss function maximizes the likelihood of the next POI v_{m+1} in the model θ as

$$\mathcal{L}_\theta = -\sum_u \sum_{m=1}^{M} \log P_\theta(v_{m+1} | X_p(u, H^m_u, \mathcal{K}_v, \mathcal{K}_u)). \tag{4}$$

For inference, the beam search is used to predict the next POI list.

4. Experiments

4.1. Experimental setup

For POI recommendation tasks, LLM-based methods have generally been effective [21]. Among those utilizing large-scale LLM such as Llama, SeCor and soft prompting [40] demonstrated the highest performance. Both SeCor and our method are fine-tuned from a pre-trained Llama2-7B and T5-small, respectively. SeCor requires an additional step of CF training. However, as shown in Table 2, the use of

[1]Note that our approach derives community labels by clustering user embeddings obtained from the social graph. Therefore, user KP is applicable only to datasets that include explicit social relations (WWW2019 and Gowalla). For datasets without social graphs (NYC and TKY), user KP cannot be constructed.

Table 1
Summary and statistics of datasets used in experiments.

Dataset	NYC	TKY	TIST2015	WWW2019	Gowalla
Source	Foursquare	Foursquare	Foursquare	Foursquare	Gowalla
Scale	Small	Small	Large	Large	Medium
Social Graph	-	-	-	Yes	Yes
Time Period	Apr 2012 –Feb 2013	Apr 2012 –Feb 2013	Apr 2012 –Sep 2013	Apr 2012 –Jan 2014	Feb 2009 –Oct 2010
Users	1,083	2,293	243,004	114,324	69,705
Venues	15,624	24,321	1,564,541	1,312,372	536,810
Check-ins	198,593	525,710	30,477,062	19,680,643	5,327,596
Ave. Check-ins	183.3	229.3	125.4	172.2	76.4
Category Types	262	226	428	441	-
Addresses	2,761	3,708	414	629	500
Lat/Lon Entries	8,380	10,773	776,281	831,263	387,750
User Edges	-	-	-	607,333	325,354

large-scale LLMs is impractical for large-scale datasets[2]. Except for methods using large-scale LLMs, the POD and collaborative filter-based seq2graph [41] performed comparable.

We compare our method [3] with P5[4] and SeCor, as they represent two contrasting baselines: P5 as a prompt-based approach and SeCor as an embedding-based approach using large-scale LLM. We primarily compare with P5 because it represents a compact, prompt-based LLM recommender that is closest in spirit to our framework. While traditional baselines such as popularity or matrix factorization would also be informative, we leave their integration for future work since our focus here is to evaluate the efficiency and effectiveness of knowledge prompting relative to existing LLM-based approaches.

We used T5-small (60.5M parameters) as the backbone, which is much smaller than Llama-7B. The model used 512 dimensions and eight multihead attention. AdamW was used with a maximum learning rate of 0.001, batch size 64, and 100 epochs. The best model was selected on the basis of the validation performance. Following previous work [21, 42], the check-ins were sorted chronologically by user. Users with fewer than 10 check-ins and venues visited only once or twice were removed. Three types of venue KG were used: address, lat/lon[5], and category such as restaurant. KP was implemented based on KP4SR[6]. The evaluation used NDCG@K and Recall@K for $K = \{5, 10, 50\}$ in a leave-one-out setting[7]. The last two venues were reserved for testing and predicted by models fine-tuned on the remaining data. We used a paired t-test with Bonferroni correction to assess the significance between P5 and our KP4POI.

Table 1 shows the statistics of four Foursquare datasets[8] (NYC, TKY, TIST2015, and WWW2019) and one Gowalla dataset[9], where Foursquare and Gowalla are location-based social networks with user check-ins. NYC and TKY are small, commonly used, and suitable for parametric studies. TIST2015 and WWW2019 are used to test scalability, because they are large-scale. Among the four datasets, only WWW2019 and Gowalla contain explicit social graphs. Since our method requires social graphs to derive community labels via clustering, user KP is only available for these two datasets

The NYC and TKY dataset [43] contains check-ins in New York City (NYC) and Tokyo (TKY), where

[2]Although we used the official implementation (https://github.com/siri-ya/SeCor), we could not reproduce the results reported in [21].
[3]https://github.com/DensoITLab/KP4POI
[4]P5 and POD are functionally equivalent, except that POD uses multiple prompt templates [28], while P5 in our implementation employs 11 templates to ensure comparability.
[5]In [21], lat/lon was converted to address. We validated the effectiveness of this.
[6]https://github.com/zhaijianyang/KP4SR
[7]We focus on accuracy metrics (NDCG, Recall) as in prior work, leaving diversity/coverage for future work.
[8]https://sites.google.com/site/yangdingqi/home/foursquare-dataset
[9]https://snap.stanford.edu/data/loc-gowalla.html

Table 2
Fine-tuning time per epoch. Estimated times are shown in parentheses. The fine-tuning time for P5 and KP4POI was calculated as the average over 100 epochs. SeCor uses Llama2-7B as backbone. SeCor's fine-tuning time excludes the collaborative filtering training step.

Method	NYC	TKY	TIST2015	WWW2019	Gowalla
P5(POD)	1.5s	3.8s	4m44s	5m50s	3m31s
KP4POI	6.0s	9.1s	6m6s	7m52s	4m40s
SeCor	10h7m	25h29m	(79d19h)	(98d9h)	(59d6h)

Table 3
Performance comparison on NYC and TKY in terms of NDCG and Recall for P5 and knowledge prompting (KP) variants, where "add" represents address, "lat/lon" is latitude and longitude, and "cat" is category. A paired t-test was conducted with P5; † indicates significance at $\alpha = 0.1$, * at $\alpha = 0.05$, and ** at $\alpha = 0.01$ levels.

	NYC				TKY			
	NDCG		Recall		NDCG		Recall	
Method	@5	@10	@5	@10	@5	@10	@5	@10
P5(POD)	0.2383	0.2437	0.3084	0.3250	0.2135	0.2243	0.2861	0.3188
KP4POI(add)	0.2395	0.2431	0.3121	0.3232	0.2162	$0.2287^{\dagger}$	$0.2904^{\dagger}$	0.3280^{**}
KP4POI(lat/lon)	0.2385	0.2436	0.3102	0.3259	0.2175	$0.2287^{\dagger}$	$0.2883^{\dagger}$	0.3227^{*}
KP4POI(add+lat/lon)	0.2399	0.2451	0.3121	0.3278	0.2144	0.2256	0.2887	0.3227
KP4POI(cat)	0.2457^{*}	0.2516^{*}	0.3075	0.3250	0.2197^{*}	0.2307^{*}	0.2909^{**}	0.3245^{**}
KP4POI(cat+add)	0.2521^{**}	0.2562^{**}	0.3112^{**}	0.3232	0.2196^{*}	$0.2291^{\dagger}$	0.2918^{*}	$0.3210^{\dagger}$
KP4POI(cat+add+lat/lon)	0.2508^{**}	0.2560^{**}	0.3130^{**}	0.3287^{**}	0.2197^{*}	$0.2290^{\dagger}$	0.2913^{**}	0.3201

each check-in includes a timestamp, lat/lon, and category. The addresses were obtained by reverse geocoding: NYC via Nominatim[10] ("street" level), TKY via Yahoo![11] ("chome" level). To avoid invading privacy and reduce the number of reverse geocoding, the coordinates were rounded to three decimal places. The TIST 2015 dataset [44] includes check-ins from 415 cities in 77 countries, where each city had at least 10K check-ins, and the city names were used as addresses. The WWW 2019 dataset [45] covers global check-ins[12] and includes check-ins and user social graphs[13]. The Gowalla dataset [46] covers global check-ins[14] and includes a social graph, where because many users had few check-ins, we retained only those with 10+ check-ins (69,705 of 196,591).

4.2. Results and Discussions

We designed four types of research questions: one concerning fine-tuning time and the applicability of LLM (4.2.1), two that address venue KP (4.2.2 and 4.2.3), and one exploring user KP (4.2.4).

4.2.1. RQ1: Can large-scale LLMs be applied to large-scale POI recommendation?

Table 2 shows the fine-tuning time required by our KP4POI using T5-small and SeCor using Llama2-7B. The gap between the two is significant. In fact, even in the experiments of [21], only 850,010 check-ins out of 5 million were used from the Gowalla dataset. Although parallelization makes it feasible to run experiments on datasets of the NYC/TKY scale, applying large-scale LLMs solely to obtain embeddings for large-scale data is impractical. Therefore, it is essential to improve performance by using compact

[10]https://nominatim.org/
[11]https://developer.yahoo.co.jp/webapi/map/openlocalplatform/v1/reversegeocoder.html
[12]This does not include city names but does include lat/lon and country codes. We used 629 major cities (TIST2015 cities + capital cities) and matched check-ins to the closest city with the same country code.
[13]Although two types of social graph are provided, we used the newer one.
[14]Because country codes are not provided, we matched each check-in to the nearest city using lat/lon.

LLMs. In addition to the significant difference in the fine-tuning time (Table 2), several practical implications emerge.

- **Scalability**: Our method handles tens of millions of check-ins with fine-tuning times under 10 minutes per epoch, making it practical for real-time model updates in production.
- **Cost-efficiency**: Unlike large-scale LLMs, which require substantial GPU memory and power, T5-small requires only modest computational resources, making it accessible to institutions with limited hardware.
- **Maintainability**: The lightweight architecture allows for rapid fine-tuning as user behavior evolves, facilitating continuous learning without downtime or extensive costs.
- **Broader applicability**: The method is suitable not only for centralized cloud systems, but also for edge deployment scenarios, promoting POI personalization in privacy-sensitive or offline-first environments.

These findings reaffirm that KP4POI that enhances small LLMs with structured knowledge is a practical and scalable direction for future POI recommendation systems.

4.2.2. RQ2: Which is more effective: address or latitude and longitude?

We compare KP4POI(add) and KP4POI(lat/lon) to assess the contribution of different types of geographical knowledge[15]. Although Table 3 shows only small differences between address and lat/lon in the NYC dataset due to the limited granularity of US street-level addresses, significant gains in TKY highlight the usefulness of the hierarchical structure of Japanese addresses (e.g., prefecture → city → chome (block)), which encodes spatial clustering and provides richer cues than raw coordinates. This also explains why the overall contribution of address-based knowledge may appear small in some datasets: when addresses lack hierarchical depth or are too coarse, they fail to capture the fine-grained distances that strongly influence tourist behaviour. In practice, however, address-based knowledge remains easier to interpret, more stable against noise, and often more accessible in privacy-sensitive or anonymized datasets where lat/lon is not disclosed. In contrast, latitude/longitude directly preserves precise spatial proximity. Thus, while their relative performance differences are modest, the two variants capture complementary aspects of geographical information, and address-level knowledge can sometimes substitute for exact coordinates in real-world applications.

4.2.3. RQ3: Which is more effective: geographical information or category?

Table 3 also includes performance metrics when KP4POI(cat) is applied to the NYC and TKY datasets. Address information contributes mainly to improvements in Recall, while category information mainly enhances NDCG, which shows that geographical information and category can have different types of useful information. The last two rows of Table 3 show the results when combining the address or lat/lon with the category. In the NYC dataset, this combination significantly improved performance in all metrics (with 1% significance), demonstrating the effectiveness of using both types of knowledge. KP4POI(cat+add+lat/lon) improved NDCG@5 by 1.3% compared to P5(POD) in NYC.

Table 4 presents the results of similar experiments conducted on a large-scale dataset. A consistent trend is also observed here: The combination of both leads to the best results from both perspectives. These results suggest that category-based knowledge tends to improve ranking precision (NDCG), while geographical knowledge often improves recall. Although not absolute, this complementary effect was observed consistently across datasets. Importantly, combining both sources of knowledge results in statistically significant improvements in all metrics (Table 3, Table 4). This demonstrates the complementary nature of the two: while semantic similarity guides "what" the user wants, the spatial context informs "where" it is realistically accessible. This robustness is particularly beneficial in large-scale settings like TIST2015 and WWW2019, where venue diversity and user heterogeneity are

[15]KP4POI without add is functionally equivalent to P5, as it only uses sequential prompts without additional knowledge. For this reason, we do not report it separately.

Table 4
Performance comparison on large-scale datasets: TIST2015 and WWW2019.

	TIST2015				WWW2019			
	NDCG		Recall		NDCG		Recall	
Method	@5	@10	@5	@10	@5	@10	@5	@10
P5(POD)	0.1815	0.1863	0.2386	0.2533	0.1905	0.1927	0.2480	0.2548
KP4POI(add)	0.1817	0.1928**	0.2413**	0.2753**	0.1915**	0.1988**	0.2508**	0.2730**
KP4POI(cat)	0.1881**	0.1937**	0.2394**	0.2565**	0.1970**	0.1999**	0.2481**	0.2570**
KP4POI(cat+add)	0.1894**	0.2003**	0.2440**	0.2775**	0.1988**	0.2058**	0.2529**	0.2743**

Table 5
Performance comparison on WWW2019 and Gowalla datasets confirming the effectiveness of user knowledge prompting. "u" denotes user community information. Note that venue category information is not available on Gowalla.

Method	WWW2019				Gowalla			
	NDCG		Recall		NDCG		Recall	
Method	@5	@10	@5	@10	@5	@10	@5	@10
P5(POD)	0.1905	0.1927	0.2480	0.2548	0.1929	0.1950	0.2552	0.2617
KP4POI(u)	0.1906	0.1946**	0.2490**	0.2612**	0.1930	0.1956†	0.2553	0.2632**
KP4POI(add)	0.1915**	0.1988**	0.2508**	0.2730**	0.1924	0.1979**	0.2555	0.2721**
KP4POI(add+u)	0.1921**	0.1993**	0.2513**	0.2732**	0.1927	0.1983**	0.2553	0.2723**
KP4POI(cat+add+u)	0.1991**	0.2061**	0.2526**	0.2741**	-	-	-	-

pronounced. For that case, enriching language models with multifaceted knowledge, both semantic and geographic, yields more reliable and context-aware POI recommendations.

4.2.4. RQ4: Does user information improve performance?

Table 5 shows the performance results when the user KP (KP4POI(u)) is applied to the WWW2019 dataset. Except for NDCG@5, all metrics show statistically significant improvements at the 1% level. This indicates that even when only facility IDs are available without lat/lon due to privacy issues, user KP significantly improves performance. When used together with category or address information, the added benefit of user KP is minimal, suggesting that social grouping can be learned implicitly through check-in records.

Table 5 also shows the corresponding results in the Gowalla dataset. However, incorporating address information leads to statistically significant improvements in NDCG@10 and Rec@10 (1% level). Similarly, the user KP significantly improves NDCG@10 and Rec@10, confirming its effectiveness. The combination of address and user information does not provide an additional benefit.

These results demonstrate the effectiveness of KP4POI(u), particularly in environments where venue metadata are limited or unavailable. Community-level embeddings serve as surrogates for missing contextual cues, enabling the model to generalize based on social affinity and collective behavior patterns. Although user KP significantly improves performance when used alone, its marginal benefit in combination with venue information suggests some redundancy between social and spatial signals. For privacy-sensitive contexts, user KP can be effective when explicit social graphs are available (e.g., in Gowalla or WWW2019). However, its utility for cold-start users in real-world scenarios without such graphs may be limited. Its efficiency and interpretability further support its inclusion in scalable POI recommendation pipelines.

Although the contribution of user-knowledge prompts appears small, this can be partly explained by the characteristics of the datasets. Only a fraction of users are connected on social graphs, and clustering from sparse interactions yields coarse community labels. Consequently, the additional signal provided by user KP is weaker than the venue-based knowledge. However, user-level information remains valuable in privacy-sensitive contexts or datasets where precise location data are unavailable, as it can still capture group-level behavioral tendencies.

5. Conclusion

In this paper, we proposed KP4POI, a knowledge-enhanced POI recommendation framework that uses compact LLMs with knowledge prompting derived from venues and users. By reformulating POI recommendation as a natural language task, our method integrates semantic and contextual knowledge without incurring the heavy computational costs of large-scale embedding-based approaches. Extensive experiments on four real-world datasets demonstrated that KP4POI achieves competitive performance while reducing training time and resource requirements. In addition, our ablation study showed that category- and geography-based knowledge prompts contribute differently to recommendation performance, and their complementary effects improve both ranking precision and recall. Although we leave for future work a more comprehensive evaluation, including additional baselines (e.g., popularity, matrix factorization), user cold-start scenarios, and broader recommendation metrics such as diversity and catalog coverage, these results highlight the practicality of incorporating structured knowledge into lightweight LLMs for POI recommendation.

Our findings indicate that KP4POI offers a scalable, interpretable, and privacy-aware solution to POI recommendation. This work opens promising directions for building more intelligent, adaptive, and efficient location-based services using LLMs augmented by external knowledge.

Declaration on Generative AI

During the preparation of this work, the author used ChatGPT-4 and writefull in order to: Grammar and spelling check. After using these services, the author reviewed and edited the content as needed and takes full responsibility for the publication's content.

References

[1] W.-Y. Zhu, W.-C. Peng, L.-J. Chen, K. Zheng, X. Zhou, Exploiting viral marketing for location promotion in location-based social networks, ACM Trans. Knowl. Discov. Data 11 (2016). URL: https://doi.org/10.1145/3001938. doi:10.1145/3001938.

[2] C. Zhang, K. Wang, POI recommendation through cross-region collaborative filtering, Knowledge and Information Systems 46 (2016) 369–387. URL: https://doi.org/10.1007/s10115-015-0825-8. doi:10.1007/s10115-015-0825-8.

[3] J.-D. Zhang, C.-Y. Chow, GeoSoCa: Exploiting geographical, social and categorical correlations for point-of-interest recommendations, in: Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '15, Association for Computing Machinery, New York, NY, USA, 2015, p. 443–452. URL: https://doi.org/10.1145/2766462.2767711. doi:10.1145/2766462.2767711.

[4] H. Wang, H. Shen, W. Ouyang, X. Cheng, Exploiting POI-specific geographical influence for point-of-interest recommendation, in: Proceedings of the 27th International Joint Conference on Artificial Intelligence, IJCAI'18, AAAI Press, 2018, p. 3877–3883.

[5] Y. Su, J.-D. Zhang, X. Li, D. Zha, J. Xiang, W. Tang, N. Gao, FGRec: A fine-grained point-of-interest recommendation framework by capturing intrinsic influences, in: 2020 International Joint Conference on Neural Networks (IJCNN), 2020, pp. 1–9. doi:10.1109/IJCNN48605.2020.9207571.

[6] R. Baral, X. Zhu, S. S. Iyengar, T. Li, ReEL: Review aware explanation of location recommendation, in: Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization, UMAP '18, Association for Computing Machinery, New York, NY, USA, 2018, p. 23–32. URL: https://doi.org/10.1145/3209219.3209237. doi:10.1145/3209219.3209237.

[7] Z. Wang, Y. Zhu, Q. Zhang, H. Liu, C. Wang, T. Liu, Graph-enhanced spatial-temporal network for next POI recommendation, ACM Trans. Knowl. Discov. Data 16 (2022). URL: https://doi.org/10.1145/3513092. doi:10.1145/3513092.

[8] J. Yu, L. Guo, J. Zhang, G. Wang, A survey on graph neural network-based next POI recommendation for smart cities, Journal of Reliable Intelligent Environments 10 (2024) 299–318. URL: https://doi.org/10.1007/s40860-024-00233-z. doi:10.1007/s40860-024-00233-z.

[9] J.-B. Griesner, T. Abdessalem, H. Naacke, POI recommendation: Towards fused matrix factorization with geographical and temporal influences, in: Proceedings of the 9th ACM Conference on Recommender Systems, RecSys '15, Association for Computing Machinery, New York, NY, USA, 2015, pp. 301–304. URL: https://doi.org/10.1145/2792838.2799679. doi:10.1145/2792838.2799679.

[10] J. Jian, Y. Shi, POI recommendation algorithm based on geographical information matrix decomposition, in: 2023 IEEE 6th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), volume 6, 2023, pp. 901–905. doi:10.1109/ITNEC56291.2023.10082387.

[11] X. Li, S. Xu, H. Jin, Z. Wang, Y. Ma, X. He, POI recommendation by deep neural matrix factorization integrated attention-aware meta-paths, Complex & Intelligent Systems 10 (2024) 8163–8177. URL: https://doi.org/10.1007/s40747-024-01596-9. doi:10.1007/s40747-024-01596-9.

[12] Y. Zou, K. Zhao, Short-term POI recommendation with personalized time-weighted latent ranking, Discover Computing (2024) 16–27. URL: https://doi.org/10.1007/s10791-024-09450-9. doi:10.1007/s10791-024-09450-9.

[13] C. Wang, M. Yuan, Y. Yang, K. Peng, H. Jiang, Revisiting long- and short-term preference learning for next POI recommendation with hierarchical LSTM, IEEE Transactions on Mobile Computing 23 (2024) 12693–12705. doi:10.1109/TMC.2024.3417405.

[14] M. A. Islam, M. M. Mohammad, S. S. Sarathi Das, M. E. Ali, A survey on deep learning based point-of-interest (POI) recommendations, Neurocomputing 472 (2022) 306–325. URL: https://www.sciencedirect.com/science/article/pii/S0925231221016106. doi:https://doi.org/10.1016/j.neucom.2021.05.114.

[15] P. Sánchez, A. Bellogín, Point-of-interest recommender systems based on location-based social networks: A survey from an experimental perspective, ACM Comput. Surv. 54 (2022). URL: https://doi.org/10.1145/3510409. doi:10.1145/3510409.

[16] Z. Wang, W. Höpken, D. Jannach, A survey on point-of-interest recommendations leveraging heterogeneous data, Information Technology & Tourism 27 (2025) 29–73. URL: https://doi.org/10.1007/s40558-024-00301-3. doi:10.1007/s40558-024-00301-3.

[17] S. Geng, S. Liu, Z. Fu, Y. Ge, Y. Zhang, Recommendation as language processing (RLP): A unified pretrain, personalized prompt & predict paradigm (P5), in: Proceedings of the 16th ACM Conference on Recommender Systems, RecSys '22, Association for Computing Machinery, New York, NY, USA, 2022, pp. 299–315. URL: https://doi.org/10.1145/3523227.3546767. doi:10.1145/3523227.3546767.

[18] P. Li, M. de Rijke, H. Xue, S. Ao, Y. Song, F. D. Salim, 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 '24, Association for Computing Machinery, New York, NY, USA, 2024, pp. 1463–1472. URL: https://doi.org/10.1145/3626772.3657840. doi:10.1145/3626772.3657840.

[19] T. Wang, C. Wang, Embracing LLMs for point-of-interest recommendations, IEEE Intelligent Systems 39 (2024) 56–59. URL: https://doi.ieeecomputersociety.org/10.1109/MIS.2023.3343489. doi:10.1109/MIS.2023.3343489.

[20] J. Long, L. Qu, G. Ye, T. Chen, Q. V. H. Nguyen, H. Yin, Unleashing the power of large language models for group POI recommendations, CoRR abs/2411.13415 (2024). URL: https://doi.org/10.48550/arXiv.2411.13415.

[21] S. Wang, B. Xie, L. Ding, X. Gao, J. Chen, Y. Xiang, SeCor: Aligning semantic and collaborative representations by large language models for next-point-of-interest recommendations, in: Proceedings of the 18th ACM Conference on Recommender Systems, RecSys '24, Association for Computing Machinery, New York, NY, USA, 2024, pp. 1–11. URL: https://doi.org/10.1145/3640457.3688124. doi:10.1145/3640457.3688124.

[22] J. Chen, H. Wang, J. Shang, GeoMamba: Toward efficient geography-aware sequential POI recommendation, IEEE Access 12 (2024) 167906–167918. doi:10.1109/ACCESS.2024.3482116.

[23] J. Zhai, X. Zheng, C.-D. Wang, H. Li, Y. Tian, Knowledge prompt-tuning for sequential recommendation, in: Proceedings of the 31st ACM International Conference on Multimedia, MM '23, Association for Computing Machinery, New York, NY, USA, 2023, pp. 6451–6461. URL: https://doi.org/10.1145/3581783.3612252. doi:10.1145/3581783.3612252.

[24] Z. Zhao, W. Fan, J. Li, Y. Liu, X. Mei, Y. Wang, Z. Wen, F. Wang, X. Zhao, J. Tang, Q. Li, Recommender systems in the era of large language models (LLMs), IEEE Transactions on Knowledge and Data Engineering 36 (2024) 6889–6907. doi:10.1109/TKDE.2024.3392335.

[25] G. Agrawal, T. Kumarage, Z. Alghamdi, H. Liu, Can knowledge graphs reduce hallucinations in LLMs? : A survey, in: K. Duh, H. Gomez, S. Bethard (Eds.), Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers), Association for Computational Linguistics, Mexico City, Mexico, 2024, pp. 3947–3960. URL: https://aclanthology.org/2024.naacl-long.219/. doi:10.18653/v1/2024.naacl-long.219.

[26] C. Xu, X. Mei, D. Liu, K. Zhao, A. S. Ding, An efficient privacy-preserving point-of-interest recommendation model based on local differential privacy, Complex & Intelligent Systems 9 (2023) 3277–3300. URL: https://doi.org/10.1007/s40747-022-00917-0. doi:10.1007/s40747-022-00917-0.

[27] Y. Tachioka, User knowledge prompt for sequential recommendation, in: Proceedings of the 18th ACM Conference on Recommender Systems, RecSys '24, Association for Computing Machinery, New York, NY, USA, 2024, pp. 1142–1146. URL: https://doi.org/10.1145/3640457.3691714. doi:10.1145/3640457.3691714.

[28] L. Li, Y. Zhang, L. Chen, Prompt distillation for efficient LLM-based recommendation, in: Proceedings of the 32nd ACM International Conference on Information and Knowledge Management, CIKM '23, Association for Computing Machinery, New York, NY, USA, 2023, pp. 1348–1357. URL: https://doi.org/10.1145/3583780.3615017. doi:10.1145/3583780.3615017.

[29] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, P. J. Liu, Exploring the limits of transfer learning with a unified text-to-text transformer, J. Mach. Learn. Res. 21 (2020).

[30] H.-C. Lai, Y.-S. Lu, M.-F. Wang, Y.-C. Chen, W.-Y. Shih, J.-L. Huang, SPENT+: A category- and region-aware successive POI recommendation model, in: 2021 22nd Asia-Pacific Network Operations and Management Symposium (APNOMS), 2021, pp. 230–233. doi:10.23919/APNOMS52696.2021.9562645.

[31] F. Yu, L. Cui, W. Guo, X. Lu, Q. Li, H. Lu, A category-aware deep model for successive POI recommendation on sparse check-in data, in: Proceedings of The Web Conference 2020, WWW '20, Association for Computing Machinery, New York, NY, USA, 2020, pp. 1264–1274. URL: https://doi.org/10.1145/3366423.3380202. doi:10.1145/3366423.3380202.

[32] X. Wang, Y. Liu, X. Zhou, Z. Leng, X. Wang, Next POI recommendation method based on category preference and attention mechanism in LBSNs, in: Web and Big Data: 6th International Joint Conference, APWeb-WAIM 2022, Nanjing, China, November 25–27, 2022, Proceedings, Part II, Springer-Verlag, Berlin, Heidelberg, 2022, pp. 12–19. URL: https://doi.org/10.1007/978-3-031-25198-6_2. doi:10.1007/978-3-031-25198-6_2.

[33] C. Song, J. Wen, S. Li, Personalized POI recommendation based on check-in data and geographical-regional influence, in: Proceedings of the 3rd International Conference on Machine Learning and Soft Computing, ICMLSC '19, Association for Computing Machinery, New York, NY, USA, 2019, pp. 128–133. URL: https://doi.org/10.1145/3310986.3311034. doi:10.1145/3310986.3311034.

[34] J. Tang, J. Jin, Z. Miao, B. Zhang, Q. An, J. Zhang, Region-aware POI recommendation with semantic spatial graph, in: 2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD), 2021, pp. 214–219. doi:10.1109/CSCWD49262.2021.9437810.

[35] X. Song, Z. Liu, L. Meng, D. Chu, J. Yu, Q. Z. Sheng, Accurate POI recommendation for random groups with improved graph neural networks and a multi-negotiation model, Scientific Reports 15 (2025) 7531. URL: https://doi.org/10.1038/s41598-025-91805-3. doi:10.1038/s41598-025-91805-3.

[36] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, E. Lefebvre, Fast unfolding of communities in large networks, Journal of Statistical Mechanics: Theory and Experiment 2008 (2008) P10008. URL: https://dx.doi.org/10.1088/1742-5468/2008/10/P10008. doi:10.1088/1742-5468/2008/10/P10008.

[37] H. Cai, V. W. Zheng, K. C.-C. Chang, A comprehensive survey of graph embedding: Problems, techniques, and applications, IEEE Transactions on Knowledge and Data Engineering 30 (2018) 1616–1637. doi:10.1109/TKDE.2018.2807452.

[38] A. Grover, J. Leskovec, node2vec: Scalable feature learning for networks, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16, Association for Computing Machinery, New York, NY, USA, 2016, pp. 855–864. URL: https://doi.org/10.1145/2939672.2939754. doi:10.1145/2939672.2939754.

[39] D. Arthur, S. Vassilvitskii, k-means++: the advantages of careful seeding, in: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA '07, Society for Industrial and Applied Mathematics, USA, 2007, pp. 1027–1035.

[40] Z. Zhang, B. Wang, Prompt learning for news recommendation, in: Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '23, Association for Computing Machinery, New York, NY, USA, 2023, pp. 227–237. URL: https://doi.org/10.1145/3539618.3591752. doi:10.1145/3539618.3591752.

[41] Y. Li, T. Chen, Y. Luo, H. Yin, Z. Huang, Discovering collaborative signals for next POI recommendation with iterative seq2graph augmentation, in: Z.-H. Zhou (Ed.), Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21, International Joint Conferences on Artificial Intelligence Organization, 2021, pp. 1491–1497. URL: https://doi.org/10.24963/ijcai.2021/206. doi:10.24963/ijcai.2021/206, main Track.

[42] X. Wang, F. Fukumoto, J. Cui, Y. Suzuki, J. Li, D. Yu, EEDN: Enhanced encoder-decoder network with local and global context learning for POI recommendation, in: Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '23, Association for Computing Machinery, New York, NY, USA, 2023, pp. 383–392. URL: https://doi.org/10.1145/3539618.3591678. doi:10.1145/3539618.3591678.

[43] D. Yang, D. Zhang, V. W. Zheng, Z. Yu, Modeling user activity preference by leveraging user spatial temporal characteristics in LBSNs, IEEE Transactions on Systems, Man, and Cybernetics: Systems 45 (2015) 129–142. doi:10.1109/TSMC.2014.2327053.

[44] D. Yang, D. Zhang, B. Qu, Participatory cultural mapping based on collective behavior data in location-based social networks, ACM Trans. Intell. Syst. Technol. 7 (2016). URL: https://doi.org/10.1145/2814575. doi:10.1145/2814575.

[45] D. Yang, B. Qu, J. Yang, P. Cudre-Mauroux, Revisiting user mobility and social relationships in LBSNs: A hypergraph embedding approach, in: The World Wide Web Conference, WWW '19, Association for Computing Machinery, New York, NY, USA, 2019, pp. 2147–2157. URL: https://doi.org/10.1145/3308558.3313635. doi:10.1145/3308558.3313635.

[46] E. Cho, S. A. Myers, J. Leskovec, Friendship and mobility: User movement in location-based social networks, in: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '11, Association for Computing Machinery, New York, NY, USA, 2011, pp. 1082–1090. URL: https://doi.org/10.1145/2020408.2020579. doi:10.1145/2020408.2020579.