From Dialogue to Destination: Geography-Aware Large Language Models with Multimodal Fusion for Conversational Recommendation

Yeming Li[1*], Chenxi Liu[2*], Jie Zou[1†], Cheng Long[3], Chaoning Zhang[1], Peng Wang[1], Yang Yang[1],

[1]University of Electronic Science and Technology of China, China
[2]Centre for Artificial Intelligence and Robotics, HKISI-CAS, Hong Kong, China
[3] College of Computing and Data Science, Nanyang Technological University, Singapore
202421080436@std.uestc.edu.cn, chenxi.liu@cair-cas.org.hk, jie.zou@uestc.edu.cn, c.long@ntu.edu.sg,
chaoningzhang1990@gmail.com, p.wang6@hotmail.com, yang.yang@uestc.edu.cn

Abstract

Conversational Recommender Systems (CRS) aim to provide personalized recommendations by interacting with users through natural language dialogue. However, in scenarios requiring deep geospatial awareness, existing methods, including those based on Large Language Models (LLMs), still face significant challenges in effectively fusing heterogeneous, multimodal geographic information with dynamic dialogue context. Simple fusion strategies struggle to resolve the asymmetric dependencies between dynamic user intent and static geographic context and fail to bridge the semantic gap between LLMs and structured geospatial data. To address these issues, we propose a framework for geography-aware CRS, named GeoCRS. Our core idea is to empower a frozen LLM with powerful geospatial reasoning capabilities by conditioning it on a dynamic, multimodal guidance signal generated by an external fusion architecture, all without altering the LLM's internal parameters. Specifically, we first design a hierarchical geographical encoder to uniformly represent heterogeneous geographic data. Subsequently, we introduce a contextual feature modulation module that asymmetrically injects the geographic context into the user's dialogue intent via a novel modulation mechanism to improve conversational recommendation via both geographic and dialogue context. Extensive experiments on public benchmark datasets demonstrate that our proposed GeoCRS significantly outperforms state-of-the-art baselines on the geography-aware conversational recommendation task.

Code — https://github.com/tsukikokyu/geocrs

Introduction

Conversational Recommender Systems (CRSs) (Li et al. 2018; Sun and Zhang 2018) have become an essential paradigm for advancing personalized recommendation by dynamically capturing user preferences through multi-turn natural language interactions. This interactive paradigm is applied in domains such as urban tourism and local service discovery (Xia et al. 2023), where effective recommendations are deeply intertwined with a user's geographical environment and movement patterns (Levandoski et al. 2012).

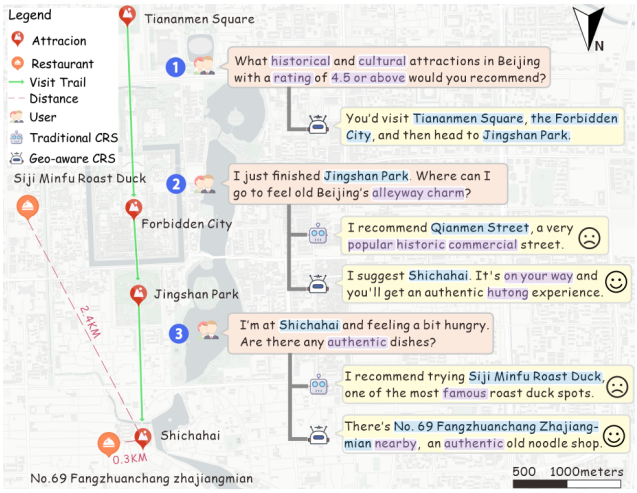


Figure 1: Traditional CRS offer suboptimal advice by ignoring location. Geo-aware CRS uses the tourist's current context and travel path to provide more relevant and timely recommendations.

In these scenarios,effective CRSs should not only understand evolving user intentions within a dialogue but also possess a deep insight into the complex spatio-temporal context. These unique requirements have given rise to the challenging subfield of geography-aware conversational recommendation (Liu et al. 2025c; Chen et al. 2025).

The Beijing travel scenario in Figure 1 illustrates the key challenges of geography-aware conversational recommendation. A user's needs combine multi-dimensional constraints: they involve explicit semantic requirements (e.g., "historical and cultural") and numerical constraints (e.g., rating ≥ 4.5), while also being inherently subject to the implicit geospatial constraints of their current location. After visiting Jingshan Park, the user wants to find a place with "alleyway charm". A traditional CRS, relying on global popularity, might recommend the distant Qianmen Street, which is in the opposite direction of the user's travel path. In contrast, a geography-aware CRS performs macro-level trajectory modeling to understand the user's path, recommending

*Equally contribution.
†Corresponding author.

the path-aligned and suitable Shichahai. Later, when the user gets hungry at Shichahai, the geography-aware CRS weighs trade-offs on a finer scale, performing micro-level proximity awareness to recommend a noodle shop just 300 meters away, instead of a more popular but distant roast duck restaurant. Therefore, an effective geography-aware CRS must not only process semantic numerical attributes reflect user intent but also perform complex and multi-level spatial reasoning.

Despite the necessity to model this complex interplay, a range of CRSs, from early knowledge-graph-based methods (Zhou et al. 2020a; Chen et al. 2019) to recent LLM-based models (Wang et al. 2022; Dao et al. 2024; Li et al. 2024; An et al. 2025), tend to either ignore geographical data or treat locations as simple text entities.

Similarly, mainstream multi-modal recommendation models are ill-suited for this task (Ye et al. 2025; Liu et al. 2024c). They are primarily designed to handle *intrinsic* item attributes like product images, not the *extrinsic* and dynamic spatial context (e.g., GPS history, relative positioning) that our problem demands, thus failing to achieve a deep multi-modal fusion of geospatial information and dialogue context (Xu et al. 2025).

Recently, the widespread success of LLMs has highlighted their substantial potential to enhance recommender systems. However, two challenges remain in effectively leveraging them for fusing dynamic dialogue context with geospatial information: **(1)** The first challenge is how to construct geospatial representations that are comprehensible to LLMs. Geographical data is inherently heterogeneous, comprising unstructured text, structured attributes, and unique spatial coordinates. As different modalities often vary in structure and information granularity, transforming these diverse data types into a coherent, information-rich representation that an LLM can understand is a critical first step (Xu et al. 2025). **(2)** The second challenge is how to leverage geospatial information without interfering with the user's primary intent, thereby delivering more precise recommendations. In decision-making, the user's dynamic intent is the primary signal, while the vast geographical history should act as a modulating background. If fused symmetrically (e.g., via simple concatenation), this key intent signal can be easily submerged by the background context, where dominant modalities can overshadow weaker ones (Gao et al. 2025). Therefore, designing an asymmetric fusion mechanism, which uses geographic context to *modulate* rather than interfere with user intent, is fundamental for accurate, geography-aware recommendations (Gao et al. 2025).

To address the aforementioned challenges, we propose GeoCRS. For the first challenge, we designed the hierarchical geographical encoder, efficiently fusing multi-source geographical data into a unified, LLM-comprehensible representation via parallel and hierarchical encoding. For the second challenge, we designed the contextual feature modulation module, which abandons traditional symmetric approaches and employs a mechanism combining attention and affine transformation to use the geographical information as a dynamic condition to modulate and enhance the user's dialogue intent. This entire process ultimately generates a guidance signal deeply fused with geographical information to direct the LLM's decision-making for recommendation.

The main contributions of this paper are as follows:

- We propose the **GeoCRS framework**, a new paradigm for geography-aware conversational recommendation that integrates LLM with external, trainable modules to fuse geospatial information and dialogue context, effectively bridging a gap in current research.
- We design a **hierarchical geographical encoder**, which efficiently creates a unified and comprehensive representation from heterogeneous geospatial data.
- We introduce a **contextual feature modulation** module, which actualizes an asymmetric mechanism where geographic context modulates user intent and prevents the primary signal from being submerged.
- Through extensive experiments on public benchmarks and our augmented CrossWOZ dataset (augmented with geographical data and which we will release publicly), we demonstrate that GeoCRS achieves state-of-the-art performance, which in turn validates the effectiveness of incorporating geospatial information into CRS.

Related Work

Conversational Recommendation

Early CRSs often utilized Knowledge Graphs (KGs) to enrich semantic context and combat data sparsity in general-domain recommendations (Christakopoulou, Radlinski, and Hofmann 2016; Chen et al. 2019; Zhou et al. 2020a; Zhang et al. 2023; Zou et al. 2024). This paradigm was subsequently adapted for geography-aware scenarios, where methods focused on building domain-specific KGs and mitigating representation noise to handle spatio-temporal complexities (Yuan et al. 2024; Liu et al. 2025c; Chen et al. 2025). The advent of LLMs has shifted the paradigm towards unified, end-to-end architectures that harness their inherent knowledge and reasoning abilities (Zhang et al. 2025; Zou et al. 2026). Many approaches reformulate the recommendation as a language generation task, using techniques like prompt learning or in-context learning to guide the LLM (Wang et al. 2022; Dao et al. 2024). Others focus on proactively asking clarifying questions (Zou et al. 2022a; Zou and Kanoulas 2020; Ma et al. 2024) to resolve ambiguities and elicit more specific user preferences (Zou, Chen, and Kanoulas 2020; Zou et al. 2022b, 2025). In the context of geography-aware recommendation, LLMs have been prompted with user trajectory data to tackle cold-start issues (Li et al. 2024). Nevertheless, these approaches struggle to effectively synthesize user intent from dynamic dialogues with spatio-temporal data and the reasoning capabilities of LLMs, which consequently limits their precision for geography-aware conversational recommendations (Liu et al. 2024b, 2025b,a).

LLM-based Multimodal Fusion for Recommendation

For multimodal recommendation, recent work with LLMs follows two primary strategies. One is modality conversion, where non-textual data like user behaviors are translated

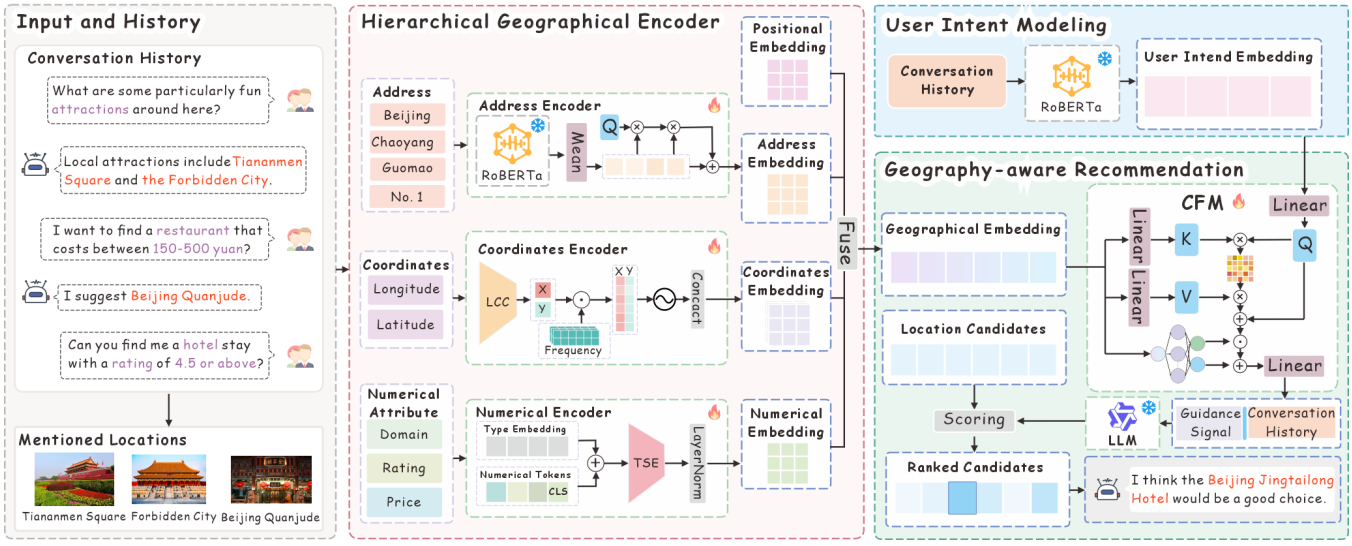


Figure 2: Overall Framework of GeoCRS. The framework processes Input and History in two streams: the Hierarchical Geographical Encoder creates a geographic representation, while User Intent Modeling encodes dialogue. The Geography-aware Recommendation module asymmetrically fuses both into a signal, steering a frozen LLM to rank candidates.

into textual descriptions for LLM comprehension (Ye et al. 2025). The other strategy is modality fusion, which integrates different data types at a semantic level (Wei et al. 2025; Wang et al. 2017, 2024; Gao et al. 2017). This includes using attention mechanisms to fuse collaborative filtering (CF) signals with text (Liu et al. 2024c), mapping CF signals into specialized tokens, or even employing the LLM to auto-construct recommendation graphs from multimodal inputs (Shan et al. 2024). While powerful, these multimodal methods are primarily designed for domains like product or movie recommendation, and their core architectures are not equipped to handle the dynamic dialogue interactions and complex spatio-temporal contexts essential for this task.

Preliminary

We first formalize the key definitions for geography-aware conversational recommendation as follows:

- **User Location**: Let $\mathcal{P} = \{p_1, p_2, \ldots, p_K\}$ denotes the universal set of locations, where K is the total number of locations. Each location $p_i \in \mathcal{P}$ is associated with a set of multimodal features, represented as $\mathbf{x}_i^m \in \mathbb{R}^{d_m}$. The modalities m include address text (a), numerical features (n), and geographical coordinates (l).
- **Conversational Session**: Let $\mathcal{S} = \{s_1, s_2, \ldots, s_M\}$ represents the set of conversational sessions, where M is the total number of sessions. Each session $s \in \mathcal{S}$ consists of a sequence of chronological utterances, denoted as $s = (c_1, c_2, \ldots, c_{T_s})$, where c_t is the utterance at turn t and T_s is the total number of turns.
- **User Interaction History**: For a user u, the interaction history $\mathcal{H}_{u,t}$ is dynamically constructed from the ongoing dialogue, evolving as a chronologically ordered sequence of visited locations $(p_{h_1}, p_{h_2}, \ldots, p_{h_k})$ up to turn t, with each $p_{h_i} \in \mathcal{P}$ inferred from the dialogue context.

Problem Definition: The objective of a geography-aware CRS is to dynamically infer user preferences by modeling the interplay between their natural language preferences, captured in dialogue, and their spatio-temporal behavior patterns, reflected in their interaction history. Formally, given the dialogue history $s_t = (c_1, c_2, \ldots, c_t)$ and the interaction history $\mathcal{H}_{u,t} = (p_{h_1}, \ldots, p_{h_k})$ for a user u up to turn t, along with the complete set of multimodal features for all locations, $\{\mathbf{x}_i^m | p_i \in \mathcal{P}\}$, the system learns a mapping function f to generate a ranked subset of candidate locations $\mathcal{R}_t$:

$$\mathcal{R}_t = f(s_t, \mathcal{H}_{u,t}, \{\mathbf{x}_i^m | p_i \in \mathcal{P}\}), \quad (1)$$

where $\mathcal{R}_t$ maximizes alignment with the user's inferred preferences.

Methodology

As illustrated in Figure 2, our framework first uses a hierarchical geographical encoder to create geographical representations from the user's interaction history , while a user intent modeling module produces the user intent representation from the conversation history. Then, the geography-aware recommendation module employs contextual feature modulation to fuse these representations, generating a guidance signal. This signal is prepended to the conversation history to guide the LLM , which finally scores candidate locations to produce the final ranking.

Hierarchical Geographical Encoding

To address LLMs' struggle with interpreting heterogeneous geographic data (Liang et al. 2025), we introduce the hierarchical geographical encoder. This module processes each location from the user's interaction history, encoding its multimodal features (textual address, coordinates, numerical attributes) in parallel and integrating them into the final geo-

contextual representation matrix, $\mathbf{E}_{\text{geo}}$. The following subsections detail these encoding and fusion strategies.

Address Encoding The textual address of a geographic location, denoted by $\mathbf{x}_i^a$, is not a flat string but is inherently structured. For example, an address such as *"No. 4, Jingshan Front Street, Dongcheng District, Beijing"* possesses a hierarchical structure, spanning from a macro-level (e.g., the city) to a micro-level (e.g., the street and number). Encoding the entire address as a single unit inevitably discards this critical multi-scale spatial structure, limiting the model's ability to perform fine-grained geographic reasoning.

To address this challenge, we propose a hierarchical address encoding mechanism. The process begins by segmenting the address string $\mathbf{x}_i^a$ into a sequence of U spatial units, $\{u_1, u_2, \ldots, u_U\}$. Each unit u_j is passed through a text encoder f_{RoBERTa} (Liu et al. 2019; Cui et al. 2019) and mean-pooled to derive a fixed-size vector $\mathbf{h}_j$:

$$\mathbf{h}_j = \mathit{MeanPool}(f_{\text{RoBERTa}}(u_j)). \tag{2}$$

We then employ an attention mechanism to aggregate the resulting sequence of unit embeddings $\mathbf{H} = (\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_U)$. We introduce a learnable query vector, $\mathbf{q} \in \mathbb{R}^{d_{\text{model}}}$, shared across all addresses and optimized during training.

The attention weight α_j for each unit embedding $\mathbf{h}_j$ is computed as:

$$\alpha_j = \frac{\exp(\mathbf{q}\mathbf{h}_j^\top)}{\sum_{k=1}^{U} \exp(\mathbf{q}\mathbf{h}_k^\top)}. \tag{3}$$

The final, unified address representation, $\mathbf{e}_{\text{address},i}$, is then produced by the weighted sum of all unit embeddings:

$$\mathbf{e}_{\text{address},i} = \sum_{j=1}^{U} \alpha_j \mathbf{h}_j. \tag{4}$$

Coordinates Encoding Raw latitude and longitude coordinates are ill-suited for LLMs because their implicit, nonlinear spatial relationships hinder geographic reasoning. We bridge this gap by encoding each coordinate pair into a rich embedding that makes multi-scale spatial relations explicit, enabling the LLM to interpret geographic positions.

(1) Local Planar Projection. Since spherical coordinates are suboptimal for city-level modeling , we apply the Lambert Conformal Conic (LCC) projection. LCC transforms WGS-84 geodetic coordinates into a 2D Cartesian plane, preserving angles and shapes well for urban-scale modeling.

For each location p_i, the LCC projection transforms its geodetic coordinates, longitude λ_i and latitude ϕ_i, into planar coordinates (x_i, y_i). The projection is defined by standard parallels ϕ_1, ϕ_2, a central meridian λ_0, and a reference latitude ϕ_0. The transformation involves the following steps:

First, compute the projection constants n and F:

$$n = \frac{\ln(\cos\phi_1 \sec\phi_2)}{\ln\left(\tan\left(\frac{\pi}{4} + \frac{\phi_2}{2}\right)\cot\left(\frac{\pi}{4} + \frac{\phi_1}{2}\right)\right)}, \tag{5}$$

$$F = \frac{\cos\phi_1 \cdot \tan^n\left(\frac{\pi}{4} + \frac{\phi_1}{2}\right)}{n}. \tag{6}$$

Then compute the polar coordinates (ρ_i, θ_i):

$$\rho_i = F \cdot \cot^n\left(\frac{\pi}{4} + \frac{\phi_i}{2}\right), \quad \theta_i = n(\lambda_i - \lambda_0). \tag{7}$$

Finally, convert (ρ_i, θ_i) to Cartesian coordinates (x_i, y_i):

$$x_i = \rho_i \sin\theta_i, \quad y_i = \rho_0 - \rho_i \cos\theta_i, \tag{8}$$

where ρ_0 is the radius at reference latitude ϕ_0.

(2) Multi-Scale Sinusoidal Encoding. Human spatial perception naturally operates across multiple scales, ranging from walkable neighbors to broader administrative regions. To mimic this property, we propose a multi-scale encoding scheme. The final representation for a location i, $\mathbf{e}_{\text{location},i}$, is formed by concatenating embeddings derived from C distinct spatial scales:

$$\mathbf{e}_{\text{location},i} = [\mathbf{e}_{i,1}; \mathbf{e}_{i,2}; \ldots; \mathbf{e}_{i,C}]. \tag{9}$$

Each scale-specific embedding $\mathbf{e}_{i,c}$ in the sequence is itself a concatenation of the individual encodings for the planar coordinates (x_i, y_i), which are generated by a sinusoidal function $\psi(\cdot)$:

$$\mathbf{e}_{i,c} = [\psi(x_i); \psi(y_i)]. \tag{10}$$

The function $\psi(\cdot)$ maps any scalar value v to a $2D$-dimensional vector using a set of D scale-dependent angular frequency pairs $\{\omega_{c,k}\}$:

$$\begin{aligned}\psi(v) = [&\sin(\omega_{c,1}v), \cos(\omega_{c,1}v), \ldots, \\ &\sin(\omega_{c,D}v), \cos(\omega_{c,D}v)]^\top,\end{aligned} \tag{11}$$

where the frequencies $\{\omega_{c,k}\}$ are defined based on a predefined maximum radius $r_{\text{max},c}$ for each scale c as follows:

$$\omega_{c,k} = (r_{\text{max},c})^{-k/D}. \tag{12}$$

This approach, inspired by the positional encoding in Transformers, captures spatial relationships from local to regional patterns, enhancing the model's capacity for fine-grained geographic reasoning.

Numerical Attribute Encoding We use a Transformer-based numerical encoder (Liu et al. 2024a) to encode key attributes $x_i^n = \{a_1, \ldots, a_{N_{\text{attr}}}\}$ into a unified representation $\mathbf{e}_{\text{numeric},i}$. First, we convert each attribute a_j into an input embedding by summing its value embedding $\mathbf{v}_j$ and a learnable type embedding $\mathbf{t}_j$. Then, a learnable $\mathbf{e}_{\text{CLS}}$ token is prepended to the sequence to aggregate attribute information (Devlin et al. 2019).

Thus, the initial input matrix $H^{(0)}$ for the Transformer encoder (TSE) is constructed as follows:

$$H^{(0)} = [\mathbf{e}_{\text{CLS}} + \mathbf{t}_0; \mathbf{v}_1 + \mathbf{t}_1; \ldots; \mathbf{v}_{N_{\text{attr}}} + \mathbf{t}_{N_{\text{attr}}}]. \tag{13}$$

To capture the deep, cross-attribute dependencies, we feed the initial embeddings $H^{(0)}$ into a standard L-layer Transformer Encoder (TSE) to obtain the final hidden states $H^{(L)}$:

$$H^{(L)} = \mathit{TransformerEncoder}(H^{(0)}). \tag{14}$$

Further, we extract the output embedding corresponding to the "[CLS]" token as $\mathbf{e}_{\text{numeric},i}$.

Geographic Information Fusion To capture the deep, cross-modal interactions between the address, coordinate, and numerical attribute representations ($\mathbf{e}_{\text{address},i}$, $\mathbf{e}_{\text{location},i}$, and $\mathbf{e}_{\text{numeric},i}$), we fuse them using a Multi-layer Perceptron (MLP). The three modality-specific embeddings are first concatenated and then fed into the MLP, f_{MLP}, to generate the final unified geo-contextual representation, $\mathbf{e}_{\text{geo},i}$.

$$\mathbf{e}_{\text{concat},i} = [\mathbf{e}_{\text{address},i}; \mathbf{e}_{\text{location},i}; \mathbf{e}_{\text{numeric},i}], \tag{15}$$

$$\mathbf{e}_{\text{geo},i} = f_{\text{MLP}}(\mathbf{e}_{\text{concat},i}). \tag{16}$$

By applying this fusion process to all locations $(p_{h_1}, \ldots, p_{h_k})$ in the user's interaction history $\mathcal{H}_{u,t}$, we construct the final geography-contextual matrix $\mathbf{E}_{\text{geo}}$:

$$\mathbf{E}_{\text{geo}} = (\mathbf{e}_{\text{geo},h_1}, \mathbf{e}_{\text{geo},h_2}, \ldots, \mathbf{e}_{\text{geo},h_k}). \tag{17}$$

User Intent Modeling

To capture evolving user preferences in a dialogue session s , we use a frozen, pre-trained RoBERTa model as the text encoder f_{RoBERTa} to derive high-level semantic representations from the dialogue history. Given $s = \{c_1, \ldots, c_{T_s}\}$, we obtain the user intent embedding $\mathbf{E}_{\text{intent}}$:

$$\mathbf{E}_{\text{intent}} = f_{\text{RoBERTa}}(c_1, c_2, \ldots, c_{T_s}), \tag{18}$$

where $\mathbf{E}_{\text{intent}} \in \mathbb{R}^{L \times d_{\text{model}}}$ denotes the contextual representation of user intent, L is the tokenized sequence length, and d_{model} is the hidden size of the encoder.

Geography-aware Recommendation

Given the user intent $\mathbf{E}_{\text{intent}}$ and geographic context $\mathbf{E}_{\text{geo}}$, our contextual feature modulation module fuses them into a guidance signal $\mathbf{E}_{\text{cond}}$. This signal serves as a soft prompt for an LLM to produce the final user representation $\mathbf{u}_s$.

Contextual Feature Modulation To effectively guide user intent through geographic context, we propose the contextual feature modulation (CFM) module. It jointly performs two complementary subtasks, Content Alignment and Contextual Modulation, within a unified architecture to decouple and enhance the cross-modal fusion process. The core function of contextual feature modulation is to treat the geographic context embedding $\mathbf{E}_{\text{geo}}$ as a dynamic condition that modulates the user intent embedding $\mathbf{E}_{\text{intent}}$, yielding the final context-aware representation $\mathbf{E}_{\text{cond}}$:

$$\mathbf{E}_{\text{cond}} = CFM(\mathbf{E}_{\text{intent}}, \mathbf{E}_{\text{geo}}). \tag{19}$$

(1) Content Alignment. The first subtask, Content Alignment, focuses on answering the question: *"Which parts of the geographic context are most relevant to the current user intent?"* To this end, we employ a cross-attention mechanism, where we treat the dialogue intent ($\mathbf{E}_{\text{intent}}$) as the query that actively "interrogates" the user's geographic history ($\mathbf{E}_{\text{geo}}$), which serves as the key-value memory bank. This process yields an aligned representation $\mathbf{E}_{\text{aligned}}$ where the most pertinent historical locations are emphasized:

$$\begin{aligned} \mathbf{A} &= Attention(\mathbf{Q} = \mathbf{E}_{\text{intent}}, \mathbf{K} = \mathbf{E}_{\text{geo}}, \mathbf{V} = \mathbf{E}_{\text{geo}}) \\ \mathbf{E}_{\text{aligned}} &= LayerNorm(\mathbf{E}_{\text{intent}} + \mathbf{A}). \end{aligned} \tag{20}$$

(2) Contextual Modulation. The aligned representation $\mathbf{E}_{\text{aligned}}$ captures relevant historical context, it has not yet been conditioned on the geographic background. The Contextual Modulation achieves this by adopting Feature-wise Linear Modulation (FiLM) (Perez et al. 2018) as the core operator. Specifically, we first aggregate the geographic context sequence $\mathbf{E}_{\text{geo}}$ by taking its first hidden state, and then a learnable projection network f_{proj} transforms this aggregated vector into a scaling vector $\boldsymbol{\gamma}$ and a bias vector $\boldsymbol{\beta}$:

$$[\boldsymbol{\gamma},\ \boldsymbol{\beta}] = f_{\text{proj}}(Aggregate(\mathbf{E}_{\text{geo}})), \tag{21}$$

where $\boldsymbol{\gamma}, \boldsymbol{\beta} \in \mathbb{R}^{d_{\text{model}}}$.These parameters are then applied to the aligned features $\mathbf{E}_{\text{aligned}}$ to generate the final conditioned representation $\mathbf{E}_{\text{cond}}$, which we term the guidance signal. Note that both $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ are single vectors that are broadcast across the sequence dimension of $\mathbf{E}_{\text{aligned}}$ to perform the element-wise modulation:

$$\mathbf{E}_{\text{cond}} = \boldsymbol{\gamma} \odot \mathbf{E}_{\text{aligned}} + \boldsymbol{\beta}. \tag{22}$$

In this affine transformation, $\boldsymbol{\gamma}$ modulates feature importance, while $\boldsymbol{\beta}$ shifts feature semantics, enabling concise, context-aware preference modeling without redundancy.

LLM-based Recommendation Our approach revolves around conditioning an LLM with dynamically generated multimodal representations to produce geography-aware recommendations. The core of this conditioning process is our contextual feature modulation module, which produces a context-aware guidance signal $\mathbf{E}_{\text{cond}}$. This signal is prepended to the dialogue history embeddings, $\mathbf{E}_{\text{intent}}$, to form the final model input $\mathbf{R}_s$:

$$\mathbf{R}_s = [\mathbf{E}_{\text{cond}}; \mathbf{E}_{\text{intent}}]. \tag{23}$$

We input $\mathbf{R}_s$ into the LLM, extracting the last token's hidden state as the user representation vector, $\mathbf{u}s \in \mathbb{R}^{d_{\text{model}}}$:

$$\mathbf{u}_s = f_{\text{LLM}}(\mathbf{R}_s; \theta_{\text{fusion}})_{[-1]}, \tag{24}$$

where θ_{fusion} consists exclusively of the learnable parameters from our external fusion architecture, since the LLM is frozen.

The resulting user representation $\mathbf{u}_s$, which encapsulates all necessary dialogue and historical context, is used to compute a recommendation score for each location p via its dot product with the corresponding learnable embedding $\mathbf{l}_p$.

We optimize the learnable parameters θ_{fusion} using a two-stage training strategy consisting of a pre-training stage for learning with ground-truth followed by an adaptation stage for predicting masked targets. The entire architecture is trained end-to-end by minimizing the cross-entropy loss against the ground-truth target p_{target}. For each instance in the training set $\mathcal{S}_{\text{train}}$, the loss is defined as:

$$\mathcal{L}(\theta_{\text{fusion}}) = -\log \frac{\exp(\mathbf{u}_s \cdot \mathbf{l}_{p_{\text{target}}})}{\sum_{p' \in \mathcal{P}} \exp(\mathbf{u}_s \cdot \mathbf{l}_{p'})}, \tag{25}$$

where the softmax denominator is computed over the universal set of locations $\mathcal{P}$, which serves as the candidate set during training.

Model	CrossWOZ							MultiWOZ						
	Recall			NDCG		MRR		Recall			NDCG		MRR	
	@1	@5	@10	@5	@10	@5	@10	@1	@5	@10	@5	@10	@5	@10
Popularity	0.0385	0.1637	0.2438	0.1010	0.1266	0.0805	0.0909	0.0303	0.1085	0.1617	0.0677	0.0846	0.0546	0.0613
BERT	0.2164	0.4923	0.5906	0.3625	0.3946	0.3193	0.3327	0.2851	0.6217	0.7285	0.4641	0.4989	0.4116	0.4261
SASRec	0.1588	0.3289	0.4337	0.2452	0.2790	0.2178	0.2316	0.2755	0.3699	0.4042	0.3255	0.3366	0.3108	0.3153
KBRD	0.0789	0.2844	0.4052	0.1830	0.2218	0.1497	0.1655	0.2627	0.4824	0.5903	0.3763	0.4114	0.3413	0.3559
KGSF	0.0988	0.2746	0.3816	0.1880	0.2227	0.1596	0.1740	0.3818	0.7233	0.8127	0.5631	0.5921	0.5098	0.5218
TG-ReDial	0.2799	0.5872	0.6782	0.4431	0.4728	0.3952	0.4076	0.4069	0.7077	0.7907	0.5683	0.5952	0.5218	0.5329
ZSCRS	0.0792	0.2276	0.3172	0.1548	0.1839	0.1309	0.1430	0.1370	0.2505	0.2909	0.1996	0.2126	0.1825	0.1878
VRICR	0.1159	0.3261	0.4710	0.2226	0.2693	0.1887	0.2079	0.2678	0.4842	0.5615	0.3805	0.4056	0.3461	0.3566
UniCRS	0.2462	0.6478	0.7459	0.4580	0.4900	0.3950	0.4083	0.5677	0.8929	0.9550	0.7464	0.7668	0.6971	0.7058
DCRS	0.2615	0.6510	0.7435	0.4615	0.5033	0.4005	0.4198	0.5780	0.9012	0.9605	0.7575	0.7761	0.7093	0.7162
GeoCRS	**0.3384***	**0.6985***	**0.7792***	**0.5298***	**0.5561***	**0.4736***	**0.4846***	**0.5873***	**0.9212***	**0.9664***	**0.7703***	**0.7851***	**0.7196***	**0.7258***

Table 1: Recommendation performance comparison on CrossWOZ and MultiWOZ datasets. The symbol "*" refers to a significant improvement compared to the best baseline at the $p < 0.05$ level using the two-tailed pairwise t-test.

Experiments

Experimental Setup

Datasets We conduct experiments primarily on the CrossWOZ dataset (Zhu et al. 2020), which we augment with coordinates (via the Amap API) and parse all address texts. In addition, we use the MultiWOZ_2.2 dataset (Zang et al. 2020) to validate generalization.

Evaluation Metrics To evaluate the recommendation performance, we adopt three widely-used ranking metrics: **Recall@k**, **NDCG@k**, and **MRR@k** (k $\in \{1, 5, 10\}$), similar to Wang et al. (2022) and Wei et al. (2025). For all metrics, higher values indicate better performance.

Baselines To evaluate the effectiveness of our model, we select the following representative baselines for comparison: Popularity, BERT (Devlin et al. 2019), SASRec (Kang and McAuley 2018), KBRD (Chen et al. 2019), KGSF (Zhou et al. 2020a), TG-ReDial (Zhou et al. 2020b), UniCRS (Wang et al. 2022) ZSCRS (He et al. 2023), implemented using the DeepSeek-V3 model, VRICR(Zhang et al. 2023) and DCRS (Dao et al. 2024). Details of the baselines are provided in the *Appendix*.

Implementation Details The detailed configurations and hyperparameters are provided in the *Appendix*. The codes are provided in the supplementary materials.

Overall Performance Comparison

Experimental results in Table 1 demonstrate the superiority of our proposed GeoCRS model. It consistently outperforms all baselines across both datasets, achieving a significant 20.9% relative improvement in Recall@1 on CrossWOZ over the strongest baseline. This might be because traditional sequential and general models (e.g., Popularity, BERT, and SASRec) lack an understanding of dialogue context, while both knowledge graph-based (e.g., KBRD, KGSF, TG-ReDial, and VRICR) and general-purpose LLM methods (e.g., ZSCRS, UniCRS, and DCRS) fail to handle the dynamic, heterogeneous, and multimodal nature of geographical information. In contrast, our GeoCRS benefits from its two key innovations that enable genuine geography-aware reasoning to improve the recommendation performance: (1) the hierarchical geographical encoder, which unifies geographic data into a rich representation for the LLM; (2) the contextual feature modulation module, which precisely modulates the dynamic dialogue intent with the geographic context via an asymmetric fusion mechanism.

Model Variant	R@1	R@5	R@10
GeoCRS w/o Intent	0.3180	0.6331	0.7262
GeoCRS w/o Geo-Context	0.2546	0.6728	0.7606
GeoCRS w/o Pre-train	0.3041	0.6679	0.7539
GeoCRS w/o Guidance	0.0526	0.1833	0.2702
GeoCRS[Concat. Fusion]	0.3160	0.6810	0.7668
GeoCRS[Symm. Fusion]	0.2513	0.6448	0.7354
GeoCRS	**0.3384**	**0.6985**	**0.7792**

Table 2: Ablation study of key components within GeoCRS.

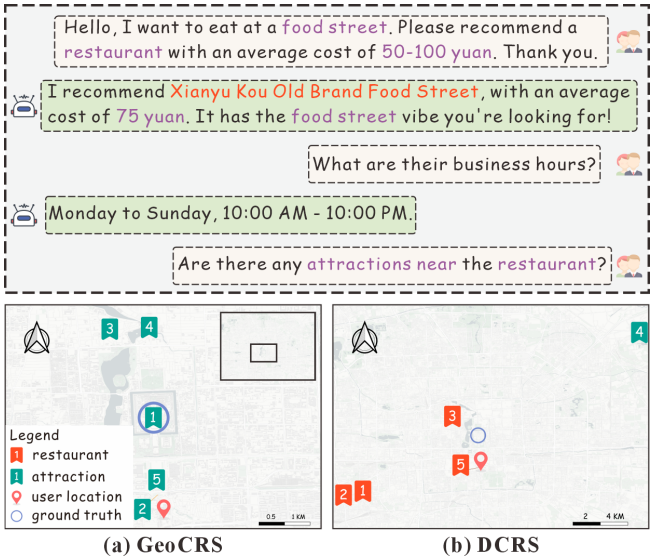
Model	R@1	R@5	R@10
UniCRS[GPT-2]	0.2462	0.6478	0.7459
DCRS[GPT-2]	0.2615	0.6510	0.7435
GeoCRS[GPT-2]	0.2796	0.6547	0.7529
GeoCRS[Qwen-1.7B]	**0.3384**	**0.6985**	**0.7792**

Table 3: Ablation study of the LLMs within GeoCRS

.

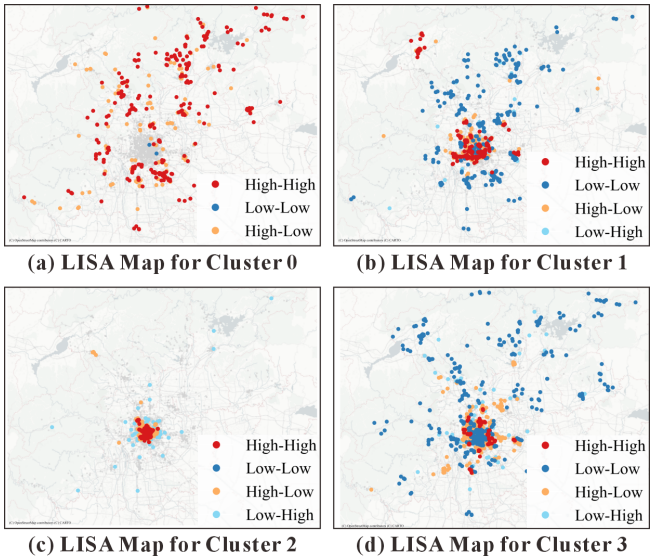
Ablation Studies

Analysis of Core Components. To quantify the contribution of each component of our GeoCRS, we conducted ablation studies on the CrossWOZ dataset for several model variants, with results in Table 2. First, removing either the dialogue intent (*GeoCRS w/o Intent*) or the whole geographic context, i.e., $\mathbf{E}_{\text{geo}}$ (*GeoCRS w/o Geo-Context*) leads to a significant performance degradation, confirming that both are



(a) GeoCRS (b) DCRS

Figure 3: A case study comparing the recommendation results of GeoCRS and DCRS. The numbers on the maps indicate the recommendation rank.



(a) LISA Map for Cluster 0 (b) LISA Map for Cluster 1

(c) LISA Map for Cluster 2 (d) LISA Map for Cluster 3

Figure 4: LISA visualization for four learned clusters: "High-High" (red) for hotspots of geographic concentration, "Low-Low" (blue) for cold spots of spatial absence, and "High-Low/Low-High" for areas of local contrast.

indispensable inputs. Similarly, replacing our asymmetric fusion module with simple concatenation (*GeoCRS[Concat. Fusion]*) or a symmetric mechanism (i.e., bidirectional attention fusion (Huang et al. 2024)) (*GeoCRS[Symm. Fusion]*) yields inferior performance, which demonstrates the superiority of our asymmetric fusion of GeoCRS. Furthermore, skipping the pre-training stage (*GeoCRS w/o Pretrain*) impairs the model's performance, highlighting the necessity of our two-stage training strategy. Finally, the model without any external guidance signal (*GeoCRS w/o Guidance*) exhibits the most substantial performance drop, decisively validating the effectiveness of our core idea to guide an LLM with external modules.

Analysis of different LLM backbones. In addition, we analyze the impact of the different LLM backbones, with results in Table 3. To isolate architectural benefits, we tested our framework with GPT2-Chinese (Du 2019), the same base model as UniCRS and DCRS. Our framework still outperforms both, proving the gain stems from our novel fusion architecture, not merely a stronger LLM. Furthermore, upgrading the base model to the more advanced Qwen3-1.7B (Yang et al. 2025) substantially improves performance, demonstrating our framework's scalability and ability to leverage more powerful language models.

Case Study

Figure 3 presents a case from the CrossWOZ test set where the user query "attractions near the restaurant" contains both attribute and spatial constraints. The baseline model fails on both counts, recommending distant restaurants instead. In contrast, GeoCRS correctly fulfills the request by generating a guidance signal that encodes both constraints and fuses user intent, directing the LLM's geography-aware reasoning.

To verify the source of GeoCRS's geographic awareness at a deeper level, we performed a spatial analysis on the learned geographic embeddings, $\mathbf{E}_{\text{geo}}$. We first used the K-Means algorithm to cluster the embeddings of all locations into four categories, and then a Global Moran's I analysis confirmed significant positive spatial autocorrelation for all four clusters, indicating a strong tendency for geographic clustering. A subsequent Local Indicators of Spatial Association (LISA) analysis further identified four distinct spatial patterns shown in figure 4. Notably, in the case of Figure 3 (a), all "attractions" recommended by GeoCRS belong exclusively to cluster 2, with LISA types consistently identified as "High-High" hotspots. This finding validates our core hypothesis: GeoCRS learns an embedding space where similarity reflects geographic proximity, enabling genuine spatial reasoning.

Conclusion

In this paper, we introduce GeoCRS, a framework that enables LLMs for geography-aware conversational recommendation. Our approach is distinguished by two key components within an external architecture: a hierarchical geographical encoder that translates geographic data into a unified, LLM-comprehensible representation, and a contextual feature modulation module that asymmetrically fuses this representation with the user's primary intent. Extensive experiments and ablation studies validate the effectiveness of our proposed GeoCRS. We acknowledge several limitations of this work, particularly the reliance on static location information. Therefore, future research could beneficially explore integrating streaming contextual data and incorporating richer modalities, such as location images or real-time event data.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (62402093), the Sichuan Science and Technology Program (2025ZNSFSC0479), the Fundamental Research Funds for the Central Universities (JBK202511020), and the InnoHK funding. This work was also supported in part by the National Natural Science Foundation of China under grants U20B2063 and 62220106008, the Sichuan Science and Technology Program under Grant 2024NSFTD0034.

References

An, G.; Zou, J.; Wei, J.; Zhang, C.; Sun, F.; and Yang, Y. 2025. Beyond whole dialogue modeling: Contextual disentanglement for conversational recommendation. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 31–41.

Chen, J.; Miao, H.; Qiu, D.; Guo, J.; Li, Y.; and Zhao, Y. 2025. Sustainability-Oriented Task Recommendation in Spatial Crowdsourcing. In *IEEE International Conference on Data Engineering*, 2712–2725.

Chen, Q.; Lin, J.; Zhang, Y.; Ding, M.; Cen, Y.; Yang, H.; and Tang, J. 2019. Towards knowledge-based recommender dialog system. *arXiv preprint arXiv:1908.05391*.

Christakopoulou, K.; Radlinski, F.; and Hofmann, K. 2016. Towards conversational recommender systems. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 815–824.

Cui, Y.; Che, W.; Liu, T.; Qin, B.; Yang, Z.; Wang, S.; and Hu, G. 2019. Pre-Training with Whole Word Masking for Chinese BERT. *arXiv preprint arXiv:1906.08101*.

Dao, H.; Deng, Y.; Le, D. D.; and Liao, L. 2024. Broadening the view: Demonstration-augmented prompt learning for conversational recommendation. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 785–795.

Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies*, 4171–4186.

Du, Z. 2019. GPT2-Chinese: Tools for training GPT2 model in Chinese language.

Gao, L.; Guo, Z.; Zhang, H.; Xu, X.; and Shen, H. T. 2017. Video captioning with attention-based LSTM and semantic consistency. *IEEE Transactions on Multimedia*, 2045–2055.

Gao, X.; Cao, B.; Zhu, P.; Wang, N.; and Hu, Q. 2025. Asymmetric reinforcing against multi-modal representation bias. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 16754–16762.

He, Z.; Xie, Z.; Jha, R.; Steck, H.; Liang, D.; Feng, Y.; Majumder, B. P.; Kallus, N.; and McAuley, J. 2023. Large language models as zero-shot conversational recommenders. In *Proceedings of the 32nd ACM international conference on information and knowledge management*, 720–730.

Huang, Q.; Zhou, Z.; Yang, K.; Lin, G.; Yi, Z.; and Wang, Y. 2024. Leret: Language-empowered retentive network for time series forecasting. In *Proceedings of the 33rd International Joint Conference on Artificial Intelligence, IJCAI-24*.

Kang, W.-C.; and McAuley, J. 2018. Self-Attentive Sequential Recommendation. In *Proceedings of the 2018 IEEE International Conference on Data Mining*, 197–206.

Levandoski, J. J.; Sarwat, M.; Eldawy, A.; and Mokbel, M. F. 2012. Lars: A location-aware recommender system. In *2012 IEEE 28th international conference on data engineering*, 450–461. IEEE.

Li, P.; de Rijke, M.; Xue, H.; Ao, S.; Song, Y.; and Salim, F. D. 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*, 1463–1472.

Li, R.; Ebrahimi Kahou, S.; Schulz, H.; Michalski, V.; Charlin, L.; and Pal, C. 2018. Towards deep conversational recommendations. In *Advances in neural information processing systems*, volume 31.

Liang, J.; Hou, S.; Zhao, A.; Xu, Q.; Xiang, L.; Li, R.; and Wu, H. 2025. Design and application of a semantic-driven geospatial modeling knowledge graph based on large language models. *Geo-spatial Information Science*, 1–20.

Liu, C.; Miao, H.; Xu, Q.; Zhou, S.; Long, C.; Zhao, Y.; Li, Z.; and Zhao, R. 2025a. Efficient Multivariate Time Series Forecasting via Calibrated Language Models with Privileged Knowledge Distillation. In *IEEE International Conference on Data Engineering*.

Liu, C.; Xu, Q.; Miao, H.; Yang, S.; Zhang, L.; Long, C.; Li, Z.; and Zhao, R. 2024a. TimeCMA: Towards llm-empowered time series forecasting via cross-modality alignment. In *AAAI*, 2025.

Liu, C.; Yang, S.; Xu, Q.; Li, Z.; Long, C.; Li, Z.; and Zhao, R. 2024b. Spatial-Temporal Large Language Model for Traffic Prediction. In *MDM*, 31–40.

Liu, C.; Zhou, S.; Xu, Q.; Miao, H.; Long, C.; Li, Z.; and Zhao, R. 2025b. Towards Cross-Modality Modeling for Time Series Analytics: A Survey in the LLM Era. In *International Joint Conference on Artificial Intelligence*.

Liu, Y.; Miao, H.; Shen, G.; Zhao, Y.; Kong, X.; and Lee, I. 2025c. SPOT-Trip: Dual-Preference Driven Out-of-Town Trip Recommendation. In *Advances in Neural Information Processing Systems*.

Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; and Stoyanov, V. 2019. RoBERTa: A Robustly Optimized BERT Pretraining Approach. *arXiv preprint arXiv:1907.11692*.

Liu, Z.; Zhang, H.; Dong, K.; and Fang, Y. 2024c. Collaborative cross-modal fusion with large language model for recommendation. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, 1565–1574.

Ma, H.; Zou, J.; Aliannejadi, M.; Kanoulas, E.; Bin, Y.; and Yang, Y. 2024. Ask or Recommend: An Empirical Study on Conversational Product Search. In *Proceedings of the 33rd*

ACM International Conference on Information and Knowledge Management, 3927–3931.

Perez, E.; Strub, F.; De Vries, H.; Dumoulin, V.; and Courville, A. 2018. Film: Visual reasoning with a general conditioning layer. In *Proceedings of the AAAI conference on artificial intelligence*.

Shan, R.; Lin, J.; Zhu, C.; Chen, B.; Zhu, M.; Zhang, K.; Zhu, J.; Tang, R.; Yu, Y.; and Zhang, W. 2024. An automatic graph construction framework based on large language models for recommendation. *arXiv preprint arXiv:2412.18241*.

Sun, Y.; and Zhang, Y. 2018. Conversational recommender system. In *The 41st international acm SIGIR conference on research & development in information retrieval*, 235–244.

Wang, B.; Yang, Y.; Xu, X.; Hanjalic, A.; and Shen, H. T. 2017. Adversarial cross-modal retrieval. In *Proceedings of the 25th ACM international conference on Multimedia*, 154–162.

Wang, X.; Zhou, K.; Wen, J.-R.; and Zhao, W. X. 2022. Towards unified conversational recommender systems via knowledge-enhanced prompt learning. In *Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining*, 1929–1937.

Wang, Z.; Gao, Z.; Yang, Y.; Wang, G.; Jiao, C.; and Shen, H. T. 2024. Geometric matching for cross-modal retrieval. *IEEE Transactions on Neural Networks and Learning Systems*.

Wei, Y.; Zou, J.; Guo, W.; Wang, G.; Xu, X.; and Yang, Y. 2025. MSCRS: Multi-modal Semantic Graph Prompt Learning Framework for Conversational Recommender Systems. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 42–52.

Xia, J.; Yang, Y.; Wang, S.; Yin, H.; Cao, J.; and Yu, P. S. 2023. Bayes-enhanced multi-view attention networks for robust POI recommendation. *IEEE Transactions on Knowledge and Data Engineering*, 2895–2909.

Xu, R.; Cheng, H.; Guo, C.; Gao, H.; Hu, J.; Yang, S. B.; and Yang, B. 2025. Mm-path: Multi-modal, multi-granularity path representation learning. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 1*, 1703–1714.

Yang, A.; et al. 2025. Qwen3 Technical Report. *arXiv preprint arXiv:2505.09388*.

Ye, Y.; Zheng, Z.; Shen, Y.; Wang, T.; Zhang, H.; Zhu, P.; Yu, R.; Zhang, K.; and Xiong, H. 2025. Harnessing multimodal large language models for multimodal sequential recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 13069–13077.

Yuan, M.; Zhang, F.; Tong, Y.; Ying, Y.; Zhuang, F.; Wang, D.; Huai, B.; Zhang, Y.; and Su, J. 2024. Light POI-guided conversational recommender system based on adaptive space. In *Proceedings of the 2024 SIAM International Conference on Data Mining*, 724–733.

Zang, X.; Rastogi, A.; Sunkara, S.; Gupta, R.; Zhang, J.; and Chen, J. 2020. MultiWOZ 2.2: A dialogue dataset with additional annotation corrections and state tracking baselines. *arXiv preprint arXiv:2007.12720*.

Zhang, X.; Xin, X.; Li, D.; Liu, W.; Ren, P.; Chen, Z.; Ma, J.; and Ren, Z. 2023. Variational reasoning over incomplete knowledge graphs for conversational recommendation. In *Proceedings of the 16th ACM international conference on web search and data mining*, 231–239.

Zhang, Z.; Liu, S.; Liu, Z.; Zhong, R.; Cai, Q.; Zhao, X.; Zhang, C.; Liu, Q.; and Jiang, P. 2025. Llm-powered user simulator for recommender system. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 13339–13347.

Zhou, K.; Zhao, W. X.; Bian, S.; Zhou, Y.; Wen, J.-R.; and Yu, J. 2020a. Improving conversational recommender systems via knowledge graph based semantic fusion. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, 1006–1014.

Zhou, K.; Zhou, Y.; Zhao, W. X.; Wang, X.; and Wen, J.-R. 2020b. Towards topic-guided conversational recommender system. *arXiv preprint arXiv:2010.04125*.

Zhu, Q.; Huang, K.; Zhang, Z.; Zhu, X.; and Huang, M. 2020. Crosswoz: A large-scale chinese cross-domain task-oriented dialogue dataset. *Transactions of the Association for Computational Linguistics*, 8: 281–295.

Zou, J.; Aliannejadi, M.; Kanoulas, E.; Han, S.; Ma, H.; Wang, Z.; Yang, Y.; and Shen, H. T. 2025. PSCon: Product Search Through Conversations. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 3659–3669.

Zou, J.; Chen, Y.; and Kanoulas, E. 2020. Towards question-based recommender systems. In *Proceedings of the 43rd international ACM SIGIR conference on research and development in information retrieval*, 881–890.

Zou, J.; Huang, J.; Ren, Z.; and Kanoulas, E. 2022a. Learning to ask: Conversational product search via representation learning. *ACM Transactions on Information Systems*, 1–27.

Zou, J.; and Kanoulas, E. 2020. Towards question-based high-recall information retrieval: Locating the last few relevant documents for technology-assisted reviews. *ACM Transactions on Information Systems*, 1–35.

Zou, J.; Kanoulas, E.; Ren, P.; Ren, Z.; Sun, A.; and Long, C. 2022b. Improving conversational recommender systems via transformer-based sequential modelling. In *Proceedings of the 45th international ACM SIGIR conference on research and development in information retrieval*, 2319–2324.

Zou, J.; Lin, C.; Guo, W.; Wang, Z.; Wei, J.; Yang, Y.; and Shen, H. T. 2026. Multi-type context-aware conversational recommender systems via mixture-of-experts. *Information Fusion*, 103638.

Zou, J.; Sun, A.; Long, C.; and Kanoulas, E. 2024. Knowledge-enhanced conversational recommendation via transformer-based sequential modeling. *ACM Transactions on Information Systems*, 42(6): 1–27.