# DiMA: Distinguishing Resident and Tourist Preferences via Multi-Modal LLM Alignment for Out-of-Town Cross-Domain Recommendation

**Fan Zhang[1,3], Jinpeng Chen[1,2,3*], Tao Wang[1,3], Huan Li[4], Senzhang Wang[5], Feifei Kou[1,3], Ye Ji[6], Kaimin Wei[7], Zhenye Yang[1,3]**

[1]School of Computer Science (National Pilot Software Engineering School), Beijing University of Posts and Telecommunications
[2]The Open Project of Anhui Provincial Key Laboratory of Multimodal Cognitive Computation, Anhui University
[3]Key Laboratory of Trustworthy Distributed Computing and Service (BUPT), Ministry of Education
[4]The State Key Laboratory of Blockchain and Data Security, Zhejiang University
[5]School of Computer Science and Engineering, Central South University
[6]TravelSky Technology Limited
[7]College of Information Science and Technology, Jinan University
zhang_fan@bupt.edu.cn, jpchen@bupt.edu.cn

## Abstract

Out-of-Town (OOT) recommendation aims to provide personalized suggestions for users in unfamiliar cities. However, OOT recommendation faces two fundamental challenges: the difficulty of reasoning across modalities, as preference signals in disparate formats such as images and text are hard to compare; and the preference deviation problem, since a user's resident and tourist preferences often diverge, rendering simple preference transfer ineffective. To address these challenges, we propose Distinguishing Resident and Tourist Preferences via Multi-Modal LLM Alignment for Out-of-Town Cross-Domain Recommendation (DiMA), a framework for re-ranking Points of Interest (POIs). To tackle the multimodal challenge, DiMA first leverages Multimodal Large Language Models and Large Language Models (LLMs) to transform heterogeneous POI data into unified semantic tags, enabling both cross-modal reasoning and efficient downstream processing. To address preference deviation, a "teacher" LLM executes a custom Chain-of-Thought (CoT) process to disentangle resident and tourist preferences from multi-city histories for re-ranking. Finally, a lightweight student model learns this CoT reasoning via Supervised Fine-Tuning and is then refined with Direct Preference Optimization to align with true user choices, with the potential to surpass the teacher. Extensive experiments on a real-world dataset demonstrate that DiMA significantly enhances the performance of baseline models in the OOT recommendation re-ranking task.

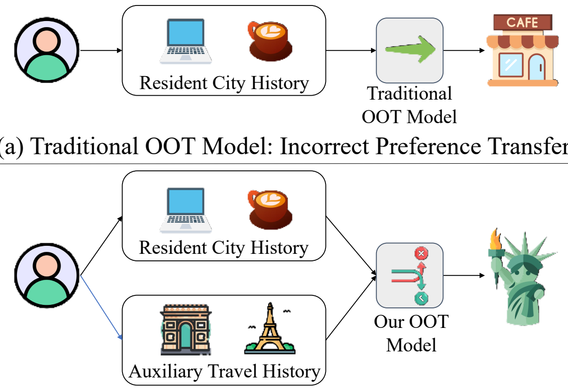
## Introduction

Cross-Domain Recommendation (CDR) (Ge et al. 2025; Li et al. 2025; Wang et al. 2025b; Rong et al. 2025) aims to alleviate the critical challenges of data sparsity and cold start by transferring knowledge from data-rich source domains to data-scarce target domains. Among these, the Out-of-Town (OOT) (Ference, Ye, and Lee 2013; Liu et al. 2024b) recommendation focuses on addressing the challenge of transferring user preferences across different cities. In this context, each city constitutes an independent domain with its unique Points of Interest (POIs) and user interaction patterns, where a user's resident city is considered the source domain, and the destination city they are visiting is the target domain. Existing OOT research (Yin et al. 2016; Ding et al. 2020; Li and Gong 2020) often addresses issues like data sparsity and interest drift by developing sophisticated models. However, these approaches often overlook two more fundamental challenges in the OOT scenario, limiting their effectiveness in real-world applications.

The first core challenge is the inherent difficulty of reasoning across modalities (Zhou, Pang, and Li 2024). The appeal of a POI is determined by a combination of its multifaceted characteristics, which are scattered across different data modalities like images (reflecting facts) and reviews (containing sentiments). To determine if a user would like a new venue, the model might need to connect a "vintage atmosphere" feature from an image the user previously liked with a "nostalgic style" mentioned in the new venue's reviews. However, since this information exists in entirely different formats, it cannot be directly compared or reasoned with, posing a key bottleneck for achieving precise preference matching. Recently, Large Language Models (LLMs) (Liu et al. 2025b,a) have shown promise in solving this problem, but how to guide them for structured reasoning and efficient deployment introduces new questions.

Building upon a unified representation of POIs, we face a second, more profound challenge: the preference deviation problem across different user scenarios. Current methods typically predict travel preferences by relying solely on a user's resident city history (Xin et al. 2022; Liu et al. 2024b). Although these methods may incorporate various debiasing techniques, they are constrained by the singularity of the information source, making it difficult to capture the distinct behavioral intents between a user's "resident mode" and "tourist mode." For instance, a user who frequently visits business-oriented and convenient restaurants in their resident city might prefer exploring places with unique scenery, local character, or historical significance while on vacation.

*Corresponding author.




(a) Traditional OOT Model: Incorrect Preference Transfer

(b) Our OOT Model: Preference Disentanglement

Figure 1: An illustration of the preference deviation problem in OOT recommendation. (a) Traditional models, which rely solely on resident history, often incorrectly transfer daily-life preferences (e.g., a functional cafe) to a tourist context. (b) Our framework leverages an auxiliary travel history (another city) to distinguish between resident and tourist preferences, enabling context-aware recommendations (e.g., an iconic landmark) through preference disentanglement.

As illustrated in Figure 1, preferences distilled from resident history can hardly predict true tourist needs.

To address the aforementioned challenges, we propose Distinguishing Resident and Tourist Preferences via Multi-Modal LLM Alignment for Out-of-Town Cross-Domain Recommendation (DiMA). Firstly, to overcome the multi-modal reasoning challenge, we design a novel multi-modal alignment and compression pipeline. This pipeline leverages Multimodal Large Language Models (MLLMs) and LLMs to map heterogeneous POI data (such as images and reviews) into a shared space of structured tags. This unified semantic foundation not only enables cross-modal reasoning but also significantly reduces the complexity and cost of subsequent LLM processing. Building on this foundation, to address the preference deviation problem, a large "teacher" LLM executes our carefully designed Chain-of-Thought (CoT) reasoning process. This process explicitly instructs the LLM to synthesize a "tourist-mode" profile from the user's resident and auxiliary travel histories, and then evaluate each candidate by integrating both semantic signals from the aligned multi-modal tags and collaborative signals from collaborative filtering (CF) scores. Finally, to ensure both efficiency and performance, we employ a two-stage alignment strategy. The student model first learns the teacher's complete CoT reasoning paradigm via Supervised Fine-Tuning (SFT). Subsequently, Direct Preference Optimization (DPO) further refines the model, aligning it with ground-truth user preferences and giving it the potential to surpass the teacher.

The main contributions are summarized as follows:

- We propose a novel multi-modal alignment and compression method that utilizes LLMs to transform heterogeneous data into unified structured tags, enabling low-cost cross-modal LLM reasoning in OOT scenarios.
- We identify and define the preference deviation problem in OOT recommendation and design a corresponding CoT reasoning mechanism that leverages an auxiliary travel history to enable the disentanglement and precise utilization of users' resident and tourist preferences.
- We develop the unified framework DiMA, which integrates these mechanisms with an advanced SFT+DPO alignment strategy, and demonstrate through experiments its effectiveness in achieving accurate and efficient recommendations in complex OOT tasks.

## Related Work

**Out-of-Town Recommendation.** Out-of-Town recommendation aims to address data sparsity and preference transfer challenges for users in new cities. Existing research has explored various strategies. For instance, models like ST-LDA (Ference, Ye, and Lee 2013) focus on modeling interest drift, while TRAINOR (Xin et al. 2021) infers high-level travel intentions. Others incorporate contextual signals, such as crowd-awareness in CAPTOR (Xin et al. 2022) or user sentiment in LSARS (Wang et al. 2017). More advanced methods like KDDC (Liu et al. 2024b) attempt to disentangle causal factors behind user choices using knowledge graphs. However, a common limitation across these approaches is their primary reliance on the user's resident city history to predict travel preferences. This overlooks the potential preference deviation between a user's "resident mode" and "tourist mode," where daily routines may not reflect exploratory interests during travel. Our work addresses this problem by incorporating an auxiliary travel history, enabling our model to explicitly distinguish between these two preference modes for more accurate OOT recommendations.

**Multi-Modal Recommendation.** The integration of multi-modal data enriches item representations in recommender systems (Yang and Yang 2024; Li et al. 2024). MMGCN (Wei et al. 2019) learns modality-specific representations on user-item graphs, while GRCN (Wei et al. 2020) refines the graph structure itself using multi-modal content. This paradigm has also been extended to cross-domain scenarios. MOTKD (Yang, Yang, and Liu 2023) uses knowledge distillation to transfer multi-modal knowledge, and P2M2-CDR (Wang et al. 2024) develops a privacy-preserving framework for this setting. However, their primary objective is to enhance recommendations for **overlapping users**. The challenge of the OOT scenario is to predict a **cold-start user's** preference in a new city. Furthermore, these approaches fuse multi-modal information into a latent vector, making it difficult to perform explicit reasoning across modalities. In contrast, DiMA aligns multi-modal information into a unified, structured tag space and enables cross-modal reasoning.

**Large Language Models in Recommendation.** LLMs are increasingly serving as the central reasoning engines in recommender systems (Ye et al. 2025). The current application paradigms can be broadly categorized. One line of work utilizes LLMs to enhance item or user feature representations, such as using them for semantic enhancement of

POIs (Cheng et al. 2025). A second, more dominant trend is to employ LLMs as powerful re-rankers or core models in sequential and session-based recommendation, leveraging their profound contextual understanding and reasoning capabilities (Du et al. 2025; Wang et al. 2025c; Liu et al. 2025c). Some works have begun to explore COT prompting specifically for recommendation tasks (Yue et al. 2025). Our work builds upon the re-ranking paradigm. However, applying general LLMs to the OOT scenario is challenging due to high inference latency. To overcome the efficiency and performance limitations, we introduce an SFT+DPO teacher-student alignment framework.

# The DiMA Framework

## Problem Formulation

Our work focuses on the re-ranking task in the OOT scenario. Formally, let $\mathcal{U}$ be the set of users and $\mathcal{P}$ be the set of POIs. Each POI $p \in \mathcal{P}$ is associated with a set of structured tags $\mathcal{T}_p$ extracted from its multi-modal information (e.g., images, reviews). For any user $u \in \mathcal{U}$, the input consists of:

- **Resident City History:** $H_u^{\text{res}} = \langle p_1, \ldots, p_m \rangle$.
- **Auxiliary Travel History:** $H_u^{\text{aux}} = \langle p'_1, \ldots, p'_n \rangle$.
- **Candidate List:** $L_u^{\text{tgt}} = [p''_1, \ldots, p''_k]$, a list of candidate POIs for the target city, generated by a baseline model.

We use the ground-truth interactions for DPO training and final evaluation. The goal is to learn a function $f_{\text{DiMA}}$:

$$L'_u = f_{\text{DiMA}}(H_u^{\text{res}}, H_u^{\text{aux}}, L_u^{\text{tgt}}) \quad (1)$$

where $L'_u$ is the re-ranked list. The objective is to optimize $f_{\text{DiMA}}$ to align the order of POIs in $L'_u$ with the user's true preferences, thereby maximizing ranking metrics.

## Overall Architecture

The DiMA framework, illustrated in Figure 2, is a multi-stage re-ranking pipeline composed of three primary modules. (1) The **Multi-Modal Alignment Module** is a universal pre-processor that uses MLLMs and LLMs to transform raw POI images and reviews into a unified set of structured tags. (2) The **Prompt Construction Module** then aggregates these tags with user histories and CF scores, embedding all information into a structured CoT prompt template. (3) Finally, the **Teacher-Student Alignment Module** uses this prompt to guide a "teacher" LLM in generating high-quality reasoning and a ranked list. This output, along with a "Winner Rank" from ground-truth data, is used to train a lightweight "student" model via a two-stage process of SFT and DPO for efficient deployment.

## Multi-Modal Alignment Module

To enable cross-modal reasoning, our first step is to map POI data (images and reviews) into a unified, structured semantic space. We employ a three-stage tagging pipeline to compress this information into a comparable set of semantic units.

**Image-to-Text Conversion.** First, we utilize a powerful MLLM, InternVL-2.5-8B-MPO (Wang et al. 2025a), to convert POI images into detailed textual descriptions.

**LLM-based Tag Extraction.** Next, DeepSeek-V3 (Liu et al. 2024a) serves as a tag extractor. Guided by prompts with strict constraints to ensure high quality and distinctiveness, it extracts two types of tags:

- Objective Entity Tags, which identify the core identity of a POI (e.g., "cafe", "pizza"), are extracted from both image descriptions and user reviews.
- Subjective Emotion Tags, which capture the affective dimension of user experiences (e.g., "cozy", "lively"), are sourced exclusively from user reviews.

**Tag Normalization.** Finally, to ensure tag consistency, we perform a normalization step. We apply lemmatization to objective tags (e.g., "pizzas" $\rightarrow$ "pizza") and stemming to emotion tags, where semantically similar words are mapped to a canonical form (e.g., "amazement" $\rightarrow$ "amazing").

This pipeline yields a unified set of normalized tags for each POI $p$, denoted as $\mathcal{T}_p = \mathcal{T}_p^{\text{img}} \cup \mathcal{T}_p^{\text{rev}} \cup \mathcal{T}_p^{\text{emo}}$, which enables cross-modal reasoning in the subsequent stages.

## Prompt Construction Module

The Prompt Construction Module, as depicted in Figure 2, aggregates all available signals and formats them into a single, structured textual input, guided by a meticulously designed CoT prompt template.

**Feature Aggregation** The module first compiles two types of features:

- Content Signals: The user's historical interactions from the resident ($H_u^{\text{res}}$) and auxiliary travel ($H_u^{\text{aux}}$) cities, along with the candidate POIs ($L_u^{\text{tgt}}$). Each POI is represented by its **metadata** (e.g., `name`, `category`, `popularity`) and its unified set of structured tags ($\mathcal{T}_p$).
- Collaborative Signals: Two CF scores ($s_{\text{cat}}$ and $s_{\text{ven}}$) for each candidate POI. To compute these, we first measure the similarity between the target user and every other user based on their shared interaction histories (in both resident and auxiliary cities) using the Jaccard index, calculated at both the *category-level* and the *venue-level*. The final scores for a candidate POI are then obtained by summing the corresponding similarities of all its visitors.

**Chain-of-Thought Prompting** All aggregated information is then embedded into a prompt template that enforces a structured reasoning process using CoT, as outlined in Algorithm 1. The design of these CoT steps is crucial for addressing the preference deviation problem by explicitly guiding the model's reasoning:

- Step 1 (Profile Synthesis): Forces the model to analyze and contrast the two histories to form a predictive hypothesis about the user's likely preferences as a tourist.
- Step 2 (Candidate Evaluation): Requires the model to evaluate each candidate POI against this synthesized tourist profile, integrating all available signals, including the aligned multi-modal tags and collaborative scores.
- Step 3 (Re-ranking): Generates the final list based on its evaluation based on the comprehensive evaluation.

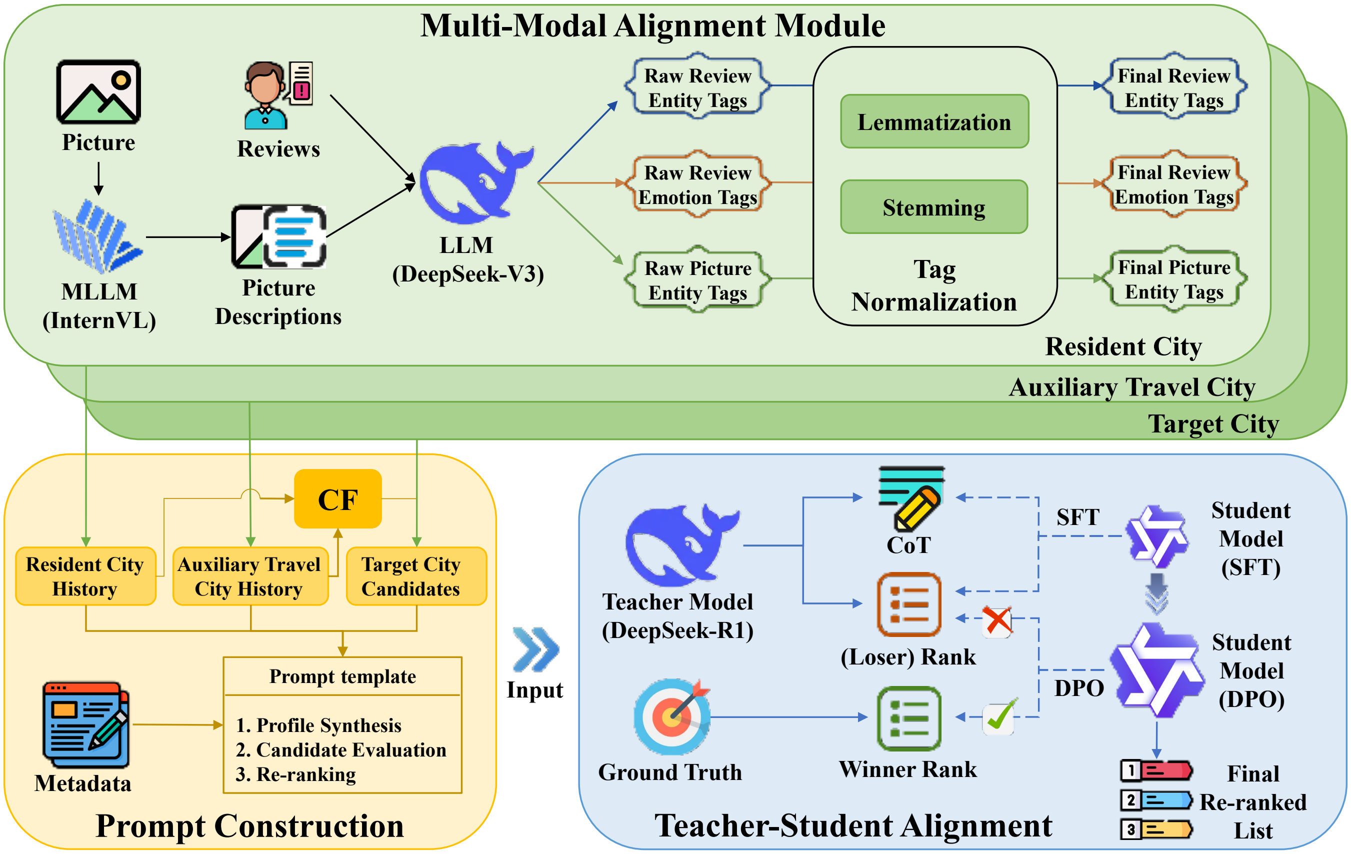



Figure 2: The overall architecture of the DiMA framework. The pipeline consists of three primary modules. (1) Multi-Modal Alignment Module: Raw POI data (images and reviews) are transformed into unified structured tags using MLLMs and LLMs, followed by a normalization step. This process is applied to all POIs across the resident, auxiliary travel, and target cities. (2) Prompt Construction Module: The module aggregates the processed user histories, target city candidates, CF scores, and other metadata, embedding them into a structured CoT prompt template. (3) Teacher-Student Alignment Module: A teacher LLM executes the CoT prompt to generate a reasoning trace and a ranked list. This, along with a "Winner Rank" derived from ground-truth data, is used to train a lightweight student model via a two-stage SFT and DPO process.

This fully constructed prompt serves as the unified input for both the teacher model's generation process and the student model's training and inference.

## Teacher-Student Alignment Module

This module forms the core of our framework, where a teacher model's complex reasoning is transferred to an efficient student model. The process involves the teacher's inference followed by the student's two-stage training.

**Teacher Model: Generating Training Data** The teacher model, DeepSeek-R1 (Guo et al. 2025), receives the structured prompt and executes the CoT process. By following the prescribed reasoning steps, the teacher generates a high-quality output comprising both the detailed reasoning text and the final ranked list. This complete output serves as the training data for the student model.

**Student Model: Two-Stage Alignment** To create an efficient yet powerful model for deployment, we train a lightweight student model, Qwen3-4B (Yang et al. 2025), using a two-stage alignment strategy.

**Stage 1: SFT for Reasoning Replication.** The first stage aims to teach the student model how to reason like the teacher. The student is trained to replicate the teacher's entire structured output by minimizing the standard auto-regressive loss (Equation 2). This step ensures the student masters the complex inference process.

$$\mathcal{L}_{\text{SFT}}(\theta) = -\sum_{i=1}^{|y_{\text{teacher}}|} \log P(y_i | y_{<i}, x; \theta) \tag{2}$$

where $\theta$ represents the student model's parameters and $y_{\text{teacher}}$ is the teacher's full response.

**Stage 2: DPO for User Alignment.** While SFT ensures the student imitates the teacher's reasoning style, its performance is capped. To potentially surpass the teacher, the second stage further refines the student using DPO (Rafailov et al. 2023) to directly align it with true user preferences.

Algorithm 1: Core Structure of the CoT Prompt Template

1: **Role Definition:** User Itinerary Predictor.
2: **Input Data Sections:** Resident History, Auxiliary Travel History, Candidate Locations.

3: **Task Steps (Reasoning):**
4: 1. **Profile Synthesis:** Analyze the resident and travel histories to form a “tourist-mode” profile.
5: 2. **Candidate Evaluation:** Evaluate candidates using all signals (tags, CF scores, etc.).
6: 3. **Re-ranking:** Justify the final ranking order.

7: **Output Format Specification:**
8: **Part 1:** Reasoning Steps (CoT text).
9: **Part 2:** Final JSON Output (a list of ranked ids).

**Preference Data Construction.** DPO requires a dataset of preference triplets $(x, y_w, y_l)$. The context $x$ is the input prompt. The dispreferred (loser) response $y_l$ is the ranked list generated by our teacher model. The preferred (winner) response $y_w$ is an ideal response constructed from ground-truth interactions by re-ranking the candidate list such that all “hit” POIs are ranked higher than “miss” POIs, while preserving the teacher’s relative ordering within each set.

**Optimization Objective.** DPO optimizes the model by maximizing the log-likelihood of the preferred responses, using the objective function defined in Equation 3.

$$\mathcal{L}_{\text{DPO}}(\theta, \theta_{\text{ref}}) = -\mathbb{E}_{(x,y_w,y_l)\sim\mathcal{D}}\Bigg[\log\sigma\Big(\beta\log\frac{\pi_\theta(y_w|x)}{\pi_{\theta_{\text{ref}}}(y_w|x)} - \beta\log\frac{\pi_\theta(y_l|x)}{\pi_{\theta_{\text{ref}}}(y_l|x)}\Big)\Bigg] \quad (3)$$

where $\pi_\theta$ is the policy being trained, $\pi_{\theta_{\text{ref}}}$ is the reference policy from SFT, and $\beta$ is a temperature parameter. This loss encourages the model to assign a higher probability to the winner response $y_w$ over the loser response $y_l$, allowing it to potentially achieve performance superior to the teacher.

### Inference Process

During inference, DiMA leverages only the lightweight final student model to ensure efficiency. For a given user request, the framework first performs the same pre-processing steps as in training, including multi-modal tagging and collaborative signal calculation for the candidate POIs. Next, the user’s histories and the processed candidates are formatted into the structured CoT prompt. This prompt is then fed into the trained student model, which autoregressively generates a response containing both the CoT reasoning and a JSON object with the final re-ranked list. Finally, this list is parsed and presented as the recommendation.

## Experiments

In this section, we conduct a series of experiments to comprehensively evaluate the performance of our proposed DiMA framework on the OOT re-ranking task.

### Dataset

Publicly available datasets for OOT recommendation rarely contain rich multi-modal information. Therefore, we first constructed a new multi-modal dataset from Foursquare’s API. We collected data from three major US cities: New York (NY), Los Angeles (LA), and San Francisco (SF). For each POI, we downloaded its top 10 images and top 10 tips. To create comprehensive OOT scenarios, we generate six distinct experimental settings by cyclically assigning the roles of resident, auxiliary travel, and target city among the three cities (NY, LA, and SF). We identified each user’s primary city of activity as their resident city. We pre-processed the dataset by retaining only users who have interaction records in all three cities and POIs that have been visited at least five times. After filtering, the final dataset contains 1,072 users (who are active in all three cities) and a total of 8,478 unique POIs distributed across New York (3,831), Los Angeles (2,516), and San Francisco (2,131).

To further validate the generalizability of our framework, we conduct experiments on a second dataset derived from the public Yelp dataset. We selected three cities with a high degree of user overlap: New Orleans (NO), Philadelphia (PH), and St. Louis (SL). Following a similar preprocessing methodology, we filtered for POIs with at least five images and five reviews, and retained only users present in all three cities. After filtering, this Yelp dataset consists of 593 users and a total of 2,158 unique POIs, distributed across New Orleans (681), Philadelphia (1,015), and St. Louis (462).

### Evaluation Metrics

We evaluate the re-ranking performance using Normalized Discounted Cumulative Gain (NDCG@K) and Mean Average Precision (MAP@K), with K set to 5. All reported results are the average of a 5-fold cross-validation.

### Baselines and Variants

We compare DiMA against several models on candidate lists generated by two state-of-the-art OOT baselines: **KDDC** (Liu et al. 2024b) and **CAPTOR** (Xin et al. 2022). The compared re-rankers include: (1) **Teacher-2C**, the teacher model (DeepSeek-R1) prompted without the auxiliary travel history; (2) **DiMA-2C**, Teacher-2C corresponding student model; (3) **Teacher-3C**, the full teacher model prompted with all three city histories; and (4) **DiMA (Ours)**, our full student model.

### Implementation Details

Our framework is implemented using PyTorch and LLaMA-Factory. We use InternVL-2.5-8B-MPO for image-to-text conversion, DeepSeek-V3 for tag extraction, DeepSeek-R1 as the teacher model, and Qwen3-4B as our primary student model. History and candidate list lengths are capped at 15/10 and 10, respectively, with a re-ranked output of 5. The SFT stage was run on two NVIDIA RTX 3090 GPUs, and the DPO stage on an NVIDIA H20 GPU. We followed the **KDDC** (Liu et al. 2024b) paper’s protocol, splitting users 8:1:1 into train/val/test to train the baseline models. We used

| Model | LA→NY→SF | | LA→SF→NY | | SF→NY→LA | | SF→LA→NY | | NY→LA→SF | | NY→SF→LA | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **N@5** | **M@5** | **N@5** | **M@5** | **N@5** | **M@5** | **N@5** | **M@5** | **N@5** | **M@5** | **N@5** | **M@5** |
| KDDC | 0.364 | 0.515 | 0.288 | 0.399 | 0.378 | 0.544 | 0.260 | 0.403 | 0.377 | 0.509 | 0.406 | 0.526 |
| Teacher-2C | 0.426 | 0.607 | 0.394 | 0.571 | 0.390 | 0.576 | 0.335 | 0.526 | 0.458 | 0.608 | 0.466 | 0.590 |
| DiMA-2C | 0.442 | 0.650 | 0.391 | 0.557 | 0.390 | 0.581 | 0.333 | 0.525 | 0.460 | 0.615 | 0.462 | 0.592 |
| Teacher-3C | 0.427 | 0.605 | 0.397 | 0.571 | 0.402 | 0.586 | 0.353 | 0.558 | 0.465 | 0.634 | **0.481** | 0.608 |
| **DiMA (Ours)** | **0.457** | **0.694** | **0.399** | **0.581** | **0.426** | **0.638** | **0.372** | **0.609** | **0.477** | **0.662** | 0.476 | **0.632** |
| CAPTOR | 0.371 | 0.528 | 0.343 | 0.492 | 0.391 | 0.590 | 0.289 | 0.444 | 0.308 | 0.407 | 0.397 | 0.535 |
| Teacher-2C | 0.400 | 0.566 | 0.415 | 0.577 | 0.405 | 0.604 | 0.311 | 0.474 | 0.453 | 0.601 | 0.445 | 0.592 |
| DiMA-2C | 0.396 | 0.555 | 0.436 | 0.607 | 0.397 | 0.596 | 0.320 | 0.491 | 0.453 | 0.607 | 0.440 | 0.585 |
| Teacher-3C | 0.394 | 0.551 | 0.426 | 0.602 | 0.403 | 0.600 | 0.317 | 0.492 | 0.466 | 0.621 | 0.442 | 0.596 |
| **DiMA (Ours)** | **0.414** | **0.601** | **0.438** | **0.615** | **0.421** | **0.632** | **0.323** | **0.500** | **0.471** | **0.653** | **0.448** | **0.600** |

Table 1: Overall performance comparison on the Foursquare dataset. We report NDCG@5 (N@5) and MAP@5 (M@5). The top part shows results on KDDC's candidates, and the bottom part on CAPTOR's. The best result in each column is in bold.

| Model | LA→NY→SF | | LA→SF→NY | | SF→NY→LA | | SF→LA→NY | | NY→LA→SF | | NY→SF→LA | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **N@5** | **M@5** | **N@5** | **M@5** | **N@5** | **M@5** | **N@5** | **M@5** | **N@5** | **M@5** | **N@5** | **M@5** |
| KDDC | 0.029 | 0.030 | 0.099 | 0.151 | 0.097 | 0.129 | 0.071 | 0.117 | 0.160 | 0.181 | 0.128 | 0.158 |
| Teacher-2C | 0.032 | 0.043 | 0.114 | 0.175 | 0.156 | 0.204 | 0.123 | 0.204 | 0.195 | 0.218 | 0.172 | 0.227 |
| DiMA-2C | 0.034 | 0.053 | 0.112 | 0.174 | **0.188** | **0.257** | 0.114 | 0.178 | 0.217 | 0.253 | 0.164 | 0.210 |
| Teacher-3C | **0.040** | **0.071** | 0.133 | 0.204 | 0.180 | 0.237 | **0.136** | 0.205 | 0.215 | 0.240 | 0.184 | 0.248 |
| **DiMA (Ours)** | 0.035 | 0.060 | **0.146** | **0.213** | 0.171 | 0.226 | 0.130 | **0.207** | **0.216** | **0.266** | **0.190** | **0.258** |
| CAPTOR | 0.043 | 0.057 | 0.089 | 0.124 | 0.131 | 0.177 | 0.063 | 0.101 | 0.187 | 0.214 | 0.092 | 0.117 |
| Teacher-2C | 0.055 | 0.062 | 0.093 | 0.159 | 0.145 | 0.181 | 0.124 | 0.202 | 0.219 | 0.252 | 0.131 | 0.160 |
| DiMA-2C | 0.061 | 0.069 | 0.090 | 0.150 | 0.143 | 0.179 | 0.093 | 0.159 | 0.230 | 0.270 | 0.126 | 0.158 |
| Teacher-3C | 0.057 | 0.062 | 0.117 | 0.157 | 0.140 | 0.176 | **0.128** | **0.211** | 0.227 | 0.267 | 0.134 | 0.157 |
| **DiMA (Ours)** | **0.092** | **0.135** | **0.124** | **0.170** | **0.148** | **0.185** | 0.085 | 0.160 | **0.241** | **0.288** | **0.135** | **0.161** |

Table 2: Overall performance comparison on the Yelp dataset. The best result in each column is in bold.

the baseline's outputs on these splits: the validation set outputs were used to generate teacher data for fine-tuning our student model, and all final results are reported on the test set. To prevent leakage, CF scores were computed on the training set only.

## Overall Performance

The main experimental results on the Foursquare and Yelp datasets are presented in Table 1 and Table 2, respectively. For each task denoted as **Resident→Auxiliary→Target** city, two-city models use only resident and target data. The analysis yields several key findings.

First, our proposed DiMA framework consistently outperforms all baselines across both datasets and most scenarios, confirming the overall effectiveness of our LLM-based re-ranking paradigm. Second, the auxiliary travel history is crucial for addressing preference deviation. The three-city models (Teacher-3C, DiMA) generally outperform their two-city counterparts, validating our hypothesis that leveraging a user's "tourist mode" history enables more accurate preference modeling. Third, our student model, DiMA, frequently surpasses its own teacher. DiMA's consistent outperformance of Teacher-3C highlights the success of our SFT+DPO alignment strategy, which refines the student with ground-truth data to achieve a higher performance ceiling. Finally, DiMA exhibits remarkable generalization. All student models were trained **only once** on the Foursquare-KDDC task. This single trained model was then applied in a zero-shot manner to re-rank candidates from CAPTOR and the entirely separate Yelp dataset. The consistent gains across these scenarios demonstrate that DiMA learns a robust, generalizable re-ranking logic, rather than overfitting to a specific candidate generator or dataset.

## Ablation Study

To investigate the contribution of each key component in DiMA, we conduct an ablation study by creating four variants. All variants are based on the KDDC baseline and the Foursquare dataset, and the results are averaged across all six OOT tasks. The variants are defined as follows:

- DiMA w/o Tags: We remove all multi-modal tags to verify the effectiveness of our multi-modal alignment.

| Model Variant | NDCG | | MAP | |
|---|---|---|---|---|
| | @3 | @5 | @3 | @5 |
| **DiMA (Full Model)** | **0.438** | **0.435** | **0.619** | **0.636** |
| DiMA w/o Tags | 0.433 | 0.432 | 0.609 | 0.631 |
| DiMA w/o CF | 0.431 | 0.422 | 0.599 | 0.613 |
| DiMA w/o CoT | 0.391 | 0.384 | 0.524 | 0.526 |
| DiMA w/o DPO | 0.424 | 0.412 | 0.578 | 0.578 |

Table 3: Ablation study results. Performance is averaged over all six OOT tasks on the KDDC baseline.

- DiMA w/o CF: We remove the two CF scores to test the importance of collaborative signals.
- DiMA w/o CoT: To isolate the value of reasoning, the student's SFT stage is modified to imitate only the teacher's final ranked list, ignoring the CoT text.
- DiMA w/o DPO: The student model is trained only with the SFT stage, omitting the DPO step.

The results of our ablation study are presented in Table 3. First, the most significant performance drop occurs in the **DiMA w/o CoT** variant. This result strongly validates our core hypothesis: teaching the student model *how to reason* by distilling the teacher's CoT process is substantially more effective than merely imitating its final ranking output. Second, **DiMA w/o Tags** and **CF** both lead to performance degradation. This confirms that both the unified semantic representation derived from multi-modal data and the behavioral signals from like-minded users are valuable information that the LLM effectively utilizes in its reasoning. Finally, the performance gap between the full DiMA model and **DiMA w/o DPO** demonstrates the effectiveness of the DPO stage. By directly aligning the model with ground-truth user choices, DPO provides a crucial refinement step that further boosts the final recommendation accuracy.

## Analysis on Student Model Scale and Efficiency

We evaluate DiMA's performance and efficiency using four student models of varying scales: Qwen2.5-1.5B, Qwen2.5-3B, Qwen3-1.7B, and Qwen3-4B. The results, averaged across all tasks on the Foursquare-KDDC setting, are presented in Figure 3. The analysis highlights a clear trade-off between performance and efficiency.

As shown in the figure, all student models, regardless of size, significantly outperform the KDDC baseline, confirming the fundamental effectiveness of our framework. Performance generally improves with model scale, with Qwen3-4B achieving the best results. A key motivation for our framework is achieving low-latency inference. The teacher model requires 5 minutes per request. In stark contrast, our student models are highly efficient when deployed on an NVIDIA RTX 3090 GPU: the top-performing Qwen3-4B responds in 10 seconds, while the most compact Qwen2.5-1.5B responds in 1 second. This demonstrates that DiMA successfully transfers the teacher's reasoning into practical models, offering a flexible choice between maximum performance and real-time inference speed.

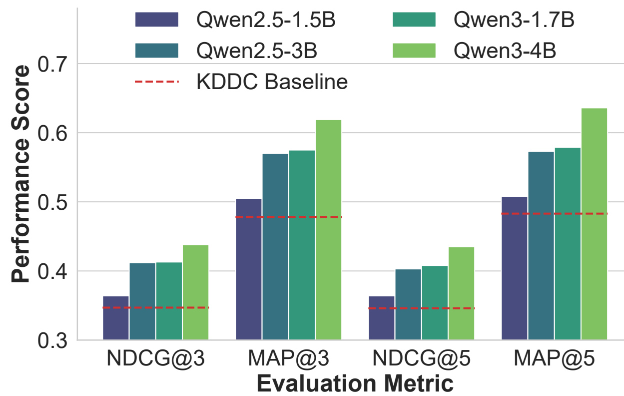



Figure 3: Performance comparison of DiMA with different student model backbones against the KDDC baseline.

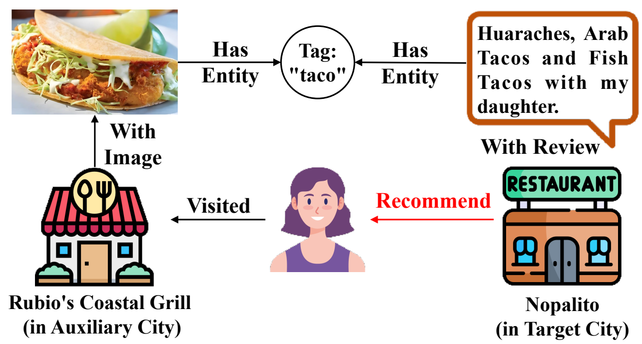



Figure 4: A case study of DiMA's cross-modal reasoning.

## Case Study: Cross-Modal Reasoning

To illustrate DiMA's cross-modal reasoning capabilities, we present a case study in Figure 4. DiMA recommended the restaurant "Nopalito" after identifying a key cross-modal link between the user's history and the candidate. As illustrated in the figure, the user had previously visited "Rubio's Coastal Grill", from which our framework extracted the entity tag "taco" from a visual source (an image). For the candidate POI "Nopalito", the same "taco" tag was extracted from a textual source (a review). This case study demonstrates that by mapping heterogeneous data into a unified tag space, DiMA successfully performs cross-modal reasoning.

## Conclusion

In this paper, we propose DiMA, a framework to address the core challenges of preference deviation and cross-modal reasoning in OOT recommendation. DiMA operates via a three-module pipeline: it first aligns POI images and reviews into unified structured tags using LLMs; then, a "teacher" LLM, guided by a CoT prompt, disentangles user preferences by leveraging multi-source histories (resident and auxiliary travel); finally, a two-stage SFT+DPO strategy efficiently transfers the teacher's capabilities to a lightweight "student" model. Our experiments demonstrate that DiMA significantly outperforms baselines across two datasets.

## Acknowledgements

This work was supported in part by the National Natural Science Foundation of China (62572075, 62572217, 62002027), in part by the Beijing Natural Science Foundation (L233034, L253004), in part by Fundamental Research Funds for the Beijing University of Posts and Telecommunications (No.2025TSQY01), in part by the Zhejiang Provincial Natural Science Foundation (No.LD24F020015), in part by the Open Project of Anhui Provincial Key Laboratory of Multimodal Cognitive Computation, Anhui University, No.MMC202408 and part by Natural Science Foundation of Guangdong Province (2025A1515010739).

## References

Cheng, J.; Wang, J.; Zhang, Y.; Ji, J.; Zhu, Y.; Zhang, Z.; and Zhao, X. 2025. Poi-enhancer: An llm-based semantic enhancement framework for poi representation learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 39, 11509–11517.

Ding, D.; Zhang, M.; Pan, X.; Yang, M.; and He, X. 2020. Modeling personalized out-of-town distances in location recommendation. In *2020 IEEE International Conference on Data Mining (ICDM)*, 112–121. IEEE.

Du, Y.; Sun, Z.; Wang, Z.; Chua, H.; Zhang, J.; and Ong, Y.-S. 2025. Active large language model-based knowledge distillation for session-based recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 11607–11615.

Ference, G.; Ye, M.; and Lee, W.-C. 2013. Location recommendation for out-of-town users in location-based social networks. In *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*, 721–726.

Ge, C.; Wang, X.; Zhang, Z.; Qin, Y.; Chen, H.; Wu, H.; Zhang, Y.; Yang, Y.; and Zhu, W. 2025. Behavior importance-aware graph neural architecture search for cross-domain recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 11708–11716.

Guo, D.; Yang, D.; Zhang, H.; Song, J.; Wang, P.; Zhu, Q.; Xu, R.; Zhang, R.; Ma, S.; Bi, X.; et al. 2025. Deepseek-r1 incentivizes reasoning in llms through reinforcement learning. *Nature*, 645(8081): 633–638.

Li, D.; and Gong, Z. 2020. A deep neural network for crossing-city POI recommendations. *IEEE Transactions on Knowledge and Data Engineering*, 34(8): 3536–3548.

Li, H.; Wang, Y.; Xiao, Z.; Yang, J.; Zhou, C.; Zhang, M.; and Ju, W. 2025. DisCo: graph-based disentangled contrastive learning for cold-start cross-domain recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 12049–12057.

Li, Z.; Liu, F.; Wei, Y.; Cheng, Z.; Nie, L.; and Kankanhalli, M. 2024. Attribute-driven Disentangled Representation Learning for Multimodal Recommendation. In *Proceedings of the 32nd ACM International Conference on Multimedia*, 9660–9669.

Liu, A.; Feng, B.; Xue, B.; Wang, B.; Wu, B.; Lu, C.; Zhao, C.; Deng, C.; Zhang, C.; Ruan, C.; et al. 2024a. Deepseek-v3 technical report. *arXiv preprint arXiv:2412.19437*.

Liu, C.; Xu, Q.; Miao, H.; Yang, S.; Zhang, L.; Long, C.; Li, Z.; and Zhao, R. 2025a. Timecma: Towards llm-empowered multivariate time series forecasting via cross-modality alignment. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 18780–18788.

Liu, P.; Guo, H.; Dai, T.; Li, N.; Bao, J.; Ren, X.; Jiang, Y.; and Xia, S.-T. 2025b. Calf: Aligning llms for time series forecasting via cross-modal fine-tuning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 18915–18923.

Liu, S.; Ding, R.; Lu, W.; Wang, J.; Yu, M.; Shi, X.; and Zhang, W. 2025c. Coherency Improved Explainable Recommendation via Large Language Model. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 12201–12209.

Liu, Y.; Shen, G.; Cui, C.; Zhao, Z.; Han, X.; Du, J.; Zhao, X.; and Kong, X. 2024b. Kddc: Knowledge-driven disentangled causal metric learning for pre-travel out-of-town recommendation. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence, Jeju Island, Republic of Korea*, 4–9.

Rafailov, R.; Sharma, A.; Mitchell, E.; Manning, C. D.; Ermon, S.; and Finn, C. 2023. Direct preference optimization: Your language model is secretly a reward model. *Advances in neural information processing systems*, 36: 53728–53741.

Rong, N.; Xiong, F.; Pan, S.; Luo, G.; Wu, J.; and Wang, L. 2025. Domain-Level Disentanglement Framework Based on Information Enhancement for Cross-Domain Cold-Start Recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 12488–12496.

Wang, H.; Fu, Y.; Wang, Q.; Yin, H.; Du, C.; and Xiong, H. 2017. A location-sentiment-aware recommender system for both home-town and out-of-town users. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, 1135–1143.

Wang, L.; Sang, L.; Zhang, Q.; Wu, Q.; and Xu, M. 2024. A privacy-preserving framework with multi-modal data for cross-domain recommendation. *Knowledge-Based Systems*, 304: 112529.

Wang, W.; Chen, Z.; Wang, W.; Cao, Y.; Liu, Y.; Gao, Z.; Zhu, J.; Zhu, X.; Lu, L.; Qiao, Y.; and Dai, J. 2025a. Enhancing the Reasoning Ability of Multimodal Large Language Models via Mixed Preference Optimization. arXiv:2411.10442.

Wang, Y.; Jian, Y.; Yang, W.; Lu, S.; Shen, L.; Wang, B.; Zeng, X.; and Zhang, L. 2025b. Towards unbiased information extraction and adaptation in cross-domain recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 12757–12765.

Wang, Z.; Du, Y.; Sun, Z.; Chua, H.; Feng, K.; Wang, W.; and Zhang, J. 2025c. Re2llm: reflective reinforcement large language model for session-based recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 12827–12835.

Wei, Y.; Wang, X.; Nie, L.; He, X.; and Chua, T.-S. 2020. Graph-refined convolutional network for multimedia recommendation with implicit feedback. In *Proceedings of the 28th ACM international conference on multimedia*, 3541–3549.

Wei, Y.; Wang, X.; Nie, L.; He, X.; Hong, R.; and Chua, T.-S. 2019. MMGCN: Multi-modal graph convolution network for personalized recommendation of micro-video. In *Proceedings of the 27th ACM international conference on multimedia*, 1437–1445.

Xin, H.; Lu, X.; Xu, T.; Liu, H.; Gu, J.; Dou, D.; and Xiong, H. 2021. Out-of-town recommendation with travel intention modeling. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, 4529–4536.

Xin, H.; Lu, X.; Zhu, N.; Xu, T.; Dou, D.; and Xiong, H. 2022. Captor: A crowd-aware pre-travel recommender system for out-of-town users. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1174–1184.

Yang, A.; Li, A.; Yang, B.; Zhang, B.; Hui, B.; Zheng, B.; Yu, B.; Gao, C.; Huang, C.; Lv, C.; et al. 2025. Qwen3 technical report. *arXiv preprint arXiv:2505.09388*.

Yang, W.; Yang, J.; and Liu, Y. 2023. Multimodal Optimal Transport Knowledge Distillation for Cross-domain Recommendation. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, 2959–2968.

Yang, W.; and Yang, Q. 2024. Multimodal-aware Multi-intention Learning for Recommendation. In *Proceedings of the 32nd ACM International Conference on Multimedia*, 5663–5672.

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*, volume 39, 13069–13077.

Yin, H.; Zhou, X.; Cui, B.; Wang, H.; Zheng, K.; and Nguyen, Q. V. H. 2016. Adapting to user interest drift for poi recommendation. *IEEE Transactions on Knowledge and Data Engineering*, 28(10): 2566–2581.

Yue, W.; Yin, Y.; Zhang, X.; Shi, B.; Liang, T.; and Wan, J. 2025. CoT4Rec: Revealing User Preferences Through Chain of Thought for Recommender Systems. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 13142–13151.

Zhou, D.; Pang, Z.; and Li, W. 2024. Learning Cross-modal Knowledge Reasoning and Heuristic-prompt for Visual-language Navigation. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, 3453–3462.