

Harnessing LLMs for Cross-City OD Flow Prediction

Chenyang Yu, Xinpeng Xie, Yan Huang, and Chenxi Qiu

{chenyangyu,xinpengxie}@my.unt.edu

{yan.huang,chenxi.qiu}@unt.edu

Department of Computer Science and Engineering, University of North Texas
Denton, Texas, USA

ABSTRACT

Understanding and predicting *Origin-Destination (OD)* flows is crucial for urban planning and transportation management. Traditional OD prediction models, while effective within single cities, often face limitations when applied across different cities due to varied traffic conditions, urban layouts, and socio-economic factors.

In this paper, by employing *Large Language Models (LLMs)*, we introduce a new method for cross-city OD flow prediction. Our approach leverages the advanced semantic understanding and contextual learning capabilities of LLMs to bridge the gap between cities with different characteristics, providing a robust and adaptable solution for accurate OD flow prediction that can be transferred from one city to another. Our novel framework involves four major components: collecting OD training datasets from a source city, instruction-tuning the LLMs, predicting destination POIs in a target city, and identifying the locations that best match the predicted destination POIs. We introduce a new loss function that integrates POI semantics and trip distance during training. By extracting high-quality semantic features from human mobility and POI data, the model understands spatial and functional relationships within urban spaces and captures interactions between individuals and various POIs. Extensive experimental results demonstrate the superiority of our approach over the state-of-the-art learning-based methods in cross-city OD flow prediction.

CCS CONCEPTS

• **Information systems** → **Location based services**; • **Computing methodologies** → **Artificial intelligence**.

KEYWORDS

Large Language Models(LLMs), Urban Computing, origin-destination, Cross-City Transferability

ACM Reference Format:

Chenyang Yu, Xinpeng Xie, Yan Huang, and Chenxi Qiu. 2024. Harnessing LLMs for Cross-City OD Flow Prediction. In *The 32nd International Conference on Advances in Geographic Information Systems (SIGSPATIAL '24)*, October 29 - November 1, 2024, Atlanta, GA, USA, 12 pages. <https://doi.org/10.1145/3557915.3560938>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

SIGSPATIAL '24, October 29 - November 1, 2024, Atlanta, GA, USA

© 2024 Association for Computing Machinery.

ACM ISBN 978-1-4503-9529-8/22/11...\$15.00

<https://doi.org/10.1145/3557915.3560938>

1 INTRODUCTION

Origin and destination (OD) matrices are invaluable tools in transportation planning and urban development. By capturing the flow of people, goods, or services between different locations within a geographic area, OD matrices provide critical insights into travel patterns, demand for infrastructure, and spatial interactions. Urban planners and policymakers utilize OD matrices to identify transportation bottlenecks [5, 30], optimize transit routes [1, 7], and allocate resources efficiently [27, 30].

Gathering accurate and comprehensive data on travel movements within a geographic area can be time-consuming and costly. Data collection efforts may face challenges such as incomplete or inconsistent data, limited data availability for certain modes of transportation or population groups, and privacy concerns related to personal travel data.

Traditional OD estimation models such as [18] and the Radiation Model [25] rely on the assumptions that flow of people or goods between two locations is proportional to their populations and inversely proportional to the square of their distance. They have been adapted to incorporate novel data sources such as social media check-ins [21], enhancing their adaptability.

Recent learning-based models excel in capturing implicit data characteristics and delivering commendable performance [9, 36]. While effective in achieving their objectives, above works primarily focus on the OD flow generation within a single city. Some recent works [28] propose to generate trajectories across cities by learning traffic patterns and vehicle behaviors from data-rich cities and extrapolating this information to generate datasets for cities lacking such data. However, it heavily depends on the vehicle trajectory and road network datasets specific to the target city. Transferring the learned patterns from these datasets to different cities is challenging due to variations in traffic conditions and urban layouts among cities.

Large language models (LLMs) [3, 6, 8, 23, 24] have demonstrated impressive capabilities across diverse tasks by recognizing patterns in vast datasets, enabling them to achieve high accuracy in various applications, from natural language processing to fields such as biomedical imaging and crop management [16, 33]. An interesting question is whether LLMs can understand and predict likely trip destinations based on a given point of interest and timeframe in the same way humans do. Has some base knowledge about trip behaviors learned across cities and once fine-tuned for flow prediction in one city and can be easily transferred to a different city? These capabilities hinge on the LLMs' ability to consider various factors and make informed guesses about trip destination as well as recognize patterns and associations in trip data. The main challenges of leveraging an LLM model are that: (1) the models were pre-trained on texts and do not have explicit knowledge of OD flows; (2) The

primary objective of the loss function of LLMs is to improve the model's ability to predict the next token in a sequence accurately, thereby enhancing its performance in generating coherent and contextually relevant text. We need to design new loss functions that take the spatial context into consideration.

1.1 Our Contributions

In this paper, we propose a framework to investigate and harness large language models' capabilities in understanding trips for OD flow prediction across cities. Our approach is driven by the following premises: (1) despite the considerable variations in their trips across different cities, people travel for specific purposes and tend to follow similar patterns in their Origin-Destination POIs [11, 17] and (2) LLMs are trained on vast amounts of text data, including information about transportation, urban planning, and human behavior, and has basic understanding of city mobility patterns.

- ▶ **Contribution 1.** We propose an *LLM based Cross-city OD (LLM-COD)* flow prediction framework. LLM-COD leverages the knowledge possessed by LLMs and enhance it to infuse understanding of origin and destination related to point of interests, time, and distance. LLM-COD captures how people interact with various POIs in the city related to time.
- ▶ **Contribution 2.** We develop a process to allow LLM-COD to transfer the knowledge to other cities and perform OD flow prediction with adaptation. We propose a new loss function that integrates POI semantics with distance which helps LLM-COD to achieve robust performance across cities. This means that even with relatively limited training data for a new city, the model can utilize the knowledge and mobility patterns learned from other cities to improve prediction accuracy.
- ▶ **Contribution 3.** We conducted extensive experiments to evaluate our method. The results demonstrate that our approach significantly outperforms traditional models and state-of-the-art learning-based methods in cross-city OD flow prediction. LLM-COD reduces RMSE significantly compared to the best state-of-the-art model especially for challenging high precision OD flow prediction (reduces by 46% for $1,000m \times 1,000m$) and is consistently better in all other commonly used metrics.

2 OVERVIEW

We consider a scenario where one city, referred to as "City A", possesses an Origin-Destination (OD) dataset, which includes the origin and destination points of various vehicle trips. In contrast, the other city, referred to as "City B", does not have access to the OD dataset. Our objective is to utilize the OD dataset from City A to train a generative model that can subsequently produce an OD dataset for City B.

Our approach is motivated by the observation that, despite significant variations in individuals' trips across different urban environments—attributable to varied transportation modes and city layouts—*people in different cities exhibit similar patterns in their Origin-Destination Points of Interest (POIs)*. For example, a universal OD pattern is the daily commute between residential areas and workplaces [15]. Additionally, educational institutions like schools,

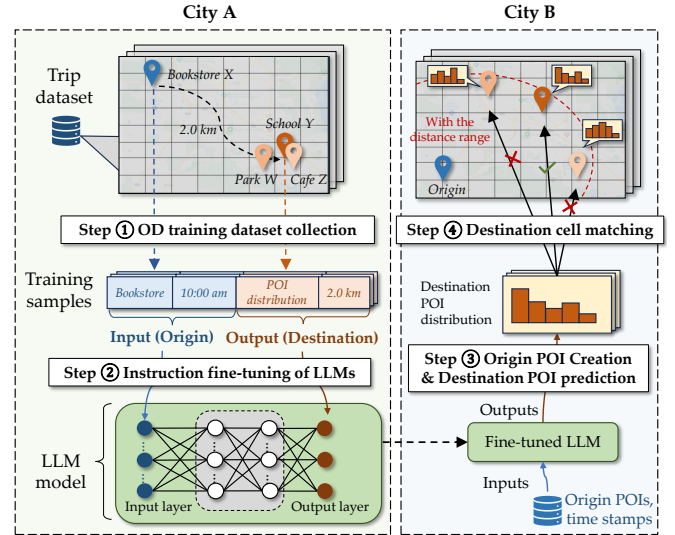


Figure 1: LLM-COD Framework.

colleges, and universities serve as major OD points in cities, creating distinct travel patterns tied to the academic year, with weekdays seeing increased travel during morning and afternoon peaks.

In recent years, *deep neural networks (DNNs)* have proven to be highly effective in modeling and analyzing daily behavior patterns, owing to their ability to discern complex patterns from vast datasets [12]. The advancement of LLMs has notably transformed many fields and enabled deeper insights into human behavior and trends. An LLM is initially pre-trained on extensive data collections and subsequently adapted to address specific downstream tasks. While initial applications were focused on tasks such as language translation [31] and sentiment analysis [6], the scope of LLMs has broadened to include a variety of fields, including the prediction of human behavior [13].

Leveraging the strong capability of LLMs to learn human behavior patterns, LLM-COD is designed to generate OD datasets across various cities. Figure 1 shows the framework of our approach, composed of the following four steps ①–④:

- ① **OD POI training dataset collection (in City A):** Given the OD vehicle trip dataset from City A, we construct an OD POI training dataset. Each sample in this dataset consists of inputs, which include a list of origin POIs and the associated timestamp, and outputs, which comprise a list of destination POIs and the permissible distance range between the origin and destination.
- ② **Instruction fine-tuning of LLM:** We take a pre-trained LLM (Llama2 + Gemma) and fine-tune the model using the OD training dataset created in ①. ② leverages the general understanding and linguistic capabilities the LLM model has learned from a vast and diverse initial training phase and then narrows its focus to our OD POI generation problem.
- ③ **Origin POI creation and their destination POI prediction (in City B):** Assuming City B possesses the origin POIs

along with their time stamps, the generative LLM model fine-tuned in ② is capable of predicting a distribution of destination POIs for each specified origin POI, as well as the acceptable distance range.

- ④ **Destination cell matching (in City B):** The resulting lists of destination POIs can then be aligned with the most similar location (represented as a grid cell), based on the highest similarity in POI distributions.

3 METHODOLOGY

In this section, we detail the four steps for generating the OD dataset of City B based on City A (Section 3.2–Section 3.5), as outlined in Figure 1. Before that, we first introduce the notations and the definitions in Section 3.1. Table 1 lists the main notations used throughout this paper.

3.1 Notations and Definitions

We use the superscript X to denote the notations specific to City X , where X can be either A or B .

Definition 3.1. (Grid) City X is divided into N^X square cells of equal size (e.g., $500\text{m} \times 500\text{m}$), denoted as $\mathcal{V}^X = \{v_1^X, \dots, v_{N^X}^X\}$, where each v_i^X represents the i -th grid cell in City X . The center coordinates of each grid cell are assumed to be known for both cities, as further discussed in Section 3.2.

Definition 3.2. (Spatial features) We consider K distinct types of POIs, detailed in Table 7 in the Appendix. Each grid cell v_i^X ($i = 1, \dots, N^X$) can be characterized by a spatial feature vector $\mathbf{u}_i^X = [u_{i,1}^X, \dots, u_{i,K}^X]$, where each $u_{i,k}^X$ ($k = 1, \dots, K$) denotes the number of the k -th type of POIs in grid cell v_i^X . We use $\mathcal{U}^X = \{u_1^X, \dots, u_{N^X}^X\}$ to represent the spatial features of the grid cells of City X .

Definition 3.3. (Trip Set) We represent each trip l by a 4-tuple $(v_{O_l}^X, t_l^X, v_{D_l}^X, c_l^X)$, including its origin cell $v_{O_l}^X$, the starting time t_l^X , its destination cell $v_{D_l}^X$, and the acceptable traveling cost c_l^X . We use $\mathcal{T}^X = \left\{ (v_{O_l}^X, t_l^X, v_{D_l}^X, c_l^X) \right\}_{l=1, \dots, L^X}$ to represent the trip set of City X , where L^X is the number of trips in \mathcal{T}^X .

Definition 3.4. (Origin Set) The origin set of city X , denoted as O^X , is defined as the set of origin cells, along with their corresponding starting time points for the trips \mathcal{T}^X in City X , i.e., $O^X = \left\{ (v_{O_l}^X, t_l^X) \right\}_{l=1, \dots, L^X}$.

Definition 3.5. (OD pair learning) Given an origin set O^X , OD pair learning is a process of predicting the destination $\hat{v}_{D_l}^X$ and traveling cost \hat{c}_l^X of each $(v_{O_l}^X, t_l^X) \in O^X$, creating the corresponding trip $(v_{O_l}^X, t_l^X, \hat{v}_{D_l}^X, \hat{c}_l^X)$.

Definition 3.6. (OD flow) The OD flow is the volume of movement between two grid cells. We use \mathcal{T}_{ij}^X to represent the set of trips from grid cell v_i^X to v_j^X

$$\mathcal{T}_{ij}^X = \left\{ (v_{O_l}^X, t_l^X, v_{D_l}^X, c_l^X) \in \mathcal{T}^X \mid v_{O_l}^X = v_i^X, v_{D_l}^X = v_j^X \right\}, \quad (1)$$

Table 1: Main notations and their descriptions

Symbol	Description
\mathcal{V}^X	The set of grid cells of City X
\mathcal{T}^X	The trip set in City X
N^X	The total number of discrete locations in \mathcal{V}^X
O^X	The origin set in City X
K	The total number of POI types
L^X	The total number of trips in City X
v_i^X	The i th grid cell in \mathcal{V}^X
σ^X	The grid cell size of City X
\mathbf{u}_i^X	$\mathbf{u}_i^X = [u_{i,1}^X, \dots, u_{i,K}^X]$ denotes the spatial feature of the cell v_i^X , where each $u_{i,k}^X$ denotes the number of type k POIs within v_i^X .
$\hat{\mathbf{u}}_i^X$	The predicted spatial feature of the cell v_i^X
$v_{O_l}^X$	The origin cell of trip l
$v_{D_l}^X$	The destination cell of trip l
c_l^X	The travel cost of trip l
t_l^X	The starting time of trip l
$f(\cdot)$	The function describing the relationship between origin and destination cells in each trip

and use $f_{ij}^X = |\mathcal{T}_{ij}^X|$ to represent the OD flow from cell v_i^X to v_j^X .

Definition 3.7. (OD matrix) The OD matrix $\mathbf{F}^X = \{f_{ij}^X\}_{N^X \times N^X}$ includes the OD flows between each pair of grid cells in City X .

Problem formulation. The cross-city OD generation problem can be considered as a domain adaptation problem[2], which aims at learning a model from a source data distribution and transferring the knowledge to a different (but related) target data distribution. Given the above notations and definitions (Definition 3.1 – 3.7), We formally define the cross-city OD generation problem in Definition 3.8:

Definition 3.8. (Cross-city OD generation problem)

- ▷ **Instance:** Given the location sets of both cities, $\mathcal{V}^A, \mathcal{V}^B$, their spatial features $\mathcal{U}^A, \mathcal{U}^B$, the trip set of City A , \mathcal{T}^A , and the origin set of City B , O^B ;
- ▷ **Question:** How to predict the OD flow matrix $\hat{\mathbf{F}}^B$ of City B .

Note that given the learned OD pairs (Definition 3.5) at City B , we can count the number of trips between each pairs of cells in City B , therefore obtaining the OD flow matrix (according to Definition 3.6 and Definition 3.7). Next, we detail our approach, LLM-COD, for the OD pair learning, including the four steps outlined in Figure 1.

3.2 Step ① - OD Training Dataset Collection in City A

To construct an *origin-destination POI* dataset for both Cities A and B , we follow a streamlined process encompassing the following three steps, as illustrated in Figure 2:

- (a) **Grid Map Creation:** Given the geographical boundary coordinates (latitude and longitude) of both Cities A and B , we define a rectangular bounding box that encompasses the area of interest. As Figure 2 (Step (a)) shows, we divide this area into square grid cells with size $\sigma^A \times \sigma^A$ (e.g., $\sigma^A = 500\text{m}$,

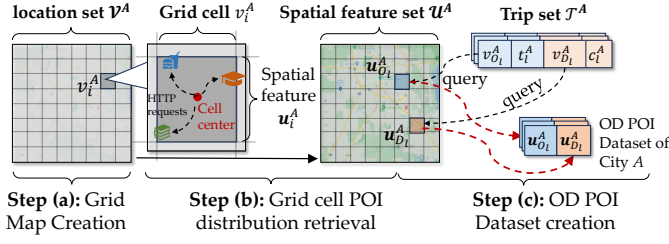


Figure 2: OD training dataset collection.

1000m, and 2000m in our experiments). For simplicity, the cell sizes are kept consistent for both City A and City B, i.e., $\sigma^A = \sigma^B$. In this step, we can obtain the location sets of both cities, \mathcal{V}^A and \mathcal{V}^B .

(b) **Grid cell POI distribution retrieval:** We retrieve the POI distributions for each cell to obtain the spatial feature sets \mathcal{U}^A and \mathcal{U}^B of the two cities. As Figure 2 (Step (b)) shows, for each grid cell v_i^A , we create its spatial feature u_i^A by identifying the POIs whose nearest center point corresponds to the center point of v_i^A . Specifically, in the experiment in Section 4, we use asynchronous HTTP requests to query nearby POIs for each cell's central point via the Map Service API.

(c) **OD POI dataset creation:** We then process the trajectory dataset to obtain the POI distributions of the origin and destination cells of each trip. As Figure 2 (Step (c)) shows, for each trip l in City A, we query POI types for its OD cells $(v_{O_l}^A, v_{D_l}^A)$ from \mathcal{U}^A created in Step (b), obtaining their corresponding POI distributions $u_{O_l}^A$ and $u_{D_l}^A$, respectively. The whole set of OD POI distributions $\{(u_{O_l}^A, u_{D_l}^A)\}_{l=1, \dots, L^A}$ is saved for further analysis and also served as the training dataset for fine-tuning the LLM models.

3.3 Step ② - Instruction-Tuning of LLMs

Our next step is to fine-tune an LLM using instructions to establish the OD pair learning in *textual format*.

However, most existing DNN-based approaches for predicting traffic flows still rely primarily on numerical data, such as road network and node features [28, 37]. As a solution, we guide the LLM with a specific instruction, as detailed in Table 2, which specifies the spatial features of the origin and destination, thereby enhancing the LLM's understanding the input format, and what the responses (or the outputs) are expected.

As depicted in Figure 3, we apply the LLM to a cross city OD flow prediction scenario guided by the instruction format in Table 2. The primary functions of LLM-COD are twofold:

- Constructing input prompt.** Specifically, for each OD trip, given $(u_{O_l}^A, t_l^A)$, we form the input sample as "*<Input> Starting place: $[v_{O_l}^A]$, Starting time: $[t_l^A]$* ".
- Training and inferencing.** For training, after we get "POIs" and "traveling cost", we will combine the two objects together and backpropagate the loss to update the parameters. For

Table 2: Instruction format

Instruction	<i><Instruction> Given the starting place and time of a taxi trajectory in [city], predict the most likely destination and how far it is from the starting point. Please use the provided "Candidate POIs" list to describe the starting place and destination. Candidate POIs: [POI list]</i>
Input Sample	<i><Input> Starting place: [Food & Cuisine, Healthcare, Shopping, Tourist Attraction], Starting time: [12:35]</i>
Response Sample	<i>"POIs": [Residential area, Company, Hotel, Cultural Venue, Shopping], "traveling cost": [1.3 kilometers]</i>
Predicted Cell Index	23

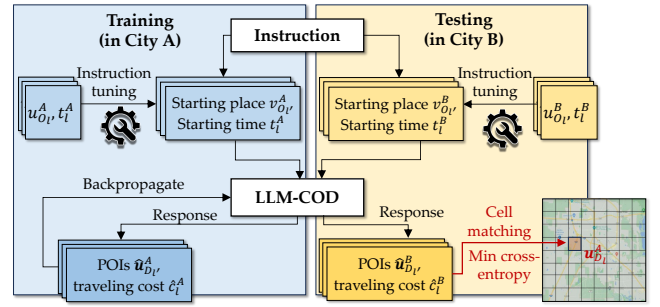


Figure 3: LLM-COD.

inference, we will perform the destination cell matching using the two objects, introduced in step ④.

We use a function f to represent the relationship between each OD pair in the LLM model,

$$\left(\hat{u}_{D_l}^X, \hat{c}_l^X\right) = f\left(u_{O_l}^X, t_l^X; \theta\right). \quad (2)$$

where

- ▷ the **inputs** include the spatial feature of the origin cell $u_{O_l}^X$ and the starting time stamp t_l^X , and
- ▷ the **outputs** include the spatial feature of the destination cell $\hat{u}_{D_l}^X$ and the accepted traveling cost \hat{c}_l^X between the origin cell and destination cell:

Here, θ is the parameters of f . Our goal is to optimize θ , achieved by fine-tuning the Language model.

Low-rank adaptation. We apply *Low-Rank Adaptation (LoRA)* [10], freezing dense layers in LLMs and updating weights with rank decomposition matrices. In LoRA, a hidden layer weight update transformation could be represented as

$$h = \mathbf{W}_0 x + \Delta \mathbf{W} x = \mathbf{W}_0 x + \mathbf{B} \mathbf{A} x, \quad (3)$$

where $\mathbf{W}_0 \in \mathbb{R}^{d \times k}$ is the original weight matrix, and $\mathbf{B} \in \mathbb{R}^{d \times r}$ and $\mathbf{A} \in \mathbb{R}^{r \times k}$ are the low-rank matrices with $r \ll \min(d, k)$. This low-rank adaptation effectively captures the necessary adjustments without the overhead of updating the entire model, as seen in its application to geoscience and other fields.

Loss function. During training, we aim to predict the spatial feature of each destination cell within the target urban area and the acceptable traveling cost to reach the destination cell. As shown in

Figure 3, LLM will output a POI list and a traveling cost. We combine the two objects together and employ *weighted cross-entropy loss* between the actual and predicted outcomes as the loss function, defined by:

$$\text{CrossEntropy} = - \left(\sum_{k=1}^K w_k \bar{u}_{D_l, k}^X \log(w_k \bar{u}_{D_l, k}^X) + \alpha \bar{c}_l^X \log(\alpha \bar{c}_l^X) \right). \quad (4)$$

Here, we use $u_{D_l, k}^X$ and $\hat{u}_{D_l, k}^X$ (c_l^X and \hat{c}_l^X) to denote the actual and forecasted numbers of POI k (actual and forecasted traveling cost) within cell $v_{D_l}^X$, respectively. In Equ. (4), each $u_{D_l, k}^X$, $\hat{u}_{D_l, k}^X$, c_l^X , and \hat{c}_l^X are normalized by the soft max function:

$$\bar{u}_{D_l, k}^X = \frac{e^{u_{D_l, k}^X}}{\sum_{k=1}^K e^{u_{D_l, k}^X} + e^{c_l^X}}, \bar{c}_l^X = \frac{e^{c_l^X}}{\sum_{k=1}^K e^{u_{D_l, k}^X} + e^{c_l^X}}, \quad (5)$$

$$\hat{u}_{D_l, k}^X = \frac{e^{\hat{u}_{D_l, k}^X}}{\sum_{k=1}^K e^{\hat{u}_{D_l, k}^X} + e^{\hat{c}_l^X}}, \hat{c}_l^X = \frac{e^{\hat{c}_l^X}}{\sum_{k=1}^K e^{\hat{u}_{D_l, k}^X} + e^{\hat{c}_l^X}}. \quad (6)$$

Each w_k ($k = 1, \dots, K$) is the weight assigned to the probability $\bar{u}_{D_l, k}^X$ and $\bar{u}_{D_l, k}^X$ and α is the weight assigned to the travel cost \bar{c}_l^X and \bar{c}_l^X .

Next, we use the fine-tuned LLM model to generate a set of OD pairs for City B in Step ③ and Step ④.

3.4 Step ③ - Origin POI Creation and Their Destination POI Prediction in City B

In this step, we first create an origin cell set O_B in City B , incorporating POIs within each cell. O_B can be established using datasets that are mostly available in major cities, such as (1) demographic data, which reflects the density of origin locations across different areas [34], (2) requested location data in LBS, which often serve as the starting points for trajectory analysis [22], (3) location data from large events, stadiums [20], or venues, which includes information on where attendees originate, and (4) public transportation data [35].

As an example, in the experiment part (Section 4), we use Tencent Location Requested (TLR) data between March 1 and March 28, 2024, sourced from Tencent's big data portal (<https://heat.qq.com>). This data captures the number of location requests from Tencent's LBSs, which include services related to social networks, games, online shopping, communication, and travel. Liu et al. [22] found the data exhibits a high correlation coefficient of 0.9 with the residential population, indicating a strong relationship between Tencent's location data and human activity patterns. This strong correlation suggests that Tencent's location-aware dataset can serve as a proxy for human activities on a short-time scale, making it suitable for defining origin cells in any city in China.

Given each generated origin POI distribution $\mathbf{u}_{O_l}^B$ in City B and their corresponding timestamps t_l^B as the inputs, the fine-tuned LLM (described by Equ. (2)) predicts the destination POI distribution $\hat{\mathbf{u}}_{D_l}^B$ along with an acceptable distance range \hat{c}_l^B between the origin and destination cell.

3.5 Step ④ - Destination Cell Matching and OD Matrix Calculation in City B

For every trip $(v_{O_l}^B, t_l^B) \in \mathcal{T}^B$ in City B , we can utilize LLM to predict the POI distribution $\hat{\mathbf{u}}_{D_l}^B$ of the destination cell and the acceptable traveling cost \hat{c}_l^B . To find the matched destination cell, we check each $v_{D_l}^B \in \mathcal{V}^B$ where the travel cost from $v_{O_l}^B$ to $v_{D_l}^B$ is no higher than \hat{c}_l^B .

Among these cells, we select the one whose normalized POI distribution $\bar{\mathbf{u}}_i^B$ exhibits the lowest cross entropy with the normalized predicted POI distribution $\bar{\mathbf{u}}_{D_l}^B$.

$$\hat{v}_{D_l} = \arg \min_{v_{D_l}^B \in \mathcal{V}^B, c(v_{O_l}^B, v_{D_l}^B) \leq \hat{c}_l^B} \text{CrossEntropy}(\bar{\mathbf{u}}_{D_l}^X, \bar{\mathbf{u}}_i^X) \quad (7)$$

Consequently, the set of OD pairs $\left\{ (v_{O_l}^B, \hat{v}_{D_l}^B) \mid v_{O_l}^B \in O_B \right\}$ constitutes the OD dataset generated in City B .

4 PERFORMANCE EVALUATION

In this section, we conduct extensive experiments to test the performance of our approach, LLM-COD, with the emphasis on answering the following 4 questions:

- RQ1: What is the performance of OD flow prediction of LLM-COD in new cities compared to the state of arts (Section 4.2)?
- RQ2: How do the design choices of our method impact performance (Section 4.3)?
- RQ3: Can the LLM-COD model robustly handle the predicting tasks with varying OD flow and trip distance distribution (Section 4.4)?
- RQ4: Are the generated OD flows spatially distributed in a practical way (Section 4.5)?

Before that, we first introduce the experiment settings in Section 4.1.

4.1 Settings

4.1.1 Dataset. Table 3 lists the statistics of the 3 real-world trip data on Beijing, Xi'an, and Chengdu. Beijing dataset [29] contain taxi trajectories; Xi'an and Chengdu datasets contain trajectories by DiDi ride-sharing¹. Table 3 also provides the time span of each trip and the number of grid cells for different grid sizes.

Table 3: Dataset information

Dataset	Beijing	Xi'an	Chengdu
#Trip	3,100,845	2,419,072	3,887,769
Time span	2009/3/2- 2009/3/25	2016/10/1- 2016/10/31	2016/10/1- 2016/10/31
#Grid Cells (1,000m × 1,000m)	250,120	14,001	30,028
#Grid Cells (2,000m × 2,000m)	140,160	8,103	17,745

¹<https://gaia.didichuxing.com/>

Furthermore, Points of Interest (POIs) data were obtained through APIs provided by Tencent map API², utilizing the latitude and longitude of different grid cells. For more detailed POI categories descriptions, please refer to the Appendix.

4.1.2 Baseline Models.

- ▶ **Random Forest (RF)** [21] is a traditional machine learning method that consists of an ensemble of multiple decision tree models. Each tree is built on a subset of the data, and the final prediction is made by aggregating the predictions of all the individual trees.
- ▶ **Gravity Model** [18] is a traditional spatial interaction model which is inspired by Newton's law of Gravitation. The spatial features of one region work as the mass and population flow between two regions follow the power-law distance decay.
- ▶ **Gradient Boosted Regression Trees(GBRT)** [26] is a machine learning method that builds an ensemble of decision trees sequentially, where each new tree is trained to correct the errors made by the previous ones.
- ▶ **GODDAG** [28] is a deep learning method designed to generate origin-destination (OD) flow data for cities where such data is unavailable. It leverages graph neural networks (GNN) to model spatial dependencies and employs domain adversarial training to transfer knowledge from a source city with ample data to a target city with scarce data, thereby improving the model's ability to generalize across different urban environments.

Table 4: Baseline models

Models	Techniques
Random Forest [21]	Tree-based Model
Gravity Model [18]	Physical Model
GBRT [26]	Tree-based Model
GODDAG [28]	Deep Learning

4.1.3 *Evaluation Metric.* We employ three widely-adopted metrics to assess the predictive efficacy of the models as Rong[28] did:

- ▶ **Root mean square error(RMSE)** RMSE is a popular metric applied in regression problems. The formula for RMSE is given by equation 8, where f_{ij}^X are the observed values, \hat{f}_{ij}^X are the predicted values, and $|\mathbf{F}^X|$ is the number of elements in the OD matrix \mathbf{F}^X , i.e., $|\mathbf{F}^X| = N^X \times N^X$.

$$RMSE = \sqrt{\frac{1}{|\mathbf{F}^X|} \sum_{i,j} (f_{ij}^X - \hat{f}_{ij}^X)^2} \quad (8)$$

- ▶ **Symmetric Mean Absolute Percentage Error (SMAPE)** SMAPE is used to measure the accuracy of forecasts. It is a variation of the Mean Absolute Percentage Error (MAPE) that addresses some of the limitations of MAPE, particularly its asymmetry and sensitivity to scale. The value range is in [0,2]. Its formula is given in equation 9 with lower values indicating better forecasting accuracy.

$$SMAPE = \frac{100\%}{|\mathbf{F}^X|} \sum_{i,j} \frac{|f_{ij}^X - \hat{f}_{ij}^X|}{(|f_{ij}^X| + |\hat{f}_{ij}^X|) / 2} \quad (9)$$

²<https://lbs.qq.com/service/webService>

- ▶ **Common Part of Commuters (CPC)** CPC measures the similarity between two data sets by comparing their common parts to the sum of their individual parts. The range is in [0,1], with a value of 1 indicating that the two commuting patterns being compared are identical and all commuters in one pattern are also present in the other pattern. Its formula is given in equation 10.

$$CPC = \frac{2 \sum_{i,j} \min(f_{ij}^X, \hat{f}_{ij}^X)}{\sum_{i,j} \hat{f}_{ij}^X + \sum_{i,j} f_{ij}^X} \quad (10)$$

4.1.4 *Experiment Settings.* The experiments were carried out on a system with 4 NVIDIA A100 GPUs, each with 40GB of memory. For training LLM based models, we used the Adam optimizer with a learning rate of 0.001, while Random Forest, Gravity Model and GBRT used the Adam optimizer with 0.0002 learning rate and weight decay of 0.001. The LLMs used is LLAMA2 7B. The experiments were conducted with a batch size of 100.

4.2 Cross City OD Flow Prediction (RQ1)

To evaluate the performance of our novel model in cross-city prediction tasks, we trained our model on Beijing datasets and performed prediction on Chengdu and Xi'an datasets, which were not seen during the training phase. The results are summarized in Table 5. We have the following observations:

- (1) **Consistent Superiority in Both Cities** Our model consistently outperforms the comparison method in both Chengdu and Xi'an. Notably, our model exhibits significant advantages in terms of RMSE, SMAPE, and CPC metrics, indicating its proficiency in maintaining the distribution of origin-destination (OD) flow in cross-city prediction scenarios. The performance in Chengdu is slightly better than Xi'an. This is attributed to the similarities between Chengdu and Beijing, including urban grid structure and population density. One of the reason that the traditional models and GODDAG underperform may be that they cannot effectively utilize the origin set of the target city because of the structure design the models, resulting in low high SMAPE and low CPC.
- (2) **LLM-CCD Excels in High Precision OD Flow Prediction** LLM-CCD reduces the RMSE of the current best model GODDAG by 46% for grid size of 1,000 m × 1,000m. The ability to predict origin-destination trip flow for high spatial resolution is critical to optimizing transportation systems.

4.3 Performance on Design Choices (RQ2)

To assess the performance of our proposed LLM-COD framework, we conducted experiments focusing on different design choices. Our evaluations utilized the Beijing to Xi'an dataset. The experiments compare the performance of our LLM-COD model under three different settings: the impact of city indicators "[city]" in the instruction format (as described in Table 2), the use of a Single POI strategy, and the impact of distance cost.

- (1) LLM-COD with Full Features: This includes all city indicators and multiple POI types for each grid cell.
- (2) LLM-COD without City Indicators: This version omits city-specific indicators in the instruction format to evaluate their

Table 5: Comparison of models for different cell sizes

Methods	Cell size = 1,000m×1,000m						Cell size = 2,000m×2,000m					
	Beijing → Chengdu			Beijing → Xi'an			Beijing → Chengdu			Beijing → Xi'an		
	RMSE	SMAPE	CPC	RMSE	SMAPE	CPC	RMSE	SMAPE	CPC	RMSE	SMAPE	CPC
Random Forest	25.32	2.00	0.00	8.72	2.00	0.00	304.85	2.00	0.00	444.25	2.00	0.00
Gravity Model	13.08	2.00	0.00	10.56	2.00	0.00	50.51	2.00	0.00	67.8	2.00	0.00
GBRT	29.12	2.00	0.00	9.07	2.00	0.00	268.09	2.00	0.00	446.26	2.00	0.00
GODDAG	4.29	2.00	0.00	3.94	2.00	0.00	26.97	2.00	0.00	28.87	2.00	0.00
LLM-COD	2.29	0.00	0.57	2.12	0.00	0.57	24.25	0.00	0.63	28.49	0.00	0.42

impact on performance. Specifically, the instruction format follows: “<Instruction> Given the starting place and time of a taxi trajectory, predict the most likely destination and how far it is from the starting point.”

- (3) LLM-COD with Single POI Strategy: Each grid cell is represented by at most one POI type. If the cell originally has multiple POIs, the policy to choose is based on social check-in scenarios, prioritizing popular locations where users frequently check in.
- (4) LLM-COD with different α (the weight assigned to the travel cost c_i^X in the loss function (Equ. (4))).

As shown in Table 6, the LLM-COD model with full features achieves an RMSE of 24.25, a SMAPE of 0.00, and a CPC of 0.42 when predicting OD flows from Beijing to Xi’an. When the city indicators are removed, the RMSE increases to 27.96, and the CPC slightly decreases to 0.40, indicating that city-specific features significantly enhance model performance. Furthermore, employing the Single POI strategy results in a higher RMSE of 31.28 and a CPC of 0.40, suggesting that representing each grid cell with multiple POIs provides a richer and more informative feature set, leading to better performance.

As shown in Figure 4, with alpha increasing, the average RMSE first decreases then increases. This trend occurs because taxi drivers often prefer destinations with shorter travel distances. However, when the weight for distance becomes too large, the model tends to overemphasize proximity, consistently choosing the nearest destinations. This approach neglects destinations that are slightly further away but are preferred for their higher quality or better reviews. This behavior is particularly evident in categories such as restaurants and hotels, where quality and reputation often outweigh mere proximity.

Table 6: Performance in Different Model Parameters

Methods	Beijing → Chengdu			Beijing → Xi'an		
	RMSE	SMAPE	CPC	RMSE	SMAPE	CPC
LLM-COD	28.49	0.00	0.63	24.25	0.00	0.42
LLM-COD (without city indicators)	32.96	0.00	0.57	27.96	0.00	0.40
LLM-COD (Single POI)	32.95	0.00	0.57	31.28	0.00	0.40

4.4 Robustness Analysis (RQ3)

To evaluate the robustness of our proposed LLM-COD model, we conducted a detailed analysis focusing on the distribution of OD

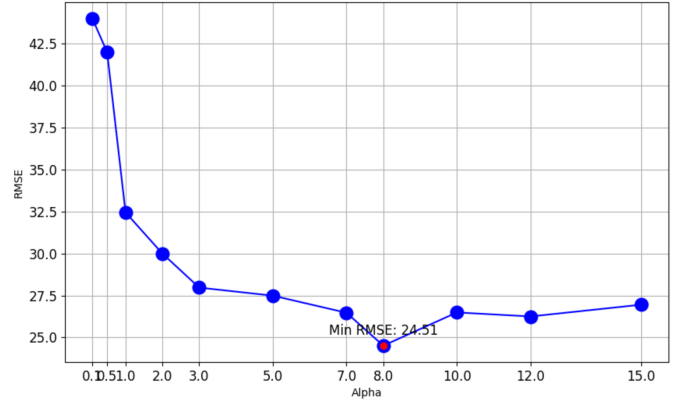
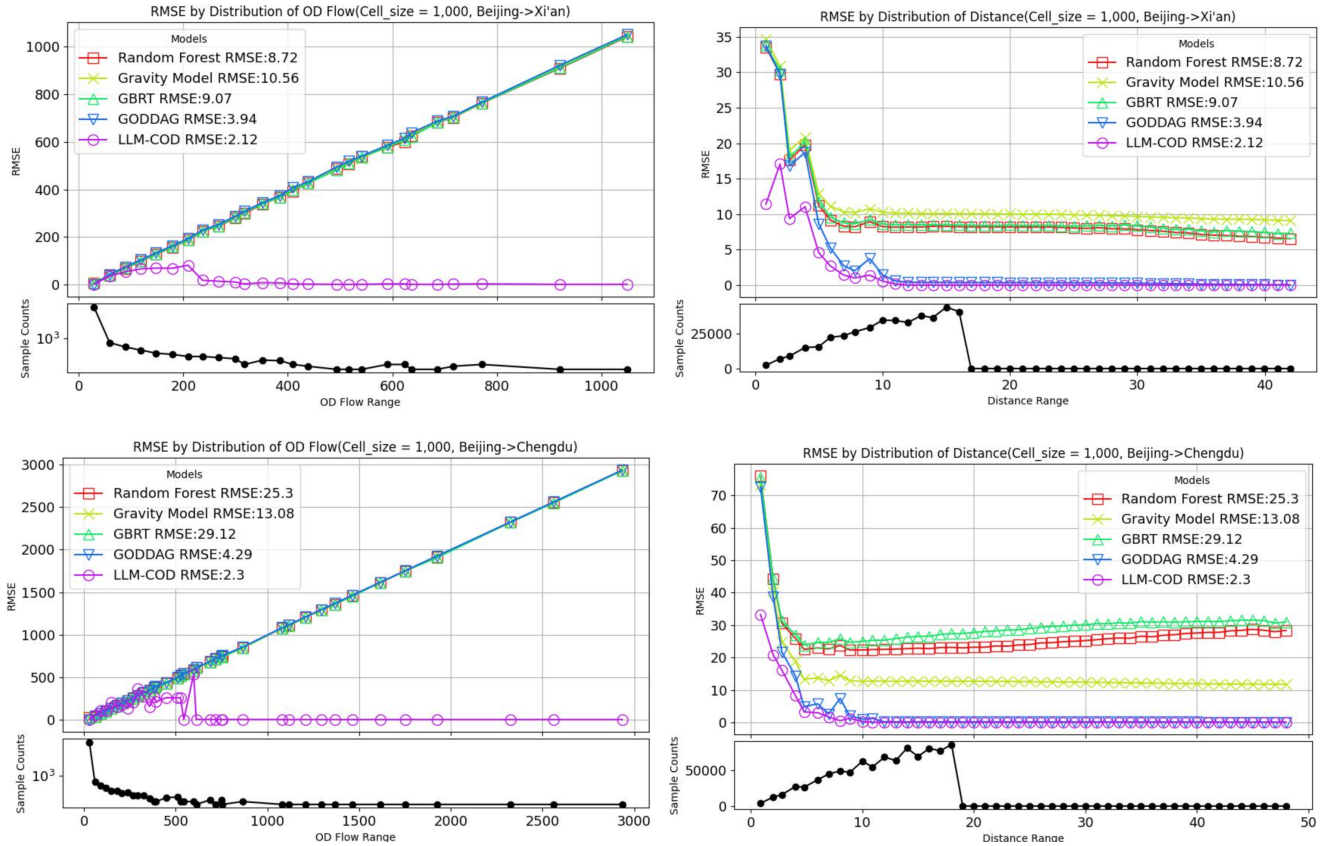


Figure 4: Effects of Alpha

flow and travel distance across the entire test dataset. We specifically examined the RMSE performance as a function of OD flow and distance, across four experimental setups: two different cell sizes (1,000m and 2,000m) and two city pairs (Beijing to Chengdu and Beijing to Xi’an). Figures 5 and 6 present the results of this robustness analysis for cell sizes 1,000m and 2,000m respectively. The key observations from these experiments are as follows:

- (1) The LLM-COD model consistently achieves a RMSE of zero as the OD flow or distance increases. This indicates that our model effectively captures and predicts high-volume and long-distance flows, maintaining accuracy even as these metrics increase. This trend is observed across all eight experiment settings, demonstrating the robustness of our model.
- (2) On the distribution of OD flow, the GODDAG model, along with other baseline models, exhibit large errors and considerable deviation from the actual distribution. These models struggle with both high and low OD flows, almost get the identical error as the flow value.
- (3) On the distribution of distance, GODDAG, represented by blue markers, achieves a RMSE of zero as the distance value increases. But it shows slightly larger errors in the mid-range distances compared to LLM-COD.
- (4) The validity of the LLM-COD model can be attributed to its ability to preserve the underlying distribution of OD flows as found in the ground truth. Unlike other models, LLM-COD leverages comprehensive city indicators and POI data, enabling it to generalize well across different cities, effectively transferring learned mobility patterns from the source city to the target city.

Figure 5: Robustness Analysis ($1,000m \times 1,000m$)

4.5 Visualization (RQ4)

To further demonstrate the effectiveness of our proposed OD generation framework and provide an intuitive comparison, we visualize the generated OD flows from our model and several baselines against the ground truth. For this comparison, we select Xi'an as the target city and Beijing as the source city. The cell size is set to 2,000m. As illustrated in Figure 7, OD flows are depicted using an arc diagram, where each arc represents a flow between two locations. Green marker stands for origin and white marker stands for destination. The brightness of the arcs indicates the volume of OD flows, with blue indicates small volumes, red indicates medium volumes, and yellow indicates the largest volumes of OD flows.

In Figure 7, the ground truth shows that most taxi trips starting from the surrounding areas to the city center. Our LLM-COD model (Figure 7f) closely replicates this spatial distribution, aligning well with the ground truth.

Conversely, the GODDAG model (Figure 7c), mainly shows blue arcs, indicating smaller volumes. However, the distribution of OD flows is incorrect, with the predicted OD flows forming an elliptical shape, suggesting that people have longer commuting ranges in the northwest-southeast direction, which is not true for Xi'an. This leads to a lower RMSE for GODDAG when the distance becomes larger, but it fails to accurately represent the actual commuting patterns. Similarly, the Random Forest (Figure 7b), gravity model (Figure 7c) and GBRT model (Figure 7d), also display elliptical

distributions of OD flows, which is inconsistent with the actual commuting patterns in Xi'an. These models show predominantly red arcs, indicating larger volumes of OD flows but also higher errors. Additionally, it is difficult to discern the true center of the city from the distributions generated by these baseline models.

This visual comparison underscores the robustness and accuracy of our LLM-COD model in replicating the true distribution of OD flows.

5 RELATED WORK

In the domain of OD flow prediction, various approaches have been developed to forecast OD flows using accessible urban data. This field has evolved from *traditional models* (Section 5.1) to advanced *learning-based models* (Section 5.2), adapting to the complexities of urban mobility patterns.

5.1 Traditional Models

Initial research predominantly employed models such as the Gravity Model [18] and the Radiation Model [25]. The Gravity Model posits that the OD flow is proportional to the population sizes of the origin and the destination and inversely proportional to the square of their distance. Recent extensions of this model have incorporated novel data sources such as social media check-ins from X [21], enhancing the model's adaptability to modern urban dynamics. The

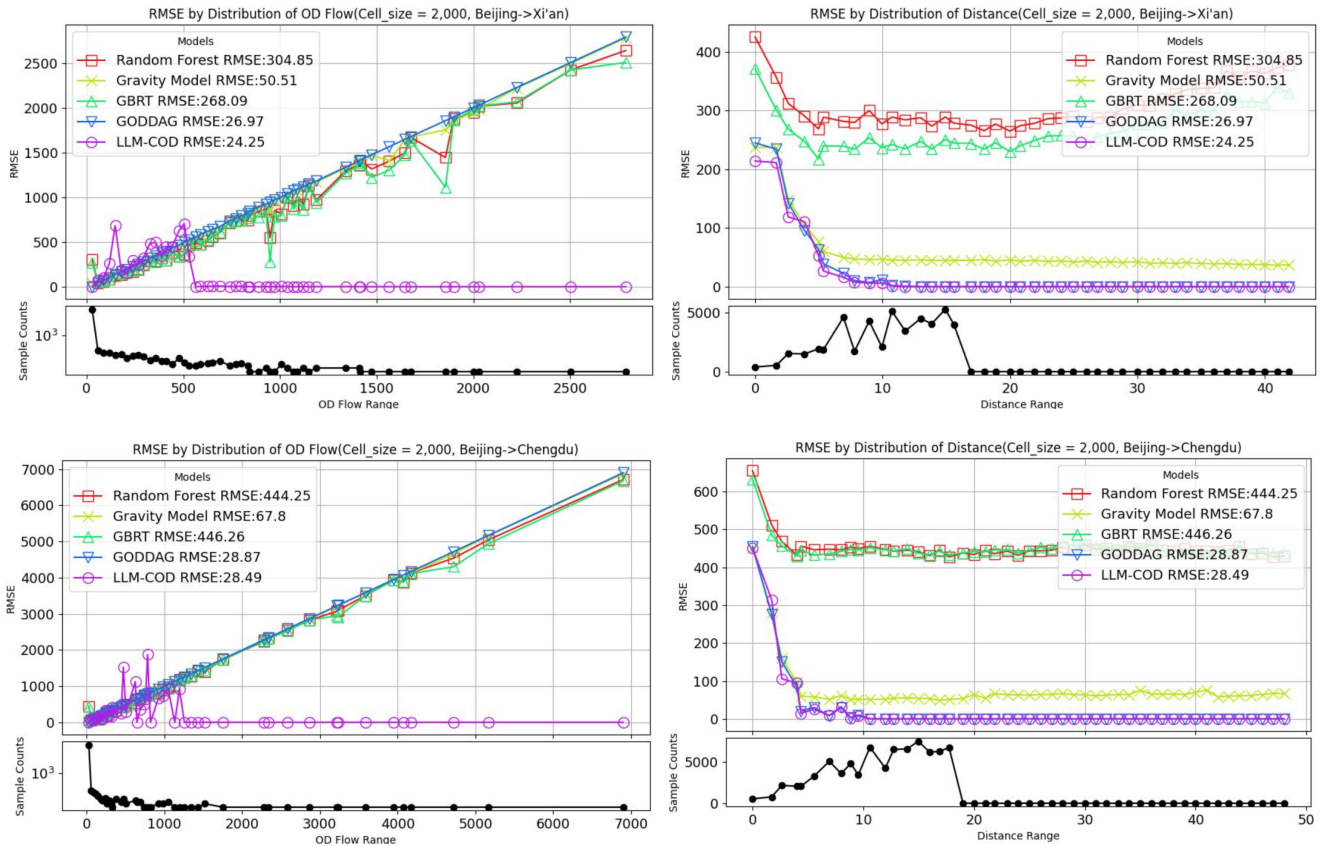


Figure 6: Robustness Analysis ($2,000m \times 2,000m$)

Radiation Model introduces a probabilistic approach [25], considering the attractiveness of intermediate regions, thus providing insights into spatial dynamics. This model has also been refined to include socioeconomic variables, such as income and education, improving its predictive performance [18]. These developments signify a methodological evolution, adapting traditional models to leverage the increasing availability of rich urban data.

Traditional models might fall short in prediction performance, yet they provide essential insights and intuition for building predictive models. Learning-based models, noted for their flexibility and complex structures, deliver commendable performance, especially in capturing implicit data characteristics.

5.2 Learning-Based Models

The advent of machine learning ushered a shift towards data-driven methodologies in OD prediction. Techniques such as Gradient Boosted Regression Trees (GBRT) [26] and Random Forest [21] have demonstrated superior capability in capturing complex interactions within mobility data. Further advancements in deep learning have led to the incorporation of graph-based methodologies, acknowledging the networked nature of urban spaces. Graph Machine Learning Ensemble (GMEL) [19] and Spatial Graph Attention Networks (Spatial GAT) [4] leverage Graph Attention Networks (GAT) [32] and Graph Convolutional Networks (GCN) [14] to model

the intricate spatial relationships essential for accurate OD flow prediction. These approaches significantly enhance the ability to interpret the influence of both direct and neighboring regions on mobility patterns.

Despite these advancements, these works all target predicting OD flows in a single city, while achieving OD prediction across cities remains an unresolved issue. Notably, OD prediction models tailored for a single city [9, 36] often depend on unique urban characteristics, such as transportation modes and urban layouts, making it challenging to apply the learned features from one city to another. The work most related to our works is proposed by Rong et al. [28], which employs unsupervised transfer learning to enhance model transferability. This approach, named GODDAG, involves unsupervised domain adaptation, which trains models within a data-rich source domain and then adapts them to an unlabeled target domain. The central concept is to align features from both the source and target domains into a shared feature space, thereby minimizing the domain discrepancies. Consequently, this method enables the application of OD prediction models, initially trained in one urban environment, to be effectively used in another without the need for labeled data in the target domain.

Leveraging the exceptional pattern recognition and inference capabilities of LLMs, our research introduces a new approach for

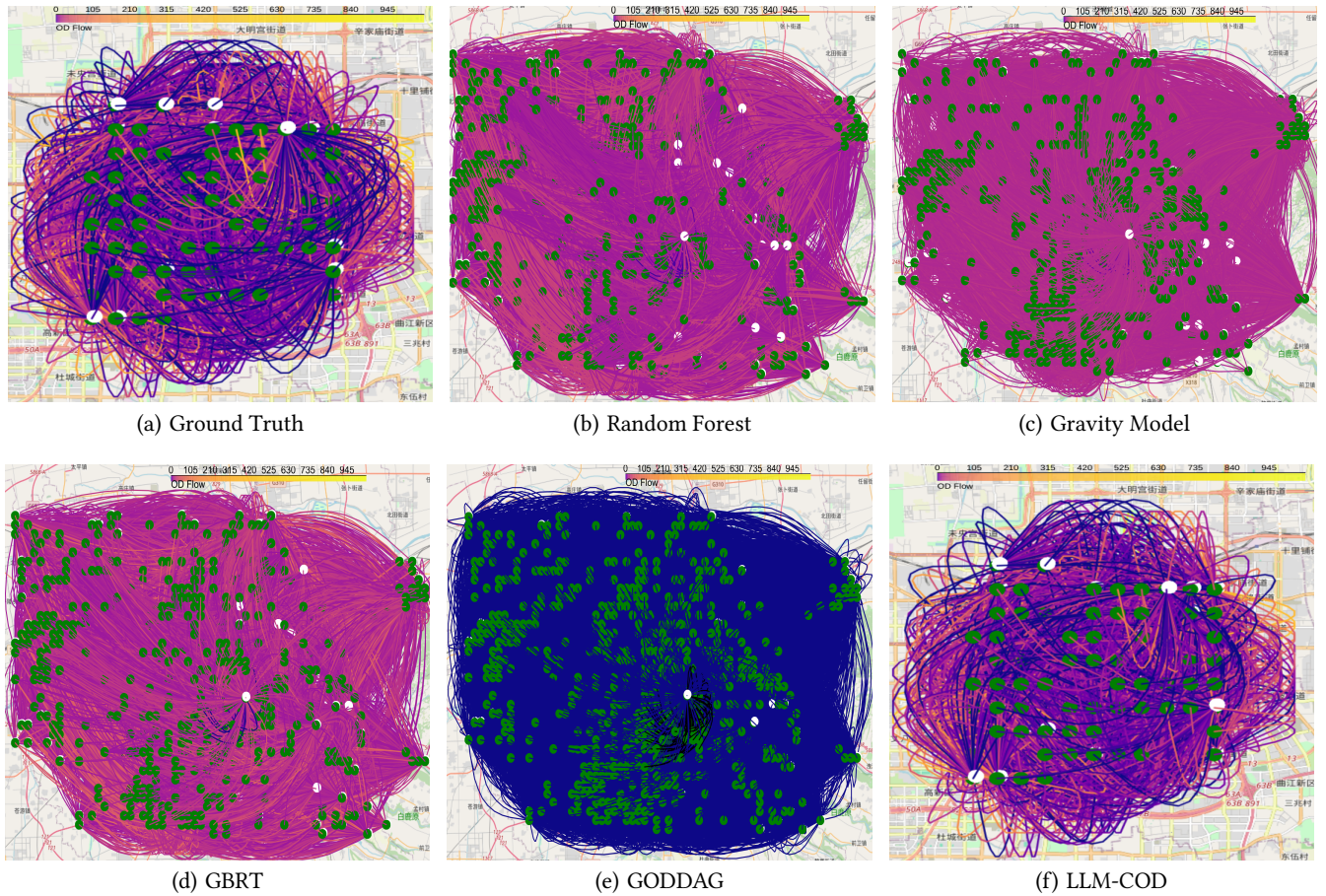


Figure 7: Visualization of the generated OD flow of ground truth and different models under cell size = 2,000m (Beijing -> Xi'an), where green marker stands for origin and white marker stands for destination. The diagrams from left to right are from (a) Ground truth, (b) Random Forest, (c) Gravity Model, (d) GBRT, (e) GODDAG, (f) LLM-COD

predicting OD flows. By combining LLMs with traditional transfer learning techniques, we enhance model generalization across diverse urban environments. This novel method is particularly beneficial for rapidly evolving cities with limited data, providing a new solution to contemporary urban planning challenges.

6 CONCLUSIONS

We introduced LLM-COD which is a new method for cross-city OD flow prediction. By fine-tuning an LLM with a new loss function, our method can understand spatial and functional relationships within urban spaces and capture interactions between individuals and various POIs. Experimental results based on data from three major cities—Beijing, Chengdu, and Xi'an—demonstrate that LLM-COD outperforms state-of-the-art learning-based methods in cross-city OD flow prediction, especially for high-volume and long-distance flows.

We envision several promising research directions to explore further. First, we are committed to enhancing the capabilities of our LLM-COD model by applying to more urban dataset. This will include integrating diverse urban datasets from different cities or

even different countries. Second, Understanding and interpreting the decision-making process of LLM-COD is another crucial area of focus. Although our model achieves high performance, it is essential to provide users with clear and understandable explanations of its predictions. In summary, our future endeavors will focus on expanding the data diversity and contextual understanding of LLM-COD, enhancing its interpretability.

7 ACKNOWLEDGEMENTS

The research was partially sponsored by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-23-2-0014. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes, not withstanding any copyright notation herein. This research was partially supported by U.S. NSF grants CNS-2136948 and CNS-2313866.

REFERENCES

- [1] Shahriar Afandizadeh Zargari, Amirmasoud Memarnejad, and Hamid Mirzahassein. 2021. Hourly origin–destination matrix estimation using intelligent transportation systems data and deep learning. *Sensors* 21, 21 (2021), 7080.
- [2] Shai Ben-David, John Blitzer, Koby Crammer, Alex Kulesza, Fernando Pereira, and Jennifer Wortman Vaughan. 2010. A theory of learning from different domains. *Machine learning* 79 (2010), 151–175.
- [3] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems* 33 (2020), 1877–1901.
- [4] Mingfei Cai, Yanbo Pang, and Yoshihide Sekimoto. 2022. Spatial Attention Based Grid Representation Learning For Predicting Origin–Destination Flow. In *2022 IEEE International Conference on Big Data (Big Data)*. IEEE, 485–494.
- [5] Sofia Cerqueira, Elisabete Arsenio, and Rui Henriques. 2022. Inference of dynamic origin–destination matrices with trip and transfer status from individual smart card data. *European Transport Research Review* 14, 1 (2022), 42.
- [6] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *North American Chapter of the Association for Computational Linguistics*.
- [7] Guido Gentile and Klaus Nökel. 2016. Modelling public transport passenger flows in the era of intelligent transport systems. *Springer Tracts on Transportation and Traffic* 10 (2016), 641.
- [8] Guangzeng Han, Weisi Liu, Xiaolei Huang, and Brian Borsari. 2024. Chain-of-Interaction: Enhancing Large Language Models for Psychiatric Behavior Understanding by Dyadic Contexts. , 392–401 pages. <https://doi.org/10.1109/ICHI61247.2024.00057>
- [9] Mingliang Hou, Feng Xia, Haoran Gao, Xin Chen, and Honglong Chen. 2022. Urban Region Profiling With Spatio-Temporal Graph Neural Networks. *IEEE Transactions on Computational Social Systems* 9, 6 (2022), 1736–1747. <https://doi.org/10.1109/TCSS.2022.3183570>
- [10] Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2021. LoRA: Low-Rank Adaptation of Large Language Models. [arXiv:2106.09685 \[cs.CL\]](https://arxiv.org/abs/2106.09685)
- [11] Liping Huang, Yongjian Yang, Xuehua Zhao, Hepeng Gao, Limin Yu, et al. 2018. Mining the Relationship between Spatial Mobility Patterns and POIs. *Wireless Communications and Mobile Computing* 2018 (2018).
- [12] Md. Ashrafur Islam, Mir Mahathir Mohammad, Sarkar Snigdha Sarathi Das, and Mohammed Eunus Ali. 2022. A survey on deep learning based Point-of-Interest (POI) recommendations. *Neurocomputing* 472 (2022), 306–325. <https://doi.org/10.1016/j.neucom.2021.05.114>
- [13] Dongjun Kang, Joonsuk Park, Yohan Jo, and JinYeong Bak. 2023. From Values to Opinions: Predicting Human Behaviors and Stances Using Value-Injected Large Language Models. [arXiv:2310.17857 \[cs.CL\]](https://arxiv.org/abs/2310.17857)
- [14] Thomas N Kipf and Max Welling. 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907* (2016).
- [15] Kevin S. Kung, Kael Greco, Stanislav Sobolevsky, and Carlo Ratti. 2014. Exploring Universal Patterns in Human Home-Work Commuting from Mobile Phone Data. *PLOS ONE* 9, 6 (06 2014), 1–15. <https://doi.org/10.1371/journal.pone.0096180>
- [16] Zhixin Lai, Jing Wu, Suiyao Chen, Yucheng Zhou, and Naira Hovakimyan. 2024. Residual-based Language Models are Free Boosters for Biomedical Imaging Tasks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. 5086–5096.
- [17] Minjin Lee and Petter Holme. 2015. Relating land use and human intra-city mobility. *PloS one* 10, 10 (2015), e0140152.
- [18] Maxime Lenormand, Aleix Bassolas, and José J Ramasco. 2016. Systematic comparison of trip distribution laws and models. *Journal of Transport Geography* 51 (2016), 158–169.
- [19] Zhicheng Liu, Fabio Miranda, Weiting Xiong, Junyan Yang, Qiao Wang, and Claudio Silva. 2020. Learning geo-contextual embeddings for commuting flow prediction. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 34. 808–816.
- [20] Michael Muchisky and Geoffrey Bingham. 2002. Trajectory forms as a source of information about events. *Perception & psychophysics* 64 (02 2002), 15–31. <https://doi.org/10.3758/BF03194554>
- [21] Nastaran Pourebrahimi, Selima Sultana, Amirreza Niakanlahiji, and Jean-Claude Thill. 2019. Trip distribution modeling with Twitter data. *Computers, Environment and Urban Systems* 77 (2019), 101354.
- [22] Jiale Qian, Zhang Liu, Yunyan Du, Fuyuan Liang, Jiawei Yi, Ting Ma, and Tao Pei. 2021. Quantify city-level dynamic functions across China using social media and POIs data. *Computers, Environment and Urban Systems* 85 (2021), 101552.
- [23] Alec Radford, Karthik Narasimhan, Tim Salimans, Ilya Sutskever, et al. 2018. Improving language understanding by generative pre-training. (2018).
- [24] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. 2019. Language models are unsupervised multitask learners. *OpenAI blog* 1, 8 (2019), 9.
- [25] Yihui Ren, Mária Ercsey-Ravasz, Pu Wang, Marta C González, and Zoltán Toroczkai. 2014. Predicting commuter flows in spatial networks using a radiation model based on temporal ranges. *Nature communications* 5, 1 (2014), 1–9.
- [26] Caleb Robinson and Bistra Dilkina. 2018. A machine learning approach to modeling human migration. In *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies*. 1–8.
- [27] Can Rong, Jingtao Ding, Zhicheng Liu, and Yong Li. 2023. Complexity-aware large scale origin-destination network generation via diffusion model. *arXiv preprint arXiv:2306.04873* (2023).
- [28] Can Rong, Jie Feng, and Jingtao Ding. 2023. GODDAG: generating origin-destination flow for new cities via domain adversarial training. *IEEE Transactions on Knowledge and Data Engineering* (2023).
- [29] Zeyuan Shang, Guoliang Li, and Zhifeng Bao. 2018. DITA: Distributed In-Memory Trajectory Analytics (SIGMOD '18). Association for Computing Machinery, New York, NY, USA, 725–740. <https://doi.org/10.1145/3183713.3183743>
- [30] Wenzhe Sun and Jan-Dirk Schmöcker. 2021. Demand estimation for public transport network planning. In *The Routledge Handbook of Public Transport*. Routledge, 289–305.
- [31] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. 2017. Attention is All you Need. In *Advances in Neural Information Processing Systems*, I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.), Vol. 30. Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fd053c1c4a845aa-Paper.pdf
- [32] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. 2017. Graph attention networks. *arXiv preprint arXiv:1710.10903* (2017).
- [33] Jing Wu, Zhixin Lai, Suiyao Chen, Ran Tao, Pan Zhao, and Naira Hovakimyan. 2024. The New Agronomists: Language Models are Experts in Crop Management. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. 5346–5356.
- [34] Lun Wu, Liu Yang, Zhou Huang, Yaoli Wang, Yanwei Chai, Xia Peng, and Yu Liu. 2019. Inferring demographics from human trajectories and geographical context. *Computers, Environment and Urban Systems* 77 (2019), 101368. <https://doi.org/10.1016/j.compenvurbsys.2019.101368>
- [35] Li Yan, Haiying Shen, Juanjuan Zhao, Chengzhong Xu, Feng Luo, and Chenxi Qiu. 2017. CatCharger: Deploying wireless charging lanes in a metropolitan road network through categorization and clustering of vehicle traffic. In *IEEE INFOCOM 2017 - IEEE Conference on Computer Communications*. 1–9. <https://doi.org/10.1109/INFOCOM.2017.8057019>
- [36] Dapeng Zhang and Feng Xiao. 2023. Dynamic Auto-Structuring Graph Neural Network: A Joint Learning Framework for Origin-Destination Demand Prediction. *IEEE Transactions on Knowledge and Data Engineering* 35, 4 (2023), 3699–3711. <https://doi.org/10.1109/TKDE.2021.3135898>
- [37] Dapeng Zhang and Feng Xiao. 2023. Dynamic Auto-Structuring Graph Neural Network: A Joint Learning Framework for Origin-Destination Demand Prediction. *IEEE Transactions on Knowledge and Data Engineering* 35, 4 (2023), 3699–3711.

A EXPERIMENTAL DETAILS DESCRIPTION

A.1 POI category Details

We collected POI data using tencent map API. The POI data we use contains 24 categories, as table 7 shows.

Table 7: POI Types and Their Frequencies

POI Type	Frequency
Residential Area	105,183
Food & Cuisine	66,158
Commercial Building	46,772
Infrastructure	38,313
Tourist Attraction	34,648
Organization	34,302
Education & School	29,849
Hotel	28,866
Shopping	22,972
Healthcare	18,315
Company & Enterprise	16,377
Industrial Park	13,789
Automobile	11,576
Real Estate Community Affiliated	11,015
Sports & Fitness	9,715
Entertainment & Leisure	9,485
Cultural Venue	8,420
Life Services	5,536
Place Name & Address	4,226
Banking & Finance	2,559
Indoor & Affiliated Facilities	574
Other Real Estate Community	2
Others	1

B LLM BACKEND COMPARISON

When we started the project, Google hadn't released Gemma. On April 16th, Google made Gemma open source. As Gemma is more powerful than LLaMA2, we conducted a comparison between Gemma-7B and the LLaMA2 model we initially used. The table below presents the results of this comparison.

Table 8: Performance in Different LLM backend

Methods	Beijing → Chengdu			Beijing → Xi'an		
	RMSE	SMAPE	CPC	RMSE	SMAPE	CPC
LLM-COD	28.49	0.00	0.63	24.25	0.00	0.42
LLM-COD(Gemma)	21.07	0.00	0.66	19.18	0.00	0.48

As shown in the Table 8, the LLM-COD model using Google Gemma significantly outperforms the one using LLaMA2. Specifically, the RMSE values for the OD flow predictions from Beijing to Chengdu and Xi'an are lower with Gemma, indicating improved accuracy. Similarly, the CPC values are higher, suggesting better alignment with the actual OD flow distribution. The experiments further demonstrate that our framework and new loss function are effective in cross-city OD flow prediction, showing increased performance with a better base LLM model and outperforming state-of-the-arts models further.