

Beyond Imitation: Generating Human Mobility from Context-aware Reasoning with Large Language Models

Chenyang Shao

Department of Electronic Engineering
Tsinghua University
Beijing, China

Fengli Xu*

Department of Electronic Engineering
Tsinghua University
Beijing, China

Bingbing Fan

Department of Electronic Engineering
Tsinghua University
Beijing, China

Jingtao Ding

Department of Electronic Engineering
Tsinghua University
Beijing, China

Yuan Yuan

Department of Electronic Engineering
Tsinghua University
Beijing, China

Meng Wang

School of Computer Science and
Information Engineering
Hefei University of Technology
Hefei, China

Yong Li*

Department of Electronic Engineering
Tsinghua University
Beijing, China

ABSTRACT

Human mobility behaviours are closely linked to various important societal problems such as traffic congestion, energy consumption, and epidemic control. However, collecting mobility data can be prohibitively expensive and involves serious privacy issues, which poses a pressing need for high-quality generative mobility models. Previous efforts focus on learning the behaviour distribution from training samples, and generate new mobility data by sampling the learned distributions. They cannot effectively capture the coherent intentions that drive mobility behavior, leading to low sample efficiency and semantic-awareness. Inspired by the emergent reasoning ability in large language models (LLMs), we propose a radical perspective shift that reformulates mobility generation as a commonsense reasoning problem. In this paper, we design a novel **Mobility Generation as Reasoning** (MobiGeaR) framework that prompts LLM to recursively generate mobility behaviour. Specifically, we design a context-aware chain-of-thoughts prompting technique to align LLMs with context-aware mobility behaviour by few-shot in-context learning. Besides, MobiGeaR employ a *divide-and-coordinate* mechanism to exploit the synergistic effect between LLM reasoning and mechanistic gravity model. It leverages the step-by-step LLM reasoning to recursively generate a temporal template of activity intentions, which are then mapped to physical locations with a mechanistic gravity model. Experiments on two real-world datasets show MobiGeaR achieves state-of-the-art performance across all metrics, and substantially reduces the size of training samples at the same time. Besides, MobiGeaR also significantly improves the semantic-awareness of mobility generation by improving the intention accuracy by 62.23% and the generated mobility data is proven effective in boosting the performance of downstream applications. The implementation of our approach is available: <https://anonymous.4open.science/r/MobiGeaR-C57C>

1 INTRODUCTION

Human mobility data often reveals the nuanced details of human’s spatio-temporal interaction with the physical environment [13]. Such datasets play crucial roles in numerous important applications, such as urban planning [31], epidemic control [7] and business site selection [18]. Traditionally, large-scale human mobility data is collected through household travel survey [26] or ubiquitous personal sensors like smartphone [37], which can be hugely expensive [17] and are associated with serious privacy risks [10, 32]. Therefore, it has been a long-standing research quest to design low-cost yet high-quality generative models for human mobility data [11, 25], unleashing its full potential in wide range of applications.

Previous efforts in designing generative human mobility models can be broadly classified into two categories — *mechanistic* and *deep learning*. On one hand, the *mechanistic* models leverage stochastic processes to characterize the probabilities of human mobility behaviours [25], such as the power law distribution of moving distance [6]. However, they are often oversimplified and only focus on certain aspects of mobility behaviour, failing to generate nuanced, accurate and comprehensive mobility data [11]. On the other hand, the *deep learning* models exploit the expressive representation power of deep neural networks to learn the probability distributions of human mobility behaviour from training samples, and generate new mobility data by sampling from the learned distributions [20]. Recent studies showed the state-of-the-art deep generative frameworks such as GAN [27, 34, 35], VAE [19] and diffusion model [38] can be adequately adapted for mobility generation. However, the *deep learning* models tend to overlook the inner mechanisms of mobility behaviour, *e.g.*, humans usually have coherent intentions in consecutive movements. As a result, the *deep learning* models have low sample efficiency to learn the underlying behaviour distributions and the generated mobility data lacks of semantic-aware intentions. Therefore, it remains a challenging research problem to design accurate, sample-efficient and semantic-aware generative models for human mobility.

*Fengli Xu and *Yong Li are the corresponding authors.

We draw inspirations from the recent revolutionary development of large language models (LLMs). The scaled-up versions of LLMs such as GPT [12] and PaLM [4] have shown emergent abilities for general purpose reasoning [28]. For example, recent study showed chain-of-thought prompting can harness LLM's power for commonsense reasoning [29]. Besides, LLM also exhibits role play ability that allows it to simulate residents of virtual village [21]. Therefore, the recent advancement of LLMs present a unique opportunity to model the inner mechanisms of human mobility as a commonsense reasoning problem, which has great potential to drastically improve the sample efficiency and semantic-awareness of generative mobility models.

To harness the reasoning power of LLMs, we propose a novel **Mobility Generation as Reasoning (MobiGear)** framework. It replaces the behaviour distribution imitation frameworks of the classic deep learning models by reformulating the mobility generation problems as a commonsense reasoning problem. Specifically, *MobiGear* recursively prompts LLM to generate next visited locations by reasoning the plausible behaviour given the individual's demographic profiles, current location and time. Such reasoning process allows *MobiGear* to tap into the commonsense reasoning and role play capabilities of LLMs to emulate the inner mechanisms of human mobility behaviour, alleviating the need of large training samples in classic deep learning models. Besides, we also design a context-aware chain-of-thought prompting technique to improve the reasoning accuracy. It leverages a few human annotated travel logs, detailing the intentions of each movement, to unlock the few-shot in-context learning capability of LLM. It allows LLM to emulate human reasoners to generate coherent intentions for consecutive movements, which can effectively model the semantic meaning of the location types associated with each movement. Finally, to reduce the potential LLM token cost, we design a novel *divide-and-coordinate* mechanism to harness the synergistic effect of LLM and *mechanistic* mobility model. Specifically, *MobiGear* is divided into two stages. The first stage solely focuses on the reasoning of high-level activity intentions. It prompts LLMs to generate temporal templates of activity intentions, such as *go to work* and *have a dinner* and their associated time periods. The second stage consequently uses a mechanistic gravity model to map the intention templates to physical locations with corresponding functions, *i.e.*, *company* and *restaurant*. The gravity model captures the movement probability based on the distance and location popularity. One intention template can be used to generate multiple mobility trajectories with different initial locations and the stochastic nature of gravity model. Therefore, *divide-and-coordinate* mechanism can substantially reduce the token cost of generating a mobility trajectory compared to pure LLM solution.

Empirically, we evaluate *MobiGear* on two mobility datasets collected from mobile applications and cellular operators. We find *MobiGear* achieves state-of-the-art performance across all metrics on both datasets. More importantly, it drastically reduces the size of training samples from 100,000 to 200 trajectories compared to deep learning baselines. Specifically, *MobiGear* uses only 8 mobility trajectories for few-shot in-context learning in first stage of LLM reasoning, and 200 trajectories for fitting the gravity model in second stage. Besides, *MobiGear* achieves superior performance in generating semantic-aware mobility data, outperforming second

best baseline by 62.23% in movement intention inference. Furthermore, experiments show *divide-and-coordinate* mechanism substantially reduces the token cost from 45592 to 300 tokens per trajectory. Finally, the generated mobility is proven effective in boosting the accuracy of mobility prediction model as data augmentation.

To summary, the contributions of our work can be summarized as follows:

- To the best of our knowledge, we are the first to formulate mobility generation as a commonsense reasoning problem with LLMs. This drastically shifted perspective unleashes LLM's reasoning power for semantic-aware and sample efficient human mobility generation.
- We propose a novel *MobiGear* model that uses context-aware chain-of-thought prompting to adapt LLM for mobility generation. Besides, it use *divide-and-coordinate* mechanism to harness the synergistic effect between LLM reasoning and mechanistic gravity model.
- Extensive experiments on real-world mobility data show *MobiGear* achieves state-of-the-art performance, even though it substantially reduces the size of training samples by several magnitudes.

2 PRELIMINARIES

2.1 Problem Definition

Mobility Behaviour. A piece of mobility behavior data consists of a sequence of points coherent in time, denoted by $x = \{s_1, s_2, \dots, s_n\}$, where each point s_i can be expressed as (t_i, l_i, e_i) . l_i denotes the spatial position, either in the form of region ID or coordinate of latitude and longitude. t_i represents the time of the visit to l_i . e_i represents the type of intention to come to this place. From the micro level, there is a close relationship between people's identity information and people's mobility behaviour. Occupation, education level, consumption level, etc. all affect people's travel patterns. For example, the mobility pattern of an IT professional is likely to be a simple "point-to-point" commute between home and the office while a delivery boy's day may involve commuting between numerous locations in the city.

Mobility Behaviour Generation. Given a real-world mobility behaviour dataset $\mathcal{X} = \{x^1, x^2, \dots, x^N\}$, where each $x^i = \{s_1^i, s_2^i, \dots, s_n^i\}$ is a piece of mobility behaviour sequence. The goal of the mission is to generate new mobility behaviour data $\mathcal{Y} = \{y^1, y^2, \dots, y^N\}$ such that the generated data exhibits similar characteristics and mobility patterns as the original real data, while also providing practical value to support downstream applications.

2.2 Background

Existing generative models for mobility behavior. Existing methods for generating movement trajectories can be broadly categorized into two main types: model-based and model-free. Model-based approaches assume that human travel habits and movement patterns can be described and modeled using a finite set of parameters with clear physical meanings. $\mathcal{G} = f_{p_1, p_2, \dots, p_n}(\cdot)$ The generation process often exhibits Markovian properties, allowing for step-by-step probability sampling: $p_{i+1} = G(p_i)$ One of the most representative works in this category is TimeGeo[17], which constructs travel patterns anchored around the home by introducing

parameters such as weekly home-based tour number, dwell rate, and burst rate. However, model-free models are often too simple to accurately model real trajectories, as real human travel patterns may be time-varying and high-order. On the other hand, model-free methods do not make any assumptions about the patterns in advance, but instead rely on deep neural networks to directly learn spatiotemporal patterns and fit data distributions from real data. Assume that the real data x_0, x_1, \dots, x_n satisfies an unknown data distribution $p_{data}(x)$ and the data generated by the deep generation model satisfy a distribution p_θ . In order to get p_θ as close to p_{data} as possible, the goal of optimization can be summarized as minimizing the expected value of the negative log-likelihood function:

$$\arg \min_{\theta} \mathbb{E}_{x \sim p_{data}(x)} [-\log p_\theta(x)] \quad (1)$$

3 MOBIGEAR FRAMEWORK

In this section, we elaborate on the framework of MobiGeaR. Our framework has two progressive stages: **Generating Intention Templates with LLM Reasoning** and **Mapping to Physical Locations with Mechanistic Model**. The whole framework is shown in Figure 1.

3.1 Generating Intention Templates with LLM Reasoning

The task at this stage is to generate a daily behaviour template that includes intentions and corresponding times, given the initial context. (E.g., [“go to sleep”, “(00:00, 08:33)”, [“go to work”, “(09:47, 17:49)”, [“eat”, “(18:45, 19:49)”, [“do shopping”, “(20:01, 20:35)”, [“go home”, “(21:40, 23:59)”]]])

To ensure that the generated events are logically arranged in time and interconnected, (For example, If a person spends long hours at work from morning till evening, it is likely that he will engage in sports or leisure activities after finishing work, such as watching a movie.) we introduce the **COT** mechanism. We don’t require the LLM to generate the complete intention template at once. Instead, we decompose the entire generation into several generation steps of individual intentions. At each generation step, we provide the preceding generation results as the basis for reasoning. Through this explicit construction of a Markov decision chain, we aim to enhance the rationality and accuracy of LLM inference.

As part of the initial context, the persona **profile** information will be provided to the LLM in a specified role play form before the generation begins and will continue to accompany all generation steps. Each profile includes occupation, gender, income level, education level. To more explicitly capture how persona attributes affect mobility patterns, we provide 8 examples. These examples include corresponding character attributes and time-event templates, all extracted from real datasets. At each time-event point of the examples provided, we inject the COT hint. Concretely, we manually annotate the thinking process behind each intention, find its connection to the prior intentions, and explain the reason for the decision. When asking, inform the LLM to think based on the preceding generation results and make decisions like the examples provided. Moreover, to ensure that the generated collective trajectories can reflect realistic city-level population mobility patterns,

we sample from the population attribute distribution of real-world dataset for individual role play, hoping that a large number of microscopic individual modellings can ultimately exhibit macroscopic regularities.

The profile serves as a static context, influencing the overall travel habits and patterns of the LLM at a macro level. To dynamically assist each generation step, we further incorporate time and date information into reasoning to improve the alignment between the generated results and time. For example, at 10 p.m., most people won’t choose to go on excursions, but rather for home. Based on our **context-aware chain of thought** mechanism, we construct a recursive generation flow. Given the context and preceding results provided in each generation step, LLM decides what to do next and the template keeps expanding until the end of time. The complete schematic diagram of our **context-aware COT** generation flow is shown in Figure 2.

Besides, we compile the distribution of the number of daily intentions from a small amount of real data (200 7-day pieces), from which we would sample as a limit on the total number of intentions to be generated before the generation process starts. Still with these data, we also count the durations distribution of each intent type. In this way, once the LLM decides the next intention during the generation step, we sample from the duration distribution of this intention and calculate the start and end times to generate the complete time-intention point. The total of 10 intentions considered in our experiment are as follows: 1. go to work 2. go home 3. eat 4. do shopping 5. do sports 6. excursion 7. leisure or entertainment 8. sleep 9. medical treatment, 10. handle the trivialities of life. In fact, the granularity of intention division can be freely controlled according to semantic requirements.

3.2 Mapping to Physical Locations with Mechanistic Model

In the previous stage, templates containing time and intentions are generated, while the selection of locations will be carried out in this stage. The home and work places can be sampled directly from addresses identified by real data, but how to correspond an abstract intention to a specific one in millions of POI locations is a challenge.

Unlike many previous approaches to location prediction and generation using deep learning methods, we decide to use a simple mechanistic model to handle this. Through the bold assumption and simple modeling of selection strategy, the transition probability between locations can be easily obtained. Considering the performance as well as implementation costs, we choose the Gravity model.

The Gravity model emphasizes the importance of distance in human migration. It considers that the magnitude of migrating traffic between two regions can be reduced to such a simple form:

$$P_{i,j} \propto \frac{P_i P_j}{r_{ij}} \quad (2)$$

where P_i and P_j donate the population of region i and region j , respectively. And r_{ij} donates the distance between region i and region j . Taking the nonlinear factors into account, it can be further expressed as:

$$P_{i,j} = K m_i m_j f(r_{ij}) \quad (3)$$

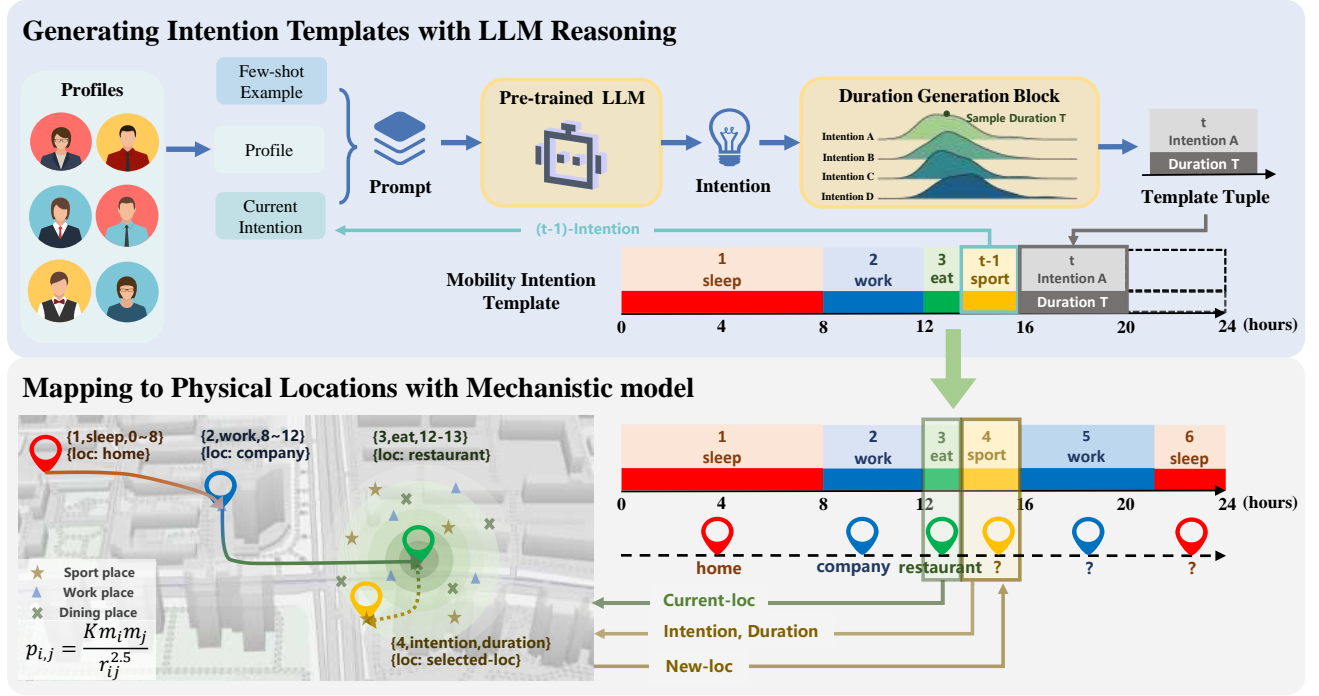


Figure 1: The overview of the proposed MobiGear framework.

where m_i is the function of P_i and is typically expressed in the form of P_i^α , as is m_j . $f(r_{ij})$ represents a decreasing function of distance. All of these functions can be fitted. In our experiment, we replaced the population num with the density of POIs. We believe that the higher the density of POIs, the more attractive they are to individuals. With the previous location as the center of the circle, the surrounding map is partitioned into concentric rings at 1km intervals. For all POIs in one ring, m_j is set to the POI density in the ring (number/ring area), and $f(r_{ij})$ is set to $r_{ij}^{-2.5}$ which can be fitted from a small amount of data. We assume that the great majority of people travel an upper distance limit of 10km, so the POI search is limited to circles within a radius of 10km.

3.3 Generation Process

Corresponding to the divide-and-coordinate framework design, the generation process is also divided into two collaborative parts.

- *Generate intention templates*: assign a role play identity to LLM, provide examples and initial time information, generate intention-time templates step-by-step with LLM reasoning.
- *Map each intention to physical location*: Before mapping a template to a specific trajectory, home and work locations are randomly sampled. For each intention in the template, Gravity model gives a choice of specific locations. Specifically, when the template switches from (t_i, l_i, e_i) to (t_{i+1}, e_{i+1}) , the mechanism model will select the next physical location by probabilistic sampling based on l_i .

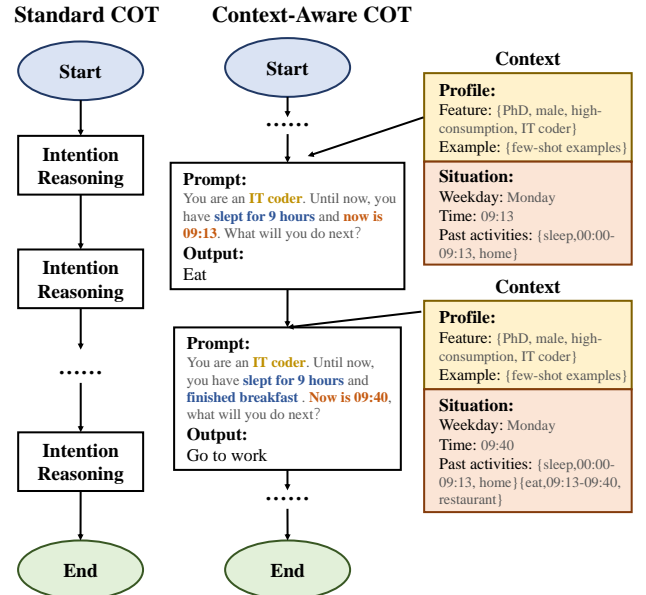


Figure 2: Generation process based on Context-Aware COT

4 EXPERIMENTS

In this section, we carry out experiments to verify the validity of our proposed model as well as to answer several research questions below:

- **RQ1:** How does the proposed model perform compared with various current models? Is it still possible to achieve similar performance without using deep learning generative models to fit distributions to large amounts of data?
- **RQ2:** Whether all the mechanisms designed during the model construction are effective and necessary?
- **RQ3:** Whether the cost of using LLMs is acceptable? How much does our synergistic approach contribute to cost reduction?

4.1 Experimental Setup

4.1.1 Datasets. Our experiments are conducted on two real-world mobility datasets collected from two data sources, Tencent and ChinaMobile, covering the geographical area of Beijing, China. The basic information of these datasets is presented in Appendix A.

4.1.2 Baselines. In order to objectively assess the performance of our model, we compare it to several of the most classical and recent baseline models. The baselines can be categorized into 2 groups, simple mechanism models: TimeGeo[17], and deep generative models: MoveSim[11], Volunteer[19] and DiffTraj[38]. Detailed information about baselines is presented in Appendix B:

4.2 Evaluation Methods

In order to make an comprehensive evaluation of the generated data, we evaluate the data quality from three dimensions: statistical evaluation, semantic evaluation and aggregation evaluation.

4.2.1 Statistical Evaluation. For the statistical metrics, we calculate the distance of the distribution of each statistic between the generated data and the real data with Jensen–Shannon divergence(JSD). The lower the JSD, the closer the two distributions are. Specifically, we will quantitatively calculate the JSDs on the following 4 metrics: **Radius**, **DailyLoc**, **IntentDist**, **G-rank**. We refer the readers to Appendix C for more details of these statistical metrics.

4.2.2 Semantic Evaluation. In previous work, assessments have been limited to the statistical level, while here we add evaluation on the semantic dimension by means of two metrics: intention accuracy(IntentAcc) and the distribution of intention categories(IntentType). One preparatory work before the semantic evaluation is needed: match intention categories for trajectory records. For the real data, we get the intention category information of each location visit record by manual labeling or POI matching. For the results we generate, they naturally contain the category distinction. For the baseline model without the concept of category, after we identify the stay at home and working from the trajectories, other intentions are randomly assigned.

Then, for the former metric, we need to calculate the similarity by direct comparison between the generated intention sequences and the real ones. Since real data inevitably possesses uncertainty and randomness, we aggregate 7 days of a person’s trajectories into 1 day by majority voting on the same time slice. Of course, we only focus on the level of intention type when aggregating. For example, if 5 of these 7 days are eating at 12:00, and 2 days are working at 12:00, then the intention of 12:00 in the aggregated result is eating. With the help of aggregation, the uncertainty of the trajectory is eliminated and a more stable and representative real

trajectory is obtained. Then the generated results are compared with the aggregated trajectory on a time-slice-by-time-slice basis. The length of the time slices is standardized to half an hour. For the latter metric, we calculate the time proportion of each category of intentions in the trajectories weighted by durations, and calculate the JSD of the proportion vectors.

4.2.3 Aggregation Evaluation. We also evaluate the authenticity of the spatial distribution of the data from an aggregated perspective. We first assess the fidelity of the generated data by two metrics, **LocFreq** and **ODSim**. The first metric is to partition the map uniformly into grids, count the frequency of trajectories visiting each grid, and calculate JSD between the frequency vectors. The second metric counts the movement of individuals between grids, calculates the OD transfer probability matrix, and computes the MSE values between the matrices. In addition to the numerical measurement, we also visualize the grid-visiting frequency difference between the generated data and real data using heatmaps, such that we can intuitively judge the quality of the generated data through the visual difference between our model and baselines.

4.3 Overall Performance(RQ1)

Tables 1 and 2 show the full results of our evaluation of the three dimensions on the Tencent and ChinaMobile datasets. It is worth mentioning that although the original dataset has a huge amount of data (Tencent has 10w user trajectories), only **200** pieces of trajectory data are used in our model, while the other baselines use all.

4.3.1 Statistical Performance. We achieved the best results in all four statistical metrics. With a CNN structure in the denoising model, DiffTraj can well capture the local and global geographic information and thus it performs well in Radius, but its continuous generation based on latitude and longitude makes it difficult to capture the patterns of discrete intentions. The trajectory VAE in the Volunteer model learns the distributions of travel time and dwell time accurately, and such an approach is fruitful in terms of capturing the patterns of discrete intentions, but does not perform well in the spatial dimension. MoveSim’s data format is discrete in both time and space, which has advantages in distinguishing different intentions. But its performance on different datasets is not consistent. As for TimeGeo, its small number of parameters may not allow for accurate modeling of complex movement patterns, but it has a relatively good performance on ChinaMobile dataset, showing the robustness on different datasets. Our excellent performance on statistical metrics proves that there are other ways to maintain the fidelity of trajectories besides fitting to large amounts of real data.

4.3.2 Semantic Performance. At the semantic level, we achieve a far superior position. In the assessment of IntentAcc, TimeGeo outperforms other deep generative models, suggesting that its explicit distinction between home and other places is effective. DiffTraj does better in the distribution of intention types, showing that DiffTraj has done a better job of capturing the simple cycle movement patterns between home and work, since the types of the other intentions are randomly assigned.

| Model | Statistical | | | | Semantic | | Aggregated | |
|----------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|-----------------|
| | Radius ↓ | DailyLoc ↓ | IntentDist ↓ | G-rank ↓ | IntentAcc ↑ | IntentType ↓ | LocFreq ↓ | ODSim ↓ |
| TimeGeo | 0.2592 | 0.2513 | 0.2040 | <u>0.0176</u> | <u>0.4639</u> | 0.1545 | 0.6931 | 7.38E-05 |
| MoveSim | 0.2235 | <u>0.0521</u> | <u>0.1010</u> | 0.0244 | 0.0956 | 0.1781 | 0.6384 | <u>5.93E-05</u> |
| VOLUNTEER | 0.5116 | 0.0560 | 0.3296 | 0.0213 | 0.1906 | 0.1620 | 0.2956 | 6.00E-05 |
| DiffTraj | <u>0.0284</u> | 0.6931 | 0.6931 | 0.0286 | 0.4035 | <u>0.0804</u> | <u>0.2872</u> | 6.38E-05 |
| MobiGear(Ours) | 0.0245 | 0.0259 | 0.0158 | 0.0046 | 0.7526 | 0.0334 | 0.2820 | 5.45E-05 |

Table 1: Performance comparison between our model and baselines on Tencent dataset. ↓ means lower is better. Bold denotes the best results and underline denotes the second-best results.

| Model | Statistical | | | | Semantic | | Aggregated | |
|----------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|-----------------|
| | Radius ↓ | DailyLoc ↓ | IntentDist ↓ | G-rank ↓ | IntentAcc ↑ | IntentType ↓ | LocFreq ↓ | ODSim ↓ |
| TimeGeo | 0.2617 | 0.2760 | <u>0.2204</u> | 0.0178 | <u>0.4558</u> | 0.1550 | 0.6388 | 8.40E-05 |
| MoveSim | 0.5374 | 0.6841 | 0.6521 | 0.0439 | 0.1215 | 0.2779 | 0.4538 | 6.59E-05 |
| VOLUNTEER | 0.4729 | 0.6510 | 0.6389 | 0.0266 | 0.1370 | 0.2203 | 0.4364 | 7.00E-05 |
| DiffTraj | <u>0.0245</u> | 0.6935 | 0.6931 | <u>0.0147</u> | 0.4463 | <u>0.0600</u> | <u>0.2943</u> | <u>5.38E-05</u> |
| MobiGear(Ours) | 0.0229 | 0.0884 | 0.0659 | 0.0035 | 0.7283 | 0.0345 | 0.2527 | 5.15E-05 |

Table 2: Performance comparison between our model and baselines on ChinaMobile dataset.



Figure 3: The difference between the frequency distribution of the generated data and the real data on the map grid, with brighter grid points indicating a larger difference.

4.3.3 Aggregation Evaluation. In this part, we also achieve the best results, indicating that the probabilistic strategy of Gravity model aligns with most people’s location selection habits. On LocFreq, DiffTraj’s performance is very similar to ours, showing strong ability of fitting spatial distribution. Volunteer’s performance is also close on Tencent dataset, reflecting the validity of the User VAE in modeling an individual’s home and work place. On ODSim, MoveSim and DiffTraj are sub-optimal on two datasets respectively. In addition to the computation of the metrics, we also use heatmaps to visualize the grid-visiting frequency difference between the generated data and real data. We restrict the visualization to a square area connected to the fifth ring road of Beijing and compare the results of our model with baselines. As shown in

Figure 3, the color depth of each grid point represents the difference in the visiting frequency compared to the real data. The closer to purple, the larger the error. The trajectories generated by our model are the closest to the real trajectories as there are very few purple or red grid color blocks on our graph. The results of DiffTraj and VAE are visually similar to ours, except that some local areas have slightly darker colors. But the images of MoveSim and TimeGeo show overall darker colors, indicating that their errors are larger.

4.4 Ablation study(RQ2)

In order to test whether the context-aware COT reasoning that we introduce in LLM prompts is effective, We remove each subcomponent from the model and then test the performance. Table3 shows the results.

As can be seen from the table, removing the profile had the greatest impact on the results. At the intention level, the accuracy of daily intentions is reduced by 40% and the JSD of the intention type distribution worsened significantly. At the level of statistical metrics, all metrics except IntentDist showed significant deterioration. DailyLoc’s JSD value went up by an order of magnitude. Removing the time has the greatest impact on Dailyloc and G-rank, suggesting that without time, it is difficult to correctly make gradual planning. The semantic level also has a relatively obvious deterioration, further emphasizing the importance of time. In contrast, removing COT has the least impact on the results, and the accuracy of intents at the semantic level is even improved, but the fitting degree of intent type distribution is decreased. The possible explanation is that after the absence of COT, more consideration is given to the correlation between time and intents, while less consideration is given to the correlation with intentions generated before and the regularity of intentions themselves. The deterioration of Radius,

| Model | Statistical | | | | Semantic | | Aggregated | |
|-----------------|-------------|------------|--------------|----------|-------------|--------------|------------|----------|
| | Radius ↓ | DailyLoc ↓ | IntentDist ↓ | G-rank ↓ | IntentAcc ↑ | IntentType ↓ | LocFreq ↓ | ODSim ↓ |
| MobiGeaR (full) | 0.0245 | 0.0259 | 0.0687 | 0.0046 | 0.7526 | 0.0334 | 0.2820 | 5.45E-05 |
| No-Profile | 0.0320 | 0.2027 | 0.0680 | 0.0108 | 0.4516 | 0.1085 | 0.2812 | 5.49E-05 |
| No-Time | 0.0286 | 0.0794 | 0.0618 | 0.0224 | 0.6758 | 0.0731 | 0.3065 | 5.54E-05 |
| No-COT | 0.0598 | 0.0673 | 0.0715 | 0.0031 | 0.7612 | 0.0420 | 0.3089 | 5.53E-05 |

Table 3: Performance comparison after removing the three modules separately

DailyLoc and IntentDist may also be due to this reason. The ablation experiments show that the three parts we introduce in prompts design are all effective and contribute to the final performance.

4.5 Analysis and Comparison of LLM Cost(RQ3)

One of our original reasons for using Gravity model for spatial selection is to reduce the cost of using LLMs. How much can we reduce the cost compared to purely using LLM for generation? To solve this question, we remove the Gravity model from the model and instead use prompts to guide LLM to decide the specific location and duration for each intention by itself. Since the LLM itself does not have access to actual map data, we will provide some alternative POI locations and their distances to the current location externally, so that the LLM can make its own selection based on the choices. After conducting multiple rounds of experiments, we count the number of tokens needed for the LLM to generate a person’s trajectory for the day. At the same time, we also evaluate the results generated by pure LLM. Both the cost and the evaluation results are shown in Figure4. It turns out that it takes 45,592 tokens to generate a person’s trajectory for a day using LLM purely, while our model costs as low as 300. The huge cost difference is not only due to the convenience of Gravity model in location selection, but also because the same template can be used to print the trajectory multiple times without destroying the spatial diversity. In our experiment, we map each time-intention template to 20 specific geographical trajectories. On the other hand, the results of LLM on several metrics are far inferior to ours, indicating that while LLM can make reasonable judgments at the intention level, its perception of geographical space is weak, and it struggles to plan its time effectively. Therefore, our use of mechanisms like the Gravity model to replace LLM in spatial-temporal decision making has proven to be efficient and successful.

4.6 Case studies

4.6.1 Case Study 1: Few-shot performance exploration. In contrast to previous deep generative models that required large amounts of data to fit distributions, our model requires only a very small amount of data. They are mainly used to count the distribution of the number of intentions versus time and to fit the exponent of distance decay in gravitational models. In our experiment, only 200 trajectories of 7 days were used. Here, we explore the effect of different amounts of data on the experiment and the results are shown in Figure 5. As can be seen from the figure, the results for most of the metrics are getting better as the amount of data used increases. Statistical metrics like IntentType and Radius are less affected by the amount of data. The former is generated autonomously by the LLM and does not rely on real data

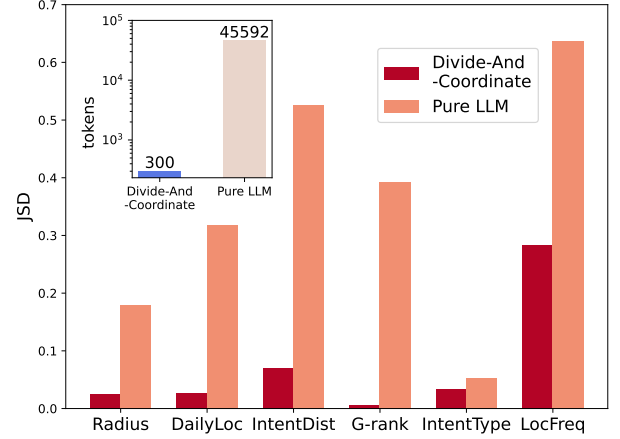


Figure 4: Comparison between the results of Pure LLM and our Divide-And-Coordinate approach

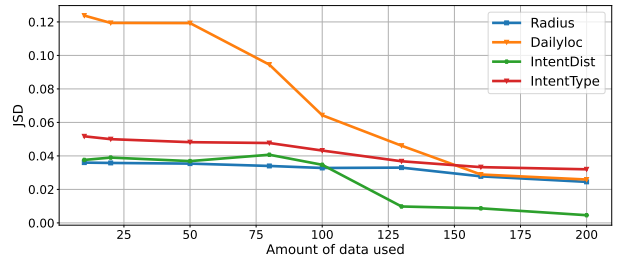


Figure 5: Impact of different training sample sizes on the performance of MobiGeaR.

distribution, while the latter does not require much real data to be fitted. However, metrics like Dailyloc and IntentDist are highly affected, with a clear trend of optimization as the amount of data rises, suggesting that the learning of these movement patterns requires a higher amount of data.

4.6.2 Case Study 2: Trajectory Prediction using generated trajectories. In this experiment, we will explore the utility of the generated data in real-world applications. Here, we consider a realistic scenario where a mobile service operator’s trajectory data cannot be shared with other companies due to issues such as privacy and permissions, and therefore cannot be applied to downstream real-world application tasks. Given the lack of data

volume, our model can generate high-quality data to augment the sparse amount of data in applications.

For the selection of specific applications, we choose the mobility prediction task which is to predict the future trajectory points based on the historical trajectory. The prediction task requires accurate mining and modelling of motion patterns in historical data, and is therefore well suited for checking whether the generated data has reasonable patterns and distribution characteristics that are close to the real ones.

We conduct this experiment on the Tencent dataset, dividing the task into geospatial prediction and semantic-level activity type prediction, both using accuracy to measure the prediction performance. The results of the experiment are shown in Figure 6. We fixed the amount of real data at 100, then added varying amounts of generated data to examine the gain in predictive performance. As can be seen from the figure, the generated data from our model all have better enhancements compared to the data generated from baseline models, which is particularly evident in the intention type prediction.

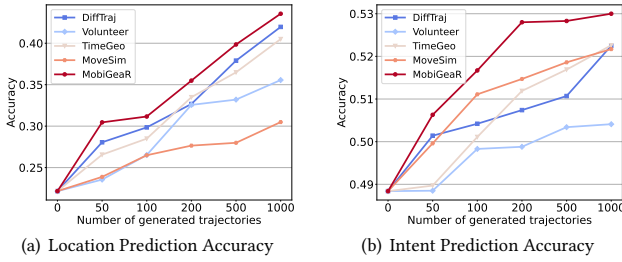


Figure 6: Performance boost in downstream mobility prediction task when different sizes of generated mobility data is provided as data augmentation.

5 RELATED WORK

5.1 Deep Generative Models for Mobility Data

Deep generative models have been widely applied to human trajectory generation problems. Feng et al. [11] incorporate a attention-based region network into the generator of a adversarial network to introduce prior information about urban structures. Long et al. [19] design two VAEs to model the locations of users' homes and workplaces as well as their temporal patterns. Zhu et al. [38] embedding conditional information in the reverse denoising process in the diffusion model to improve the fidelity of trajectories. Fundamentally, deep generative models are aimed at capturing and replicating the complex distributions of real-world data. GANs imitate data through a powerful generator under the supervision of a discriminator. VAEs map input data to a latent space through an encoder, and then map latent variables back to the original data space through a decoder, maximizing the likelihood to ensure that the data distribution generated by the decoder is consistent with the real one. Meanwhile, diffusion models restore the original data step by step through a reverse denoising process from noise. Their basic principles can be summarized as follows: first, thoroughly learn the data, and then generate new instances with similar distributions and features. In contrast, our model does not invest effort in the

first step, relying mainly on LLM inference to generate movement patterns that are consistent with real humans.

5.2 LLM-driven Behavioural and Survey Data Generation

LLMs have increasingly shown their potential in the domain of behavioral and survey generation, marking an innovative leap in how data can be acquired. Recent works [3, 5, 15, 21, 22, 24] demonstrate the ability of LLMs to generate complex, nuanced survey questions and scenarios, potentially replacing traditional methods of survey creation and administration. At the heart of this capability is the LLMs' advanced role play capability, which enables them to assume the perspectives of various respondents, thereby generating more dynamic and engaging survey content. This function not only allows for the creation of diverse and tailored survey materials but also opens the door to understanding intricate behavioral patterns through simulated interactions. However, despite these promising advancements, there is a growing concern regarding the inherent biases that may be perpetuated or even exacerbated by LLMs. [1, 14, 23] As these models learn from vast datasets, they are susceptible to mirroring the biases present in their training data. This could lead to skewed question generation, potentially influencing the responses in ways that reinforce existing stereotypes or biases, thereby affecting the validity of the survey outcomes. We are the first to introduce LLMs into mobile behaviour generation. Combining the powerful role-playing capabilities of LLMs with context-aware reasoning will open up new possibilities for semantically rich trajectory generation.

5.3 LLM-empowered Data Augmentation

Large Language models (LLMs) have been leveraged in several innovative ways to improve the quality and volume of datasets, ultimately enhancing model performance across a range of applications. Dai et al.[8] use LLM to reformulate each training utterance into multiple conceptually similar but semantically different samples, resulting in a high-quality expansion of the clinical NLP dataset. Whitehouse et al. [30] generate new and diverse training data by prompting cases from the original multilingual general knowledge dataset. Acharya et al.[2] use few-shot prompting to generate item descriptions of movies or books for downstream recommendation tasks. While Hsieh et al. [16] make LLMs play the role of a teacher, through COT and other techniques to let LLMs output judgement rationales as the support of predicted labels. These rationales provide rich, detailed thought processes and specialised domain knowledge, greatly reducing the requirements of data volume during small model training. However, despite the widespread use of LLMs in data generation, no one has yet ventured into the field of generating mobility behaviour data and our solution is the first attempt, offering an effective solution to the privacy challenges encountered in downstream applications of trajectory data.

5.4 Reasoning with LLMs

As articulated in [9], the human mental system is composed of a fast, instinctive and intuitive System I and a slow and thoughtful System II. Current LLMs achieves excellent performance on System I, leaving much to be desired in reasoning involved in System II. Through

a simple 'chain of thought prompting' method, Wei et al. [29] efficiently enhance the reasoning ability of LLMs. Building upon this, Yao et al.[33] expand the chain of thought structure into a tree, enhancing the ability for deliberate decision making through multiple reasoning paths and self-evaluation choices. Zhang et al.[36] mimic the human thought process in a cumulative and iterative manner, improving the efficiency by breaking down the task into smaller components. Inspired by these works, we build context-aware COT to guide LLMs to emulate the inner mechanisms of human mobility behaviour and generate reasonable intentions.

6 CONCLUSION

In this article, we propose a new paradigm for generating mobile behaviours, called MobiGeaR. Unlike the prevalent deep generative framework models that fit data distributions from data, our model applies LLMs' role play and few-shot learning capabilities to individual modeling. Through the context-aware reasoning prompt method, our model can generate sequences of intention-level templates that conform to common human lifestyle habits, then a mechanism model is used to map intentions to specific POI locations. Our model has extremely low data requirements and generates results with rich semantics.

In the future, we plan to use higher-quality data, such as more detailed user profiles and more comprehensive behavioural records, for more accurate user modelling and finer-grained behavioural differentiation. Leveraging the understanding capabilities of LLMs, we can also incorporate more factors into the generation process, such as holidays, weather, individual preference and road conditions, to make the generated results more realistic.

REFERENCES

- [1] Alberto Acerbi and Joseph M Stubbersfield. 2023. Large language models show human-like content biases in transmission chain experiments. *Proceedings of the National Academy of Sciences* 120, 44 (2023), e2313790120.
- [2] Arkadeep Acharya, Brijraj Singh, and Naoyuki Onoe. 2023. Llm based generation of item-description for recommendation system. In *Proceedings of the 17th ACM Conference on Recommender Systems*. 1204–1207.
- [3] Gati V Aher, Rosa I Arriaga, and Adam Tauman Kalai. 2023. Using large language models to simulate multiple humans and replicate human subject studies. In *International Conference on Machine Learning*. PMLR, 337–371.
- [4] Rohan Anil, Andrew M Dai, Orhan Firat, Melvin Johnson, Dmitry Lepikhin, Alexandre Passos, Siamak Shakeri, Emanuel Taropa, Paige Bailey, Zhifeng Chen, et al. 2023. Palm 2 technical report. *arXiv preprint arXiv:2305.10403* (2023).
- [5] Lisa P Argyle, Ethan C Busby, Nancy Fulda, Joshua R Gubler, Christopher Rytting, and David Wingate. 2023. Out of one, many: Using language models to simulate human samples. *Political Analysis* 31, 3 (2023), 337–351.
- [6] Dirk Brockmann, Lars Hufnagel, and Theo Geisel. 2006. The scaling laws of human travel. *Nature* 439, 7075 (2006), 462–465.
- [7] Lin Chen, Fengli Xu, Zhenyu Han, Kun Tang, Pan Hui, James Evans, and Yong Li. 2022. Strategic COVID-19 vaccine distribution can simultaneously elevate social utility and equity. *Nature Human Behaviour* 6, 11 (2022), 1503–1514.
- [8] Haixing Dai, Zhengliang Liu, Wenxiong Liao, Xiaoke Huang, Yihan Cao, Zihao Wu, Lin Zhao, Shaochen Xu, Wei Liu, Ninghao Liu, Sheng Li, Dajiang Zhu, Hongmin Cai, Lichao Sun, Quanzheng Li, Dinggang Shen, Tianming Liu, and Xiang Li. 2023. AugGPT: Leveraging ChatGPT for Text Data Augmentation. *arXiv:2302.13007* [cs.CL]
- [9] Kahneman Daniel. 2017. *Thinking, fast and slow*.
- [10] Yves-Alexandre De Montjoye, Laura Radaelli, Vivek Kumar Singh, and Alex “Sandy” Pentland. 2015. Unique in the shopping mall: On the reidentifiability of credit card metadata. *Science* 347, 6221 (2015), 536–539.
- [11] Jie Feng, Zeyu Yang, Fengli Xu, Haisu Yu, Mudan Wang, and Yong Li. 2020. Learning to simulate human mobility. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*. 3426–3433.
- [12] Luciano Floridi and Massimo Chiriatti. 2020. GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines* 30 (2020), 681–694.
- [13] Marta C Gonzalez, Cesar A Hidalgo, and Albert-László Barabási. 2008. Understanding individual human mobility patterns. *nature* 453, 7196 (2008), 779–782.
- [14] Thilo Hagendorff and Sarah Fabi. 2023. Human-Like Intuitive Behavior and Reasoning Biases Emerged in Language Models—and Disappeared in GPT-4. *arXiv preprint arXiv:2306.07622* (2023).
- [15] Perttu Hämäläinen, Mikke Tavast, and Anton Kunnari. 2023. Evaluating large language models in generating synthetic HCI research data: a case study. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–19.
- [16] Cheng-Yu Hsieh, Chun-Liang Li, Chih-Kuan Yeh, Hootan Nakhost, Yasuhisa Fujii, Alexander Ratner, Ranjay Krishna, Chen-Yu Lee, and Tomas Pfister. 2023. Distilling step-by-step: outperforming larger language models with less training data and smaller model sizes. *arXiv preprint arXiv:2305.02301* (2023).
- [17] Shan Jiang, Yingxiang Yang, Siddharth Gupta, Daniele Veneziano, Shounak Athavale, and Marta C González. 2016. The TimeGeo modeling framework for urban mobility without travel surveys. *Proceedings of the National Academy of Sciences* 113, 37 (2016), E5370–E5378.
- [18] Yanchi Liu, Chuanren Liu, Xinjiang Lu, Mingfei Teng, Hengshu Zhu, and Hui Xiong. 2017. Point-of-interest demand modeling with human mobility patterns. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*. 947–955.
- [19] Qingyue Long, Huandong Wang, Tong Li, Lisi Huang, Kun Wang, Qiong Wu, Guangyu Li, Yanping Liang, Li Yu, and Yong Li. 2023. Practical Synthetic Human Trajectories Generation Based on Variational Point Processes. In *KDD 2023*.
- [20] Massimiliano Luca, Gianni Barlacchi, Bruno Lepri, and Luca Pappalardo. 2021. A survey on deep learning for human mobility. *ACM Computing Surveys (CSUR)* 55, 1 (2021), 1–44.
- [21] Joon Sung Park, Joseph O’Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. 2023. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–22.
- [22] Chen Qian, Xin Cong, Cheng Yang, Weize Chen, Yusheng Su, Juyuan Xu, Zhiyuan Liu, and Maosong Sun. 2023. Communicative agents for software development. *arXiv preprint arXiv:2307.07924* (2023).
- [23] Shibani Santurkar, Esin Durmus, Faisal Ladhak, Cinoo Lee, Percy Liang, and Tatsunori Hashimoto. 2023. Whose opinions do language models reflect? *arXiv preprint arXiv:2303.17548* (2023).
- [24] Murray Shanahan, Kyle McDonnell, and Laria Reynolds. 2023. Role play with large language models. *Nature* 623, 7987 (2023), 493–498.
- [25] Chaoming Song, Tal Koren, Pu Wang, and Albert-László Barabási. 2010. Modelling the scaling properties of human mobility. *Nature physics* 6, 10 (2010), 818–823.
- [26] Peter R Stopher and Stephen P Greaves. 2007. Household travel surveys: Where are we going? *Transportation Research Part A: Policy and Practice* 41, 5 (2007), 367–381.
- [27] Huandong Wang, Changzheng Gao, Yuchen Wu, Depeng Jin, Lina Yao, and Yong Li. 2023. PateGail: a privacy-preserving mobility trajectory generator with imitation learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 37. 14539–14547.
- [28] Jason Wei, Yi Tay, Rishi Bommasani, Colin Raffel, Barret Zoph, Sebastian Borgeaud, Dani Yogatama, Maarten Bosma, Denny Zhou, Donald Metzler, et al. 2022. Emergent abilities of large language models. *arXiv preprint arXiv:2206.07682* (2022).
- [29] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems* 35 (2022), 24824–24837.
- [30] Chenxi Whitehouse, Monojit Choudhury, and Alham Fikri Aji. 2023. LLM-powered Data Augmentation for Enhanced Crosslingual Performance. *arXiv preprint arXiv:2305.14288* (2023).
- [31] Fengli Xu, Yong Li, Depeng Jin, Jianhua Lu, and Chaoming Song. 2021. Emergence of urban growth patterns from human mobility behavior. *Nature Computational Science* 1, 12 (2021), 791–800.
- [32] Fengli Xu, Zhen Tu, Yong Li, Pengyu Zhang, Xiaoming Fu, and Depeng Jin. 2017. Trajectory recovery from ash: User privacy is not preserved in aggregated mobility data. In *Proceedings of the 26th international conference on world wide web*. 1241–1250.
- [33] Shunyu Yao, Dian Yu, Jeffrey Zhao, Izhak Shafran, Thomas L Griffiths, Yuan Cao, and Karthik Narasimhan. 2023. Tree of thoughts: Deliberate problem solving with large language models. *arXiv preprint arXiv:2305.10601* (2023).
- [34] Yuan Yuan, Jingtao Ding, Huandong Wang, Depeng Jin, and Yong Li. 2022. Activity trajectory generation via modeling spatiotemporal dynamics. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 4752–4762.
- [35] Yuan Yuan, Huandong Wang, Jingtao Ding, Depeng Jin, and Yong Li. 2023. Learning to Simulate Daily Activities via Modeling Dynamic Human Needs. In *Proceedings of the ACM Web Conference 2023*. 906–916.
- [36] Yifan Zhang, Jingqin Yang, Yang Yuan, and Andrew Chi-Chih Yao. 2023. Cumulative reasoning with large language models. *arXiv preprint arXiv:2308.04371* (2023).
- [37] Yu Zheng, Xing Xie, Wei-Ying Ma, et al. 2010. GeoLife: A collaborative social networking service among user, location and trajectory. *IEEE Data Eng. Bull.* 33, 2 (2010), 32–39.
- [38] Yuanshao Zhu, Yongchao Ye, Shiyao Zhang, Xiangyu Zhao, and James Yu. 2023. DiffTraj: Generating GPS Trajectory with Diffusion Probabilistic Model. In *Thirty-seventh Conference on Neural Information Processing Systems*.

APPENDIX

A DATASET

Tencent Dataset. This dataset is collected from a social network platform and records users' mobility trajectories. Additionally, users' profiles, such as income level, gender, occupation, education level and age, are collected through digital surveys. We randomly select 200 7-day trajectories from the whole dataset to fine-calibrate the POI matching and manually annotate specific intention types. All the experiments on MobiGeaR use only these 200 trajectories.

Mobile Dataset. This dataset is collected from a local mobile operator and records the time and location of users' connections to nearby cellular base stations. Similarly, the user profiles are collected through surveys. For each record in this dataset, we match the location of the visit to the nearest POI based on latitude and longitude coordinates, and then characterize the intention of this visit by the category attributes of the POI. We also use only 200 pieces of data to carry out experiments on MobiGeaR.

The basic information about the trajectory data for both datasets is shown in the following table:

| Datasets | Tencent | ChinaMobile |
|--------------------|-------------------------------------|--------------------------------|
| Duration | October 1, 2019 - December 31, 2019 | July 1, 2017 - August 31, 2017 |
| City | Beijing | Beijing |
| #Users | 100000 | 1246 |
| #Records | 100000 | 1246 |
| #Trajectory Points | 297363263 | 4163651 |

Table 4: Basic information about the trajectories of the two datasets

The distribution of character profiles for both datasets is shown in the following tables:

| User Profile | Category |
|--------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Income | Low (11.44%), Slightly Low (18.12%), Medium (28.67%), Slightly High (18.19%), High (3.36%), Uncertain (20.22%) |
| Gender | Male (54.67%), Female (44.68%), Uncertain (0.65%) |
| Education | Bachelor's degree (25.90%), High school diploma (21.70%), Master's degree (9.94%), Elementary school diploma (13.30%), Junior high school diploma (13.05%), Associate degree (8.02%), Doctoral degree (5.27%) |
| Age | 0-30 (44.50%), 30-40 (30.40%), 40-60 (23.55%), 60-99 (1.55%) |
| Job(top8) | IT Engineer (11.57%), Online Sales (8.37%), Training Instructor (7.34%), Investment (7.31%), Finance (6.97%), Copywriting (6.91%), Front Desk Reception (6.15%), Technical Worker (6.05%) |

Table 5: Distribution of profiles in Tencent dataset

| User Profile | Category |
|--------------|--------------------------------------------------------------------------------------------------------------------|
| Income | Low (23.27%), Relatively Low (20.30%), Medium (36.44%), Relatively High (15.97%), High (4.01%) |
| Gender | Male (63.56%), Female (36.44%) |
| Education | Bachelor degree (58.43%), High school degree (21.03%), Master's degree (11.32%), Junior high school degree (9.23%) |
| Age | 0-30 (22.71%), 30-40 (28.01%), 40-60 (40.85%), 60-99 (8.43%) |

Table 6: Distribution of profiles in ChinaMobile dataset

B BASELINE

- **TimeGeo.** This work models human mobility behavior as a chain of probabilistic choices. It defines the weekly home-based tour number, dwell rate and burst rate to model the temporal choices and employs EPR to model the spatial choices.
- **MoveSim.** This model utilizes a generative adversarial framework, wherein the discriminator is continually reinforced to supervise the generator in producing high-quality trajectory data. To introduce prior knowledge about urban spatial structures, an attention-based region network is designed within the generator. Furthermore, leveraging the spatial continuity and temporal periodicity migration patterns, both the generator and discriminator are pretrained to accelerate the learning process.
- **DiffTraj.** This work proposes a spatio-temporal diffusion probabilistic model for trajectory generation, the core idea of which is to reconstruct and synthesize geographic trajectories from white noise through a reverse trajectory denoising process. UNet is used to embed conditional information and accurately estimate the noise level in the inverse process.
- **Volunteer.** This work uses two vae for joint modeling, the first user VAE is used to learn the distribution of an individual's residence and workplace, and the second VAE decouples travel time and dwell time to accurately model an individual's movement trajectory.

C STATISTICAL EVALUATION METRICS

- **Radius.** radius of gyration, which represents the spatial range of the user's daily activities.
- **DailyLoc.** daily visited locations, which is calculated as number of different locations visited per day for each of the user.
- **IntentDist.** number of intentions per day, which differs from Dailyloc in that there is no de-duplication of locations.
- **G-rank.** number of visits to different locations, which is calculated as the top-100 visiting frequency among all the locations.