

Predicting Human Mobility in Disasters via LLM-Enhanced Cross-City Learning

Yinzhou Tang, Huandong Wang, Xiaochen Fan, and Yong Li

Abstract—The vulnerability of cities to natural disasters has increased with urbanization and climate change, making it more important to predict human mobility in the disaster scenarios for downstream tasks including location-based early disaster warning and pre-allocating rescue resources, etc. However, existing human mobility prediction models are mainly designed for normal scenarios, and fail to adapt to disaster scenarios due to the shift of human mobility patterns under disaster. To address this issue, we introduce DisasterMobLLM, a mobility prediction framework for disaster scenarios that can be integrated into existing deep mobility prediction methods by leveraging LLMs to model the mobility intention and transferring the common knowledge of how different disasters affect mobility intentions between cities. This framework utilizes a RAG-Enhanced Intention Predictor to forecast the next intention, refines it with an LLM-based Intention Refiner, and then maps the intention to an exact location using an Intention-Modulated Location Predictor. Extensive experiments illustrate that DisasterMobLLM can achieve a 32.8% improvement in terms of Acc@1 and a 35.0% improvement in terms of the F1-score of predicting immobility compared to the baselines. The code is available at <https://github.com/tsinghua-fib-lab/DisasterMobLLM>.

Index Terms—Mobility Prediction, Disaster, Large Language Models

I. INTRODUCTION

WITH the rapid urbanization [1] and climate change [2], cities across the world are becoming increasingly vulnerable to natural disasters (e.g., heavy rains), exposing more human lives and properties at risk. To tackle these challenges, a fundamental research problem is to predict human mobility during disaster scenarios, which can support a wide spectrum of downstream emergency response tasks including location-based early disaster warning [3]–[5], pre-allocating rescue resources [6], and planning humanitarian relief [7], etc.

As a classic machine learning problem, human mobility prediction has been studied for decades; however, most existing work [8], [9] has focused on normal scenarios rather than disaster scenarios. As illustrated in Fig. 1(a) and (b), we employ two representative algorithms trained in the normal scenario, i.e., DeepMove [8] and Flashback [10], to predict human mobility in normal scenarios and disaster scenarios, respectively. Their performance in disaster scenarios significantly decreases compared with normal scenarios, with an average relative performance gap of 46.4% and 24.5% in terms of accuracy and mean reciprocal rank, respectively.

Y. Tang, H. Wang, X. Fan, and Y. Li are with the Department of Electronic Engineering, Beijing National Research Center for Information Science and Technology (BNRist), Tsinghua University, Beijing 100084, China (E-mail: wanghuandong@mail.tsinghua.edu.cn).

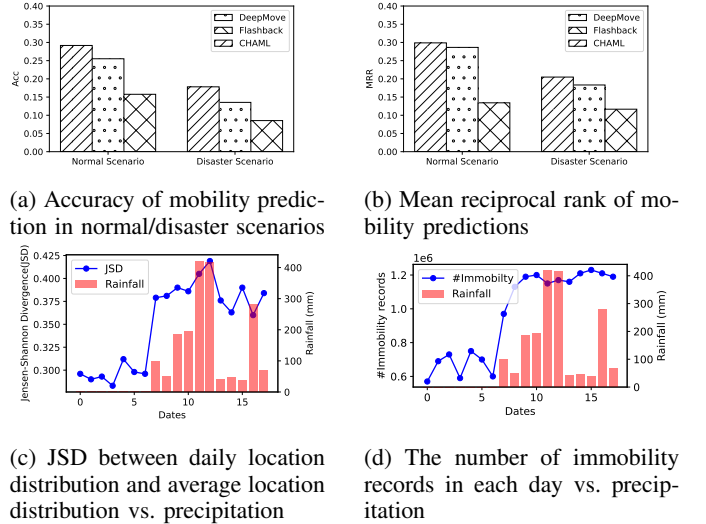


Fig. 1: Analysis of mobility prediction performance of existing algorithms trained in the normal scenario (i.e., DeepMove and Flashback) as well as cross-city transfer learning algorithm (i.e., CHAML), and the cause of their underperformance.

This underperformance arises from the significant difference in mobility patterns between normal and disaster scenarios in terms of spatial distribution (see Fig. 1(c)) and immobility probability [11] (see Fig. 1(d)). Therefore, mobility prediction in disaster scenarios remains an open problem.

However, developing a specialized mobility prediction model for disaster scenarios is also a challenging task. The most critical issue lies in the diversity of disaster situations, e.g., different rainfalls ranging from once-in-a-decade to once-in-a-century, which are usually beyond our experience. It indicates the lack of available mobility data for a specific disaster in a particular city to train the specialized model before the disaster [12]. A promising solution to this problem is to leverage data from similar disasters in other cities, transferring knowledge to train the disaster-specialized model effectively. However, we also observed that simply utilizing mobility data of similar disasters from other cities through transfer learning techniques still cannot achieve satisfactory performance. Specifically, the CHAML algorithm [13] still exhibits a substantial relative performance gap of 38.9% and 31.4% under disaster scenarios compared to normal scenarios in terms of accuracy and mean reciprocal rank (see Fig. 1(a)). The largely different distribution of spatial venues in source cities and target cities makes it difficult to effectively transfer

knowledge. Further, existing cross-city mobility prediction algorithms [14], [15] also fail to utilize the massive trajectory data in normal scenarios of source cities. An ideal approach should effectively utilize trajectory data from the source in both normal and disaster scenarios while decoupling distribution differences of spatial venues to enable efficient transfer of common knowledge describing the shift of mobility patterns (e.g., from the perspective of mobility intention) between normal scenarios and disaster scenarios.

In this paper, we propose **DisasterMobLLM**, an LLM-based cross-city transfer learning framework for mobility prediction in disaster scenarios. In this framework, we leverage LLMs to transfer knowledge about how disasters influence human mobility at the intention level, facilitating the understanding and prediction of mobility intentions across different cities in disaster scenarios. Specifically, we propose an RAG-Enhanced Intention Predictor to retrieve similar trajectories with the target user and target disaster situations as external knowledge. We also propose an intention translator and predictor that aligns intention modality and language modality to enhance LLMs to better forecast the subsequent intention. In the subsequently introduced LLM-based Intention Refiner, we utilize an intention-incorporated prompt based on Chain-of-Thought (CoT) reasoning and a disaster-level-aware soft prompt to refine the predicted intentions incorporating the shift of intention from normal scenarios to disaster scenarios. Finally, we introduce an Intention-Modulated Location Predictor combining an arbitrary base mobility prediction model and intention information to predict the next location of user movement in disaster scenarios.

Our contributions can be summarized as follows:

- We propose DisasterMobLLM, an LLM-based cross-city transfer learning framework, which utilizes LLM-adapted intention embeddings to modulate arbitrary base models to predict human mobility in disaster scenarios. It uses LLM to understand how different disasters affect human mobility and transfer it across cities and scenarios.
- We design an Intention-CLIP to achieve the modality alignment between intention modality and language modality to enhance the ability to understand human mobility intention for LLMs and predict the next intention.
- Extensive experiments demonstrate that, compared to baselines, DisasterMobLLM achieve a 32.8% improvement in Acc@1, a 28.3% improvement in MRR, and a 35.0% improvement in the F1-score for immobility, effectively compensating for the performance decrease in disaster scenarios of base models in normal scenarios.

II. RELATED WORKS

Traditional Deep Learning Models in Mobility Prediction

Since mobility prediction tasks have been widely studied in recent years [9], [16], [17], there are many mobility prediction models using deep learning methods. Feng et. al. proposed DeepMove [8] to use RNN and attention to extract time and location embeddings for trajectory prediction. Flashback [10] enhances the prediction performance of RNNs by emphasizing periodic and contextual information. STiSAN+ [18]

further refines predictions by considering both daily and hourly timing along with spatial relationships. SSDL [19] represents the complex human mobility semantics in different hidden spaces by decoupling the time-invariant and time-varying factors, and the POI graph structure is designed to explore the heterogeneous cooperation signals in the historical mobility. STAR [20] captures spatio-temporal correspondence by designing different spatio-temporal maps and building a stay branch to simulate the stay time in different locations, which is ultimately optimized through adversarial training.

There are also works using transfer learning to extract knowledge from other cities to enhance the prediction in the current city. CHAML [13] considers both city-level and user-level difficulty to enhance conditional sampling in the meta-training process. Xu et. al. propose CrossPred [14] that takes into account multiple features for cross-city POI matching and uses maximum mean difference (MMD) to enhance shared POI features between cities.

However, these models do not consider the mobility pattern shift in disaster scenarios. Thus, a series of mobility prediction models specially designed for disaster scenarios are proposed. Song et. al. [21] proposed a system using HMM to model human behavior and use MDP to generate human mobility in disaster scenarios. DeepMob [22] uses multi-modal learning to combine the information of behavior and mobility by a shared representation in the deep neural network and achieves a performance gain in both behavior and mobility prediction. WS-BiGNN [23] represents mobile and network search records as link prediction tasks by constructing a bipartite graph, and introduces a time-weighted module and a search-mobile memory module to achieve the prediction of individual irregular mobility.

LLM-enhanced Models in Mobility Prediction Recently, as LLMs have proved to have great advantages in understanding human behavior, more and more researchers have begun to explore LLMs for mobility prediction [24]. Wang et. al. [25] introduced LLMMob to capture both long-term and short-term mobility patterns by considering "historical" and "contextual" stays, and incorporating time information for more accurate predictions. LLM4POI [26] can understand the intrinsic meaning of context information while maintaining the original format of heterogeneous location data, thus avoiding information loss and realizing POI recommendation. Agent-Move [27] enhances mobility prediction by mining individual patterns, modeling urban effects, and extracting shared population behaviors. However, these models are designed for mobility prediction in normal scenarios and are unable to handle the mobility shift in disaster scenarios.

III. PROBLEM DEFINITION

DEFINITION 1 (Location): A location r is defined as an area of a specific geographical space such as an administration, a street block, or a community. It is associated with attributes including POI category, POI number, etc. Different locations may also have interactions and relationships based on the road network and transportation facility of the city. Based on the definition above, the human mobility trajectory can be defined as follows.

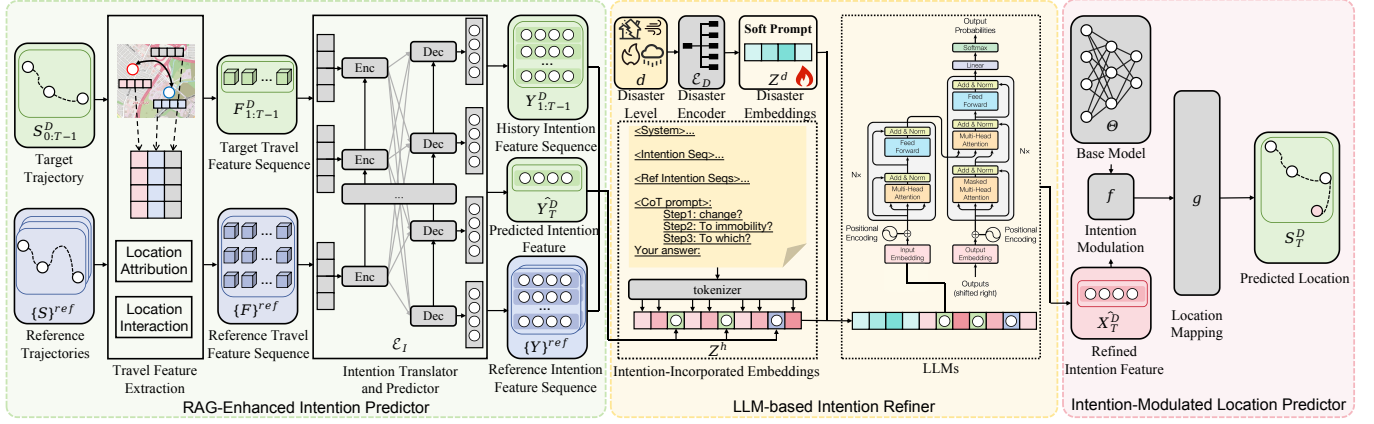


Fig. 2: Overall framework of DisasterMobLLM

DEFINITION 2 (Human Mobility Trajectory): A human mobility trajectory is formulated as a location sequence $S_{0:t} = \{r_0, r_1 \dots r_t\}$ in which r_t indicates the location where the individual is located at timestamp t .

Based on the mobility prediction task [8], [10], [11], [27], [28], we define the task of human mobility prediction in disaster scenarios as follows.

PROBLEM 1 (Human Mobility Prediction in Disaster Scenario): For a given city that experiencing a disaster, we have a user's mobility trajectory $S_{0:T-1}^D$ and a set of historical trajectories of this city in normal scenarios $\{S\}^N$, the goal is to predict the next location r_T of the user. This prediction can be enhanced by incorporating external information such as the disaster level d and mobility trajectories $\{S\}^{ref}$ from other cities in normal and disaster scenarios.

IV. METHOD

A. Overall Framework

To predict human mobility in disaster scenarios, we propose an LLM-driven transfer learning framework named DisasterMobLLM. This framework enhances existing deep mobility prediction models with intention modeled by LLMs and transfers knowledge of how disaster affects mobility patterns among cities and scenarios. The deep mobility prediction model used in the framework is referred to as the base model. In addition, the city experiencing the disaster is denoted as the target city, and other cities are denoted as source cities.

The overall structure of our framework is shown in Fig. 2, which is composed of three modules with the information flow between them shown in Fig. 3. Specifically, we initially introduce an RAG-Enhanced Intention Predictor to extract historical intention sequences from user trajectories with unsupervised learning and predict the next intention. In this part, we train an intention translator and predictor \mathcal{E}_I on the data of normal scenarios in both source and target cities as shown in the green rectangle of Intention Prediction in Fig. 3. Then we introduce the LLM-based Intention Refiner to refine the predicted intention. In this module, we fine-tune an LLM to understand how different disasters affect human mobility patterns in the data of disaster scenarios of source cities

as shown in the orange rectangle of Intention Refinement in Fig. 3. Finally, we use an Intention-Modulated Location Predictor with a deep-learning mobility prediction model as the base model to predict the specific location of the next timestamp with intention modulation. In this part, we train the base model in normal scenarios of the target city to learn how intention affects the prediction of the next location in the target city as shown in the red rectangle of Location Prediction in Fig. 3.

Thus, the framework can transfer the knowledge of how different disaster affects human mobility patterns in different cities and predict the next location in disaster scenarios.

B. RAG-Enhanced Intention Predictor

In this module, we train an intention translator and predictor \mathcal{E}_I to predict the next intention of the given trajectory. Furthermore, we also retrieve a set of reference trajectories to address the data sparsity problem of trajectories in the target city for the target disaster.

1) *Intention Translator and Predictor:* As mentioned in Sec. I, the mobility patterns shift in the disaster scenario, and it is difficult to transfer between cities. Thus, we convert the trajectory $S_{0:T-1}^D$ to the intention level and predict the next intention X_T^D using an intention translator and predictor \mathcal{E}_I .

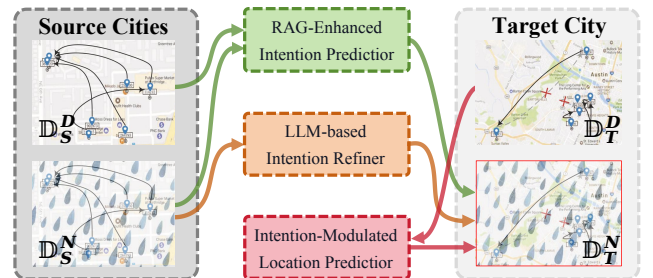


Fig. 3: Information flow in our framework, in which \mathbb{D}_S^D , \mathbb{D}_S^N , \mathbb{D}_T^D , \mathbb{D}_T^N refers to the data in normal scenarios in the source cities, data in disaster scenarios in the source cities, data in normal scenarios in the target cities, data in disaster scenarios in the target cities, respectively.

Since the intention is a concept related to travel between locations, i.e. location switch, as defined in [29], [30], we convert the trajectory $S_{0:T-1}^D$ to a travel feature sequence $F_{1:T-1}^D$ through travel feature extraction [29], [31], [32]. Specifically, given a travel between two locations in a source city, we extract the location attribution of these two locations including the number of each POI category in the location, the closest distance to the nearest transportation facility POI, etc., and interactions between two locations such as the number of each type of road between two locations as the travel feature. Then we map the feature of each travel in the travel feature sequences to a shared intention space to obtain the intention feature sequences in intention modality $X_{1:T-1}^D$ for each trajectory. This process, defined as intention mapping, includes a Transfer Component Analysis (TCA) [33] on all travel feature sequences in all source cities and an unsupervised clustering. We consider each cluster as an intention following [31] and use the feature of the central sample of each cluster as the target intention feature, thus obtaining an intention feature sequence for each trajectory. Aside from that, in the intention feature sequences, we separately assign an immobility embedding to all travels in which the location is not changed. We define the data form obtained from the above method as intention modality.

However, the LLMs cannot understand these intention features because they are pre-trained in the language modality [34] while the intention feature sequence is in the intention modality. Thus, we use a transformer that receives a travel feature sequence $F_{1:T-1}^D$ and outputs the embedding of the next intention Y_T^D at language modality to align the intention feature sequence to the language modality. This process can be formulated as follows:

$$Y_T^D = \mathcal{E}_I(F_{1:T-1}^D), \quad (1)$$

in which $F_{1:T-1}^D$ refers to any travel feature sequences including the given trajectories and reference trajectories. It should be noticed that the output is the next intention Y_T^D because we use the time series modeling capability of the transformer to predict the next intention.

2) *Intention-CLIP*: To train the intention translator and predictor \mathcal{E}_I , we propose **Contrastive Language-Intention Pre-training (Intention-CLIP)** illustrated in Fig. 4. It achieves the prediction of the next intention and the alignment of the language-intention modality at the same time. Inspired by [35]–[37], we design an attention mechanism based on LLM’s vocabulary table to achieve alignment. We firstly pass the vocabulary table of the LLM $V \in \mathcal{R}^{N_V \times D_V}$ through a weight allocation function h to obtain the matrix of intention-related prototypes $P \in \mathcal{R}^{N_P \times D_V}$ indicating a compressed intention feature. The weight allocation function h represents the weights of all combinations of the intention-related vocabularies to the prototypes. This process can be formulated as follows:

$$P = h(V), \quad (2)$$

in which N_V refers to the number of LLM vocabulary, D_V refers to the dimension of LLM vocabulary embeddings and

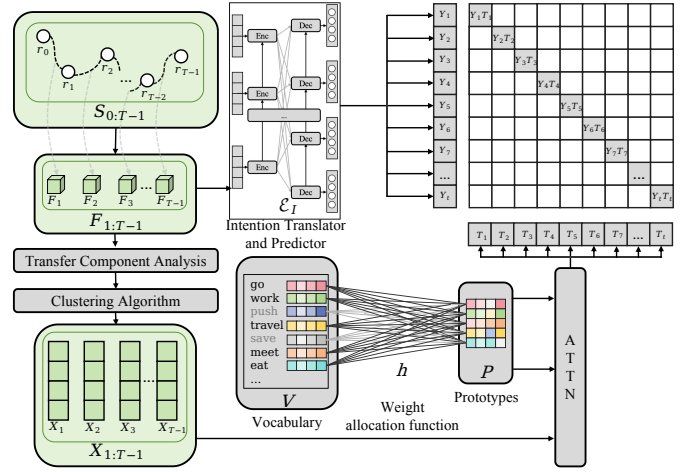


Fig. 4: Illustration of Intention-CLIP

N_P refers to the number of prototypes, which is a hyper-parameter. Then, we use an attention mechanism to obtain the relationship between the intention feature and prototypes by using X_t as query and P as key and value to obtain the intention embedding at the language modality T_t .

$$T_t = \text{softmax} \left(\frac{X_t P^T}{\sqrt{d_k}} \right) P. \quad (3)$$

When conducting Intention-CLIP, we obtain the trajectory embedding by feeding the first $T-1$ elements in the location intention-oriented sequence into the intention translator and predictor \mathcal{E}_I to obtain the prediction of the next intention Y_t . We then update the weights of the intention translator and predictor, the attention mechanism, and the weight allocation function by optimizing the InfoNCE loss [34] and cross-entropy loss as follows:

$$\mathcal{L}_C = \text{InfoNCE}(T_t, Y_t) + \text{CE}(T_t, Y_t), \quad (4)$$

in which *InfoNCE* refers to the InfoNCE loss and *CE* refers to the Cross-entropy loss. In addition, for all immobility travels, we initialize their embedding as the embedding after passing “stay still” to the tokenizer of the LLM and optimize this embedding together using \mathcal{L}_C . Thus, as \mathcal{E}_I is trained in \mathbb{D}_S^N and \mathbb{D}_T^N as mentioned in Fig. 3, it can learn the intention transfer patterns in normal scenarios and output the feature of the next intention in the language modality Y_T given an intention-oriented feature sequence $X_{0:T-1}$.

3) *Trajectory Retrieval*: To address the sparsity of trajectories in the target city of the target disaster, we retrieve reference trajectories set $\{S^{ref}\}$ that are most similar to $S_{0:T-1}^D$ from the database as external knowledge for the prediction. It includes trajectories that are most similar to $S_{0:T-1}^D$ at the intention level among all trajectories with disaster level d in the source cities $\{S^{D,cs}\}^{ref}$, and trajectories that are most similar to the intention feature sequence of $S_{0:T-1}^D$ among all trajectories in the target city in normal scenarios or other disaster scenarios $\{S^{D',ct}\}^{ref}$. We use the travel feature extraction and intention mapping to obtain their intention feature sequences $X_{1:T-1}^D$ and X' and use Dynamic

TABLE I: Performance in disaster scenarios in target city. The best results are in bold and the suboptimal results are underlined.

Model	Acc@1	Acc@10	MRR	NDCG@5	NDCG@10	Pre@Immob	Rec@Immob	F1@Immob
LSTM	0.0381	0.1892	0.0436	0.0729	0.1109	0.1698	0.1375	0.1518
GRU	0.0367	0.1649	0.0359	0.0682	0.0903	0.1624	0.1334	0.1465
DeepMove@LSTM	0.1354	0.2796	0.1832	0.1917	0.3198	0.3358	0.3297	0.3325
Flashback@LSTM	0.0854	0.2143	0.1167	0.1633	0.2874	0.2781	0.2219	0.2468
STISAN+	0.1872	0.3198	0.2187	0.2465	0.3650	0.3681	0.3819	0.3749
HMM+MDP	0.1532	0.3127	0.1970	0.2166	0.3319	0.3613	0.3597	0.3605
DeepMob	0.2180	0.4290	0.2501	0.2789	0.3701	0.3871	0.3540	0.3699
CHAML	0.1377	0.2097	0.1224	0.1401	0.2998	0.2412	0.2588	0.2496
CATUS	0.1781	0.2297	0.1402	0.1621	0.3176	0.2776	0.2809	0.2792
LLM4POI	0.1186	0.2413	0.1522	0.1612	0.2976	0.2971	0.3087	0.3031
ST-MoE-BERT	0.2119	0.4187	0.2478	0.3198	0.3699	0.3387	0.3290	0.3337
Ours@Flashback	0.1579	0.3342	0.1398	0.2022	0.3376	0.4207	0.4398	0.4301
Ours@DeepMove	0.2587	0.4982	0.2905	0.3097	0.3578	0.4863	0.4901	0.4882
Ours@STISAN+	0.2897	0.5118	0.3209	0.3283	0.4601	0.5012	0.5114	0.5062

Time Wrapping [38] to measure the similarity between two trajectories. The process can be formulated as:

$$\{S\}^{ref} = \{S^{D,C_S}\}^{ref} \cup \{S^{D',C_T}\}^{ref}. \quad (5)$$

A more detailed description of the trajectory retrieval process can be found in Appendix VII-A.

C. LLM-based Intention Refiner

Although we have initially predicted the feature of the next intention \hat{Y}_T^D in the RAG-Enhanced Intention Predictor, it only captures the common patterns of human mobility in normal scenarios because \mathcal{E}_I is trained on the dataset for normal scenarios in source cities. Since LLM is proven to be powerful in understanding human behaviors, we use it to refine the predicted next intention feature \hat{Y}_T^D based on the reference intention feature sequences $\{Y\}^{ref}$ and the target disaster level d to get the refined intention feature \hat{X}_T^D . Specifically, we introduce a hybrid prompting strategy including an intention-incorporated prompt with immobility modeling and a disaster-level-aware soft prompt [39] to stimulate the understanding of human behaviors of LLM in refining the intention.

1) *Intention-Incorporated Prompt*: Different from the traditional prompting strategy of LLMs which inputs the tokenized embedding of the text instruction, we design an intention-incorporated prompt inspired by [40]–[42] to better inspire the LLMs to understand intention information. In this prompt, we encode the text instructions into token embeddings using the tokenizer of the LLM while directly using the intention feature sequence as the embeddings for intention tokens to construct the intention-incorporated embeddings Z^h .

Furthermore, to stimulate the LLM’s ability to understand human behavior and to better guide LLMs to understand immobility in the disaster scenario at the intention level, we introduced a Chain-of-Thought prompt framework [43] to refine the intention. The thinking process contains three steps. Firstly, we guide the LLM to decide whether the predicted next intention result is right. If not, we guide LLM to distinguish whether the next intention is immobility. If not, we guide LLMs to decide which intention should be selected as the next intention given all the non-immobility intention embeddings. See Appendix VIII-D for more details.

2) *Disaster-Level-Aware Soft Prompt*: In order to address the heterogeneity of mobility patterns between different disasters, we use a disaster encoder \mathcal{E}_D to construct an adaptive embedding Z^d given the disaster level d to model the disaster level. The disaster encoder is instantiated as a look-up table and the encoding process is formulated as:

$$Z^d = \mathcal{E}_D(d), \quad (6)$$

in which Z^d refers to the disaster embedding. We use the disaster embedding Z^d in a soft prompt paradigm to model the heterogeneity of different disasters. Different from the traditional hard prompt, a soft prompt adjusts the behaviors of the LLM by learning a set of trainable embeddings to be used as prefixes for the input sequences. This approach can significantly reduce the number of parameters that need to be updated while maintaining or improving the performance of the model.

Thus, the construction process of the input embedding sequence I can be formulated as follows:

$$I = \text{Concat}(Z^d, Z^h). \quad (7)$$

When fine-tuning the LLM to learn how different disaster affects human mobility patterns, we use prefix-tuning [44] to fine-tune the LLM without changing most of the parameters. We use Eq. 7 to build the input embedding sequence I and input it into the body of LLM without the tokenizer to obtain the predicted next intention label. Then, we use it to index the corresponding intention feature in intention modality to obtain the feature for the next intention \hat{X}_T^D for the given trajectory in the target city. Then we use cross-entropy loss to optimize the model. The loss function can be formulated as follows:

$$\mathcal{L}_P = CE(\hat{X}_T^D, X_T^D). \quad (8)$$

D. Intention-Modulated Location Predictor

After obtaining the refined intention feature \hat{X}_T^D , we use the Intention-Modulated Location Predictor to predict the specific location where the individual will go in the next timestamp r_T with intention modulation. As we highlighted in Sec. I, existing models have achieved relatively accurate human mobility prediction in normal scenarios, and intentions can help us to better transfer the prediction in different cities and different scenarios.

Existing deep learning human mobility prediction models can be considered as a multi-classification task for all candidate locations. Their model structure can be divided into a mobility embedding extraction network $\mathbf{H} = \Theta(S)$ to extract embedding for location and the last classification layer to map the mobility embedding to the location $\hat{r}_t = \text{softmax}(\text{MLP}(\mathbf{H}))$, in which Θ refers to the mobility prediction model without the last classification layer, $\mathbf{H} \in \mathcal{R}^{D_H \times 1}$ refers to the embedding for the classification task and D_H refers to the hidden dimension. We use an intention modulation function f to incorporate the intention X into existing human mobility prediction models to predict the location. This function can be instantiated into three operations including element-wise production, concatenation, and attention mechanism. The process can be formulated as follows:

$$r_T = \text{softmax}(g(f(\mathbf{H}, X))), \quad (9)$$

in which g refers to the location mapping function and r_T refers to the prediction of the next location. See Appendix VII-B for a more detailed description of the intention modulation function.

When training the Intention-Modulated Location Predictor, we compute the cross-entropy loss \mathcal{L}_M and use it to optimize the model, which can be formulated as follows:

$$\mathcal{L}_M = CE(r_T, r_T), \quad (10)$$

in which r_T refers to the ground truth next location of the given trajectory.

V. EXPERIMENTS

A. Datasets

Through collaboration with a popular mobile application vendor, we collect 7 real-world human trajectory datasets in 7 cities during periods of natural disasters, whose statistics are presented in Appendix VIII-A. In particular, each dataset contains historical requested positioning data from mobile users, exhibiting spatial and temporal characteristics of human mobility in disaster scenarios. We split them by setting Zhongshan as the target city and the other cities as source cities.

B. Baselines and Experimental Settings

Baselines Different categories of baselines are compared with DisasterMobLLM. Specifically, we select traditional mobility prediction models including LSTM [45], GRU [46], DeepMove [8], Flashback [10], and STiSAN+ [18]. We also select HMM+MDP [21] and DeepMob [22] which are specially designed for disaster scenarios. Furthermore, we select models using cross-city knowledge to enhance the prediction performance including CHAML [13] and CATUS [47]. Except for traditional deep mobility prediction models, we also use mobility prediction models enhanced by LLM, such as LLM4POI [26] and ST-MoE-BERT [48]. A more detailed description of baselines can be found in Appendix VIII-B.

Metrics To compare the performance of DisasterMobLLM and all types of baseline methods, we choose average Accuracy@N (i.e., Acc@N and $N = 1, 10$), Mean Reciprocal Rank (MRR), and Normalized Discounted Cumulative Gain (NDCG) as general metrics in evaluations. Moreover, to further delve into the performance enhancement for immobility, we use Precision (Pre@Immob), Recall (Rec@Immob), and F1-score (F1@Immob) for experimental studies with labeled immobility timestamps. A more detailed introduction of the metrics can be found in Appendix VIII-C.

Experimental Settings To implement a fair comparison between different algorithms, we train LSTM, GRU, DeepMove, Flashback, and STiSAN+ on \mathbb{D}_T^N , train HMM+MDP and DeepMove on one part of \mathbb{D}_T^D , and train CHAML and CATUS on \mathbb{D}_S^D . For LLM4POI and ST-MoE-BERT, we directly add an instruction with the disaster level and train them on \mathbb{D}_T^N . For DisasterMobLLM, we apply it to different base models to conduct comprehensive evaluations, in which we use Llama-3-8B¹ as the LLM of our framework. Finally, for all models, we conduct testing in the other parts of \mathbb{D}_T^D .

C. Overall Performance

The performance comparison of all models is shown in Table I. In comparison with the best-performing baseline methods, DisasterMobLLM with STiSAN+ as the base model shows remarkable enhancements, with a 32.8% improvement in Acc@1, a 19.3% improvement in Acc@10, a 28.3% improvement in MRR, a 2.6% improvement in terms of NDCG@5, a 24.3% improvement in NDCG@10, a 29.4% improvement in Pre@Immob, a 33.9% improvement in Rec@Immob, and a 35.0% improvement in F1-score@Immob. The key reason is that our method takes the impact of disasters on human mobility patterns into account and we explicitly model the immobility in disaster scenarios. We stimulate the power of LLM in understanding how disaster affects human mobility patterns by using a hybrid tokenize strategy. Moreover, DisasterMobLLM transfers cross-city knowledge at the intention level instead of directly transferring the entire mobility trajectory, thus achieving more accurate prediction results.

D. Impact of Base Models in Different Scenarios

In this part, we evaluate the capability of DisasterMobLLM to capture shifts in human mobility during disaster scenarios and compensate for performance degradation across different base models. First, we directly use base models to make predictions on trajectories in normal scenarios. Second, we apply base models for human mobility prediction in disaster scenarios. Third, we integrate base models with intention modulation to make predictions in disaster scenarios. The results shown in Table II illustrate that the performance of different baseline methods decreases significantly in disaster scenarios, for up to 82.2% in terms of Acc@1, 65.2% in Acc@10, and 61.4% in MRR. Once combined with the proposed intention modulation, the performance of the conventional models in

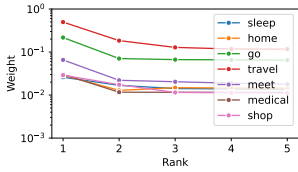
¹<https://huggingface.co/meta-llama/Meta-Llama-3-8B>

TABLE II: Predictive performance of DisasterMobLLM constructed upon different base models in different scenarios, with relative performance differences in terms of different indicators compared to the normal scenarios shown in parentheses.

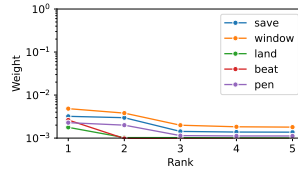
Model	Normal Scenario			Disaster Scenario			Disaster Scenario with Intention Modulation								
	Acc@1	Acc@10	MRR	Acc@1	Acc@10	MRR	MUL			CONCAT			ATTN		
LSTM	0.0607 (--)	0.1527 (--)	0.0820 (--)	0.0181 (-70.2%)	0.0892 (-41.6%)	0.0436 (-46.8%)	0.0581 (-4.3%)	0.1293 (-15.3%)	0.0651 (-20.6%)	0.0626 (+3.1%)	0.1688 (+10.5%)	0.0879 (+7.2%)	0.0640 (+5.4%)	0.1597 (+4.6%)	0.0884 (+7.8%)
GRU	0.0307 (--)	0.1527 (--)	0.0786 (--)	0.0167 (-45.6%)	0.0649 (-57.4%)	0.0359 (-54.3%)	0.0226 (-26.3%)	0.0975 (-36.1%)	0.0601 (-23.5%)	0.0241 (-21.4%)	0.0953 (-37.5%)	0.0575 (-26.8%)	0.0313 (+1.9%)	0.1595 (+4.4%)	0.0738 (-6.1%)
DeepMove	0.2552 (--)	0.4902 (--)	0.2864 (--)	0.1354 (-46.9%)	0.2796 (-43.0%)	0.1832 (-36.0%)	0.2523 (-1.1%)	0.4018 (-18.0%)	0.2695 (-5.9%)	0.2141 (-16.1%)	0.4537 (-7.4%)	0.2294 (-19.9%)	0.2587 (+1.3%)	0.4982 (+1.6%)	0.2905 (+1.4%)
Flashback	0.1576 (--)	0.3287 (--)	0.1342 (--)	0.0854 (-45.8%)	0.1143 (-65.2%)	0.1167 (-13.0%)	0.1534 (-2.0%)	0.3154 (-4.1%)	0.1354 (+0.8%)	0.1521 (-3.4%)	0.3129 (-5.0%)	0.1333 (-0.6%)	0.1579 (-0.1%)	0.3342 (+1.6%)	0.1398 (+4.1%)
STISAN	0.2891 (--)	0.5107 (--)	0.3198 (--)	0.1872 (-35.2%)	0.3198 (-37.3%)	0.2187 (-31.6%)	0.2761 (-4.4%)	0.4909 (-3.8%)	0.2981 (-6.7%)	0.2799 (-3.1%)	0.4981 (-2.4%)	0.3004 (-6.0%)	0.2897 (+0.2%)	0.5118 (+0.2%)	0.3209 (+0.3%)

TABLE III: The predictive performance of our model after removing different modules. The best performance is given in bold.

Model	Acc@10	MRR	Pre@Immob	Rec@Immob	F1@Immob
w/o RAG	0.3148	0.2691	0.3312	0.3187	0.3248
w/o Soft-prompt	0.3387	0.2770	0.3568	0.3248	0.3400
w/o Immobility	0.2896	0.2226	0.2197	0.2613	0.2387
w/o LLM Refining	0.3166	0.2531	0.2769	0.2903	0.2576
Ours	0.5118	0.3209	0.5012	0.5114	0.5062



(a) Weights for intention vocabulary examples.



(b) Weights for non-intention vocabulary examples.

Fig. 5: Top 5 weights of vocabularies in prototypes.

disaster scenarios will improve by up to 13.5% in Acc@1, 12.4% in Acc@10, and 7.8% in MRR.

E. Intention Analysis

Intention-CLIP Analysis To evaluate that our Intention-CLIP can train the intention translator and predictor to learn the intention knowledge in the alignment process, we select both intention-related words (i.e., "go" and "travel") and intention-unrelated words to compare top 5 weights for different prototypes [32]. Intuitively, the linear layer between vocabularies and prototypes can be regarded as a matrix, which indicates how much a vocabulary contributes to different prototypes. Thus, we select the top-5 contributions of intention-related and intention-not-related vocabularies. The results in Fig. 5 show that the top-5 intention-related vocabularies have higher weights than those intention-unrelated vocabularies. Moreover, the top-1 vocabulary's weight is significantly higher than the weight of the second-top vocabulary, showing that the Intention-CLIP can also distinguish different intentions reasonably.

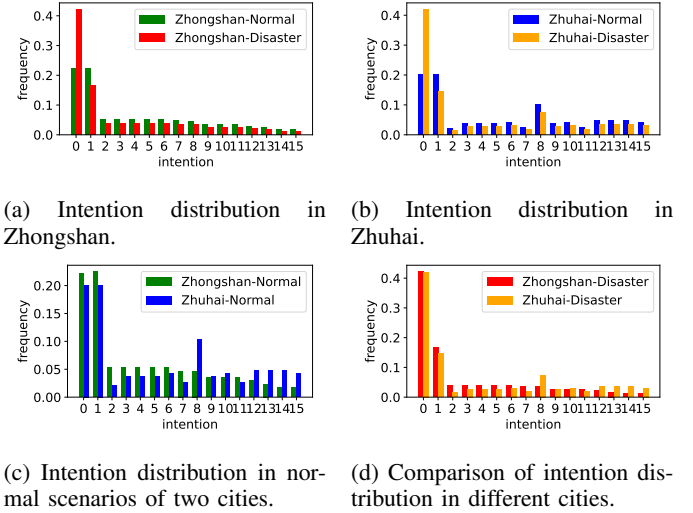


Fig. 6: Comparison of intention distribution between cities.

TABLE IV: Comparison of intention prediction in normal (Norm.) and disaster (Dis.) scenarios.

Scenarios	Acc	Pre@Immob	Rec@Immob	F1@Immob
Norm.	0.693	0.472	0.491	0.481
Dis. w/o LLM	0.438	0.418	0.428	0.422
Dis.	0.763	0.719	0.774	0.745

Intention Prediction Accuracy To evaluate the performance of the intention translator and predictor and the performance of LLM refinement *at intention level*, we compare the performance of intention prediction in both normal and disaster scenarios. The results are shown in Table IV, where the intention prediction accuracy will decrease in disaster scenarios by 58.2% in terms of Acc. In addition, the prediction of immobility intention also decreases when the scenario shifts from normal to disaster. Nevertheless, with the help of powerful LLMs, our method shows better performance in intention prediction even in disaster scenarios while achieving desirable performance in immobility prediction, with enhancements in Acc@1 by up to 10.1%.

Intention Distribution in Different Cities and Scenarios To further explore how our model can successfully transfer mobility knowledge from different cities and disasters by modeling at the intention level, we compare the intention distribution between different cities and disasters. The results shown in Fig. 6 indicate that the intention will heavily shift

when a disaster occurs. Meanwhile, the intention distribution between different cities is relatively smaller in both normal and disaster scenarios.

F. Ablation Study

We also conduct an ablation study to quantify how each module of DisasterMobLLM affects its overall performance. In the *w/o* RAG variant, we remove the retrieval part for reference trajectories, indicating that the model will predict without external knowledge. In the *w/o* Soft-Prompt variant, we remove the soft prompt, indicating that the LLM can only acquire disaster information from the prompt in a few-shot learning fashion. We remove the special design of immobility in the *w/o* Immobility variant. In the *w/o* LLM Refining variant, we remove the LLM-based Intention Refiner and directly use the intention embedding predicted by the Intention Translator and Predictor fine-tuned on the disaster dataset as the modulation of the base model. The experimental results in Table III indicate that deleting any module results in a performance decrease, highlighting the necessity of each module to accurately predict mobility in disaster scenarios.

VI. CONCLUSION

In this paper, we propose a mobility prediction model in a disaster scenario using cross-city knowledge and specially designed at the intention level. Extensive experiments illustrate that DisasterMobLLM can compensate for the performance decrease in disaster scenarios of normal scenario mobility prediction models. Future directions can consider testing our model on more kinds of disasters, including epidemics, wildfires, etc. It is also possible to consider how government policies affect the mobility patterns in disaster scenarios, such as movement restriction and evacuation intervention.

APPENDIX

VII. FRAMEWORK DETAIL

TABLE V: The statistical information of the datasets.

Usage	Dataset	D_n	D_d	$\#S_n$	$\#S_d$	M_p
Source	Qingyuan	7	14	4,858,142	8,000,537	127.08
	Shaoguan	7	14	2,812,149	4,764,873	97.04
	Zhuhai	7	6	3,113,070	6,558,081	75.46
	Wuzhou	7	8	2,609,726	2,448,882	96.43
	Guilin	6	8	9,733,964	15,686,556	83.39
	Hezhou	7	14	4,568,532	6,124,276	88.48
Target	Zhongshan	7	13	16,227,196	30,816,592	90.08

A. Detailed Description of Trajectory Retrieval

To address the sparsity of trajectories in the target city of the target disaster, we retrieve reference trajectories that are most similar to the given trajectory $S_{0:T-1}^D$ from the trajectory database as external knowledge for the prediction. The set of these reference trajectories is formulated as $\{S\}^{ref}$.

Specifically, we initially construct a trajectory database, which includes human mobility trajectories for all cities in

all disaster scenarios and normal scenarios in history. For the given trajectory $S_{0:T-1}^D$ and the disaster level d , we retrieve two categories of trajectories from the database. The first category is trajectories that are most similar to the given trajectory at the intention level among all trajectories with disaster level d in the source cities, which is formulated as $\{S^{D,C_S}\}^{ref}$. This approach addresses the data sparsity issue for the target disaster, as intuitively, even if the target disaster hasn't occurred in the target city, it is much more likely to have occurred in one of the source cities. The second category is the set of trajectories that are most similar to the given trajectory's intention feature sequence among all trajectories in the target city in normal scenarios or other disaster scenarios, which is formulated as $\{S^{D',C_T}\}^{ref}$. Thus, we address the heterogeneity of mobility between different cities. This is due to the heterogeneity of mobility patterns across different cities, which indicates that the knowledge of mobility obtained from other cities is not necessarily fully applicable to the target city.

Specifically, for the given trajectory $S_{0:T-1}^D$ and each trajectory S' in the database, we use the travel feature extraction and intention mapping to obtain their intention feature sequences $X_{1:T-1}^D$ and X' . Then we calculate the similarity between these two sequences using Dynamic Time Wrapping [38] and select the set of the top k most similar trajectories in the above two categories. Then we can get the set of reference trajectories as follows:

$$\{S\}^{ref} = \{S^{D,C_S}\}^{ref} \cup \{S^{D',C_T}\}^{ref}. \quad (11)$$

B. Detailed Description of Intention Modulation Function

We use the intention modulation function g to incorporate intention X into existing human mobility prediction models. The intention modulation function can be instantiated into three operations including element-wise production, concatenation, and attention mechanism as follows:

$$\text{Element-wise Product: } \hat{r}_T = \text{softmax}(g(\mathbf{H} \circ X)), \quad (12)$$

$$\text{Concatenation: } \hat{r}_T = \text{softmax}(g(\mathbf{Cat}(\mathbf{H}, X))), \quad (13)$$

$$\text{Attention: } \hat{r}_T = \text{softmax}(g(\mathbf{Attn}(\mathbf{H}, X))). \quad (14)$$

VIII. EXPERIMENTS DETAIL

A. Statics of the Dataset

Table V shows the statics of our dataset in which D_n and D_d refer to the numbers of days in normal and disaster scenarios, respectively. In addition, $\#S_n$ and $\#S_d$ are numbers of trajectories in normal and disaster scenarios and M_p is the maximum precipitation during the disaster. Note that the precipitation data is collected from the CHIRPS dataset [49] and there are five categorized disaster levels of rainfall.

B. Detailed Description of Baselines

We use following baselines to evaluate the performance of our DisasterMobLLM:

- **LSTM** [45] uses cell states and gating mechanisms to address gradient issues in traditional RNNs.

TABLE VI: Structure of the intention-incorporated prompt.

System Prompt:

You are now a discriminator of predicted human intentions in disaster scenarios, tasked with determining whether the given user's possible next intention is right based on the user's previous intention sequence, disaster level, a possible next intention and other reference intention sequences. Note that:

1. The intentions are token embeddings, that contain the intention information. It is wrapped in the "", which means that any token in two "" refers to the intention embedding rather than the text token.
2. Disaster level is divided into:
 - (1)"no disaster": Indicates that there is no effect on human mobility.
 - (2)"minor disaster": An individual's daily travel plans may not be affected much, but alternatives should be considered.
 - (3)"general disaster": Human mobility may be affected by disasters, causing certain activities to be adjusted or canceled for safety reasons.
 - (4)"severe disaster": Human movement will be greatly affected by the disaster, and the probability of staying still will increase.
3. Reference intention sequences are selected sequences from an intention sequence RAG. They are the most similar sequences to the given intention sequence. You can refer to these sequences to generate your answers but don't copy them exactly.
4. The given possible intention is not necessarily accurate, you need to judge whether it needs to change.
5. Both the user's previous intention sequence and the other reference intention sequences are in the format of a list like ['intention embedding 1', 'intention embedding 2'...].

Intention Sequence Prompt:

Disaster Level: <disaster level>.

Intention embedding sequence: <representation of the intention sequence>.

The given possible next intention embedding for this sequence is <representation of the predicted next intention>.

Reference Intention Sequence Prompt:

You need to refer to the following sequence to distinguish whether the given possible next intention embedding is right:

Reference Sequence 1:

Disaster Level: <disaster level>.

Intention embedding sequence: <representation of the intention sequence>.

The given possible next intention embedding for this sequence is <representation of the predicted next intention>.

...

Reference Sequence <k>:

Disaster Level: <disaster level>.

Intention embedding sequence: <representation of the intention sequence>.

The given possible next intention embedding for this sequence is <representation of the predicted next intention>.

CoT Prompt:

Let's think step by step. You need to answer each of the three questions below, and if the answer to the first question is "yes", the following questions will be output as "None":

- (1) Is the given possible next intent embedding right?
- (2) If the answer to the previous question is "yes", this answer is set to "None". If the answer to the previous question is "no", please answer: Given the current disaster level, should the next intention be "stay still"?
- (3) If the answer to the previous question is "yes", this answer is set to "stay still". If the answer to the previous question is "no", you need to give the index of the correct next intention embedding.

The indexes and embeddings of the intentions you can choose from are <intention representation of all intentions>.

Answer Prompt:

Now give your answer. You should output the answer in a ["no", "no", "2"] format, nothing else.

Your answer:

- **GRU** [46] learns long-term dependencies and reduces complexity by merging gates and using hidden states as cell states.
- **DeepMove** [8] combines RNNs and attention mechanisms to predict human mobility by capturing spatial-temporal patterns.
- **Flashback** [10] integrates long-term history data and short-term behavioral patterns, emphasizing periodic and contextual information, to accurately predict users' next location.
- **STISAN+** [18] captures sequential dependencies by assigning attention weights among different elements, improving the performance in location services for handling spatial-temporal factors and modeling individual behavior patterns. The methods above are designed for normal scenarios, which lack of special design for disaster scenarios. There are also some fundamental models which

are designed for mobility prediction in disaster scenarios.

- **HMM+MDP** [21] combines HMM and MDP to simultaneously predict behaviors and mobility paths of individuals after the occurrence of disasters.
- **DeepMob** [22] enables collaborative prediction of both features by jointly modeling behavior and mobility. However, their performance is restricted by limited knowledge of human mobility across different cities.
- **CHAML** [13] is a framework for cross-city POI search recommendations that tackles data scarcity and sample diversity in cold-start cities by integrating difficult-aware meta-learning and city-level learning strategies.
- **CATUS** [47] is a pre-trained model that learns common transition patterns across cities using self-supervised tasks, enhanced with class transition samplers and implicit-explicit policies. However, these methods fail to account

for disaster scenarios.

- **LLM4POI** [26] predicts human mobility by converting location social network data into Q&A format and combines historical tracks with key query similarity to leverage contextual information.
- **ST-MoE-BERT** [48] is a spatial-temporal framework that integrates BERT models and a Mixture of Experts (MoE) architecture for predicting long-term human mobility patterns across cities. While they make use of the capability of LLMs to comprehend human behaviors, they are not originally designed for disaster scenarios.

C. Detailed Introduction of Metrics

In our evaluation process, we choose different metrics to evaluate how different models perform in predicting human mobility in disaster scenarios. Specifically, we choose two types of metrics to evaluate the performance. We use Average Accuracy (Acc@N), Mean Reciprocal Rank (MRR) and Normalized Discounted Cumulative Gain (NDCG) to evaluate the overall performance of all locations, which can be formulated as follows:

$$\text{Acc@N} = \frac{1}{Q} \sum_{q=1}^Q \mathbb{I}(\text{correct answer in top } N), \quad (15)$$

$$\text{MRR} = \frac{1}{Q} \sum_{q=1}^Q \frac{1}{\text{rank}_q}, \quad (16)$$

in which Q refers to the number of samples and \mathbb{I} is an indication function that returns 1 if the condition is true and 0 otherwise, and rank_q refers to the rank of the first correct answer in the query q . Furthermore, the NDCG can be defined as the division of Discounted Cumulative Gain (DCG) and Ideal Discounted Cumulative Gain (IDCG), which can be formulated as follows:

$$\text{NDCG} = \frac{\text{DCG}}{\text{IDCG}}, \quad (17)$$

in which DCG and IDCG can be formulated as:

$$\text{DCG} = \sum_{i=1}^n \frac{2^{\text{rel}_i} - 1}{\log_2(i + 1)}, \quad (18)$$

$$\text{IDCG} = \sum_{i=1}^{|R|} \frac{2^{\text{rel}_{(i)}} - 1}{\log_2(i + 1)}, \quad (19)$$

in which R is a list of results sorted by true relevance.

Furthermore, we also use Precision (Pre@Immob), Recall (Rec@Immob), and F1-score (F1@Immob) with respect to immobility timestamps to evaluate the prediction performance in immobility. Specifically, the metrics can be formulated as:

$$\text{Precision} = \frac{TP}{TP + FP}, \quad (20)$$

$$\text{Recall} = \frac{TP}{TP + FN}, \quad (21)$$

$$\text{F1-score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}, \quad (22)$$

in which TP , FP , and FN refer to True Positive, False Positive, and False Negative samples, respectively.

D. Intention-Incorporated Prompt

To better guide the LLM to understand human mobility intentions, we propose an intention-incorporated prompt based on the CoT framework and an immobility modeling. The intention-incorporated prompt contains five parts: the system prompt, intention sequence prompt, reference intention sequences prompt, CoT prompt, and the answer prompt. The detailed prompt is shown in Table VI.

In the system prompt, we set the LLM as an intention discriminator to distinguish whether the given intention is right and which intention should be refined if the given one is wrong. Then, we use an intention sequence prompt and a reference intention sequence prompt as shown in Table VI to input the intention sequence and external knowledge. Specifically, we replace the tokenized embedding of the "intention embedding sequence" and "the given possible next intention embedding for this sequence", which is colored blue, with intention feature sequence $Y_{1:T-1}$ and predicted intention feature \hat{Y}_T , respectively, while passing the rest text instructions to the tokenizer of the LLM to generate the other part of the input embedding of LLMs. Furthermore, we also guide the LLM to think step by step to decide whether to refine the given intention to immobility and other intentions. Finally, we use an answer prompt to guide the LLM in giving its answer.

IX. ADDITIONAL RESULTS

A. Inference Time Analysis

Model	LLM4POI	ST-MoE-BERT	DisasterMobLLM
Time (s)	5.61	1.03	5.67

TABLE VII: Comparison of the inference time between different LLM-Based Mobility Prediction Models.

In order to evaluate the practicality of our algorithm, we analyze the inference time of both our model and the LLM-Based Mobility Prediction Model baselines, with hardware comprising two A100 GPUs. For our model and LLM4POI, we use the Llama-3.1-8B as the backbone while using BERT with 110 million parameters as the backbone of ST-MoE-BERT. The results in Table VII indicate that although the inference time of our model is a little bit slower than the baselines, the inference time is still at the same magnitude. This indicates the practicality of our framework.

REFERENCES

- [1] L. Sun, J. Chen, Q. Li, and D. Huang, "Dramatic uneven urbanization of large cities throughout the world in recent decades," *Nature communications*, vol. 11, no. 1, p. 5366, 2020.
- [2] R. Leichenko, "Climate change and urban resilience," *Current opinion in environmental sustainability*, vol. 3, no. 3, pp. 164–168, 2011.
- [3] K. M. Rahman, T. Alam, and M. Chowdhury, "Location based early disaster warning and evacuation system on mobile phones using open-streetmap," in *2012 IEEE conference on open systems*. IEEE, 2012, pp. 1–6.

- [4] X. Zhao, Y. Xu, R. Lovreglio, E. Kuligowski, D. Nilsson, T. J. Cova, A. Wu, and X. Yan, "Estimating wildfire evacuation decision and departure timing using large-scale gps data," *Transportation research part D: transport and environment*, vol. 107, p. 103277, 2022.
- [5] R. Zhang, D. Liu, E. Du, L. Xiong, J. Chen, and H. Chen, "An agent-based model to simulate human responses to flash flood warnings for improving evacuation performance," *Journal of Hydrology*, vol. 628, p. 130452, 2024.
- [6] J. Tang, J. Wang, J. Li, P. Zhao, W. Lyu, W. Zhai, L. Yuan, L. Wan, and C. Yang, "Predicting human mobility flows in response to extreme urban floods: A hybrid deep learning model considering spatial heterogeneity," *Computers, Environment and Urban Systems*, vol. 113, p. 102160, 2024.
- [7] X. Song, Q. Zhang, Y. Sekimoto, R. Shibasaki, N. J. Yuan, and X. Xie, "Prediction and simulation of human mobility following natural disasters," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 8, no. 2, pp. 1–23, 2016.
- [8] J. Feng, Y. Li, C. Zhang, F. Sun, F. Meng, A. Guo, and D. Jin, "Deep-move: Predicting human mobility with attentional recurrent networks," in *Proceedings of the 2018 world wide web conference*, 2018, pp. 1459–1468.
- [9] M. Luca, G. Barlacchi, B. Lepri, and L. Pappalardo, "A survey on deep learning for human mobility," *ACM Computing Surveys (CSUR)*, vol. 55, no. 1, pp. 1–44, 2021.
- [10] D. Yang, B. Fankhauser, P. Rosso, and P. Cudre-Mauroux, "Location prediction over sparse user mobility traces using rnns," in *Proceedings of the twenty-ninth international joint conference on artificial intelligence*, 2020, pp. 2184–2190.
- [11] E. Moro, D. Calacci, X. Dong, and A. Pentland, "Mobility patterns are associated with experienced income segregation in large us cities," *Nature communications*, vol. 12, no. 1, p. 4633, 2021.
- [12] W. Li, Q. Wang, Y. Liu, M. L. Small, and J. Gao, "A spatiotemporal decay model of human mobility when facing large-scale crises," *Proceedings of the National Academy of Sciences*, vol. 119, no. 33, p. e2203042119, 2022.
- [13] Y. Chen, X. Wang, M. Fan, J. Huang, S. Yang, and W. Zhu, "Curriculum meta-learning for next poi recommendation," in *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, 2021, pp. 2692–2702.
- [14] S. Xu and D. Guan, "Crosspred: A cross-city mobility prediction framework for long-distance travelers via poi feature matching," in *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, 2024, pp. 4148–4152.
- [15] J. Ding, G. Yu, Y. Li, D. Jin, and H. Gao, "Learning from hometown and current city: Cross-city poi recommendation via interest drift and transfer learning," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 3, no. 4, pp. 1–28, 2019.
- [16] J. Wang, X. Kong, F. Xia, and L. Sun, "Urban human mobility: Data-driven modeling and prediction," *ACM SIGKDD explorations newsletter*, vol. 21, no. 1, pp. 1–19, 2019.
- [17] C. Zhang, K. Zhao, and M. Chen, "Beyond the limits of predictability in human mobility prediction: Context-transition predictability," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 5, pp. 4514–4526, 2022.
- [18] Y. Jiang, Y. Yang, Y. Xu, and E. Wang, "Spatial-temporal interval aware individual future trajectory prediction," *IEEE Transactions on Knowledge and Data Engineering*, 2023.
- [19] Q. Gao, J. Hong, X. Xu, P. Kuang, F. Zhou, and G. Trajcevski, "Predicting human mobility via self-supervised disentanglement learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 36, no. 5, pp. 2126–2141, 2023.
- [20] Y. Wang, T. Zheng, S. Liu, Z. Feng, K. Chen, Y. Hao, and M. Song, "Spatiotemporal-augmented graph neural networks for human mobility simulation," *IEEE Transactions on Knowledge and Data Engineering*, 2024.
- [21] X. Song, Q. Zhang, Y. Sekimoto, and R. Shibasaki, "Prediction of human emergency behavior and their mobility following large-scale disaster," in *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2014, pp. 5–14.
- [22] X. Song, R. Shibasaki, N. J. Yuan, X. Xie, T. Li, and R. Adachi, "Deepmob: learning deep knowledge of human emergency behavior and mobility from big and heterogeneous data," *ACM Transactions on Information Systems (TOIS)*, vol. 35, no. 4, pp. 1–19, 2017.
- [23] J. Xue, T. Yabe, K. Tsubouchi, J. Ma, and S. V. Ukkusuri, "Predicting individual irregular mobility via web search-driven bipartite graph neural networks," *IEEE Transactions on Knowledge and Data Engineering*, 2024.
- [24] Z. Zhang, Y. Sun, Z. Wang, Y. Nie, X. Ma, P. Sun, and R. Li, "Large language models for mobility in transportation systems: A survey on forecasting tasks," *arXiv preprint arXiv:2405.02357*, 2024.
- [25] X. Wang, M. Fang, Z. Zeng, and T. Cheng, "Where would i go next? large language models as human mobility predictors," *arXiv preprint arXiv:2308.15197*, 2023.
- [26] P. Li, M. de Rijke, H. Xue, S. Ao, Y. Song, and F. D. Salim, "Large language models for next point-of-interest recommendation," in *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2024, pp. 1463–1472.
- [27] J. Feng, Y. Du, J. Zhao, and Y. Li, "Agentmove: Predicting human mobility anywhere using large language model based agentic framework," *arXiv preprint arXiv:2408.13986*, 2024.
- [28] H. Xue, F. Salim, Y. Ren, and N. Oliver, "Mobtcast: Leveraging auxiliary trajectory forecasting for human mobility prediction," *Advances in Neural Information Processing Systems*, vol. 34, pp. 30380–30391, 2021.
- [29] F. Yi, L. Yin, H. Wen, H. Zhu, L. Sun, and G. Li, "Mining human periodic behaviors using mobility intention and relative entropy," in *Advances in Knowledge Discovery and Data Mining: 22nd Pacific-Asia Conference, PAKDD 2018, Melbourne, VIC, Australia, June 3-6, 2018, Proceedings, Part I 22*. Springer, 2018, pp. 488–499.
- [30] S. Li, J. Feng, J. Chi, X. Hu, X. Zhao, and F. Xu, "Limp: Large language model enhanced intent-aware mobility prediction," *arXiv preprint arXiv:2408.12832*, 2024.
- [31] T. He, J. Bao, R. Li, S. Ruan, Y. Li, L. Song, H. He, and Y. Zheng, "What is the human mobility in a new city: Transfer mobility knowledge across cities," in *Proceedings of The Web Conference 2020*, 2020, pp. 1355–1365.
- [32] H. Yuan, Y. Qian, R. Yang, and M. Ren, "Human mobility discovering and movement intention detection with gps trajectories," *Decision Support Systems*, vol. 63, pp. 39–51, 2014.
- [33] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE transactions on neural networks*, vol. 22, no. 2, pp. 199–210, 2010.
- [34] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark *et al.*, "Learning transferable visual models from natural language supervision," in *International conference on machine learning*. PMLR, 2021, pp. 8748–8763.
- [35] M. Jin, S. Wang, L. Ma, Z. Chu, J. Y. Zhang, X. Shi, P.-Y. Chen, Y. Liang, Y.-F. Li, S. Pan *et al.*, "Time-llm: Time series forecasting by reprogramming large language models," *arXiv preprint arXiv:2310.01728*, 2023.
- [36] D. Misra, A. Goyal, B. Runwal, and P. Y. Chen, "Reprogramming under constraints: Revisiting efficient and reliable transferability of lottery tickets," *arXiv preprint arXiv:2308.14969*, 2023.
- [37] C.-H. H. Yang, Y.-Y. Tsai, and P.-Y. Chen, "Voice2series: Reprogramming acoustic models for time series classification," in *International conference on machine learning*. PMLR, 2021, pp. 11808–11819.
- [38] M. Müller, "Dynamic time warping," *Information retrieval for music and motion*, 2007.
- [39] G. Qin and J. Eisner, "Learning how to ask: Querying lms with mixtures of soft prompts," *arXiv preprint arXiv:2104.06599*, 2021.
- [40] J. Tang, Y. Yang, W. Wei, L. Shi, L. Xia, D. Yin, and C. Huang, "Higpt: Heterogeneous graph language model," in *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2024, pp. 2842–2853.
- [41] R. Ye, C. Zhang, R. Wang, S. Xu, Y. Zhang *et al.*, "Natural language is all a graph needs," *arXiv preprint arXiv:2308.07134*, vol. 4, no. 5, p. 7, 2023.
- [42] J. Tang, Y. Yang, W. Wei, L. Shi, L. Su, S. Cheng, D. Yin, and C. Huang, "Graphgpt: Graph instruction tuning for large language models," in *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2024, pp. 491–500.
- [43] J. Wei, X. Wang, D. Schuurmans, M. Bosma, F. Xia, E. Chi, Q. V. Le, D. Zhou *et al.*, "Chain-of-thought prompting elicits reasoning in large language models," *Advances in neural information processing systems*, vol. 35, pp. 24824–24837, 2022.
- [44] X. L. Li and P. Liang, "Prefix-tuning: Optimizing continuous prompts for generation," *arXiv preprint arXiv:2101.00190*, 2021.
- [45] S. Hochreiter, "Long short-term memory," *Neural Computation MIT-Press*, 1997.
- [46] K. Cho, "Learning phrase representations using rnn encoder-decoder for statistical machine translation," *arXiv preprint arXiv:1406.1078*, 2014.

- [47] K. Sun, T. Qian, C. Li, X. Ma, Q. Li, M. Zhong, Y. Zhu, and M. Liu, “Pre-training across different cities for next poi recommendation,” *ACM Transactions on the Web*, vol. 17, 2023.
- [48] H. He, H. Luo, and Q. R. Wang, “St-moe-bert: A spatial-temporal mixture-of-experts framework for long-term cross-city mobility prediction,” *arXiv preprint arXiv:2410.14099*, 2024.
- [49] C. Funk, P. Peterson, M. Landsfeld, D. Pedreros, J. Verdin, S. Shukla, G. Husak, J. Rowland, L. Harrison, A. Hoell *et al.*, “The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes,” *Scientific data*, vol. 2, no. 1, pp. 1–21, 2015.