HuMob Predictor: Towards a Generalizable Model for Multi-City Individual Mobility Prediction

Guangyue Li
Hong Kong Polytechnic University
Hong Kong, China
guangyue.li@connect.polyu.hk

Yuxiao Luo
Hong Kong Polytechnic University
Hong Kong, China
yuxiao.luo@connect.polyu.hk

Shen Liang
Hong Kong Polytechnic University
Hong Kong, China
shen.liang@connect.polyu.hk

Jinghan Wang
Hong Kong Polytechnic University
Hong Kong, China
jinghhan.wang@connect.polyu.hk

Xiaogang Guo
Wuhan University
Wuhan, China
guoxg1006@whu.edu.cn

Ling Yin
Shenzhen Institute of Advanced Technology, CAS
Shenzhen, China
yinling@siat.ac.cn

Luliang Tang
Wuhan University
Wuhan, China
tll@whu.edu.cn

Yang Xu*
Hong Kong Polytechnic University
Hong Kong, China
yang.ls.xu@polyu.edu.hk

ABSTRACT

Multi-city human mobility prediction is critical for GIScience and urban computing. However, pronounced spatiotemporal heterogeneity and training instability on large-scale, multi-city datasets restrict model generalization. To address these challenges, we propose the Human Mobility Predictor (HuMob Predictor), a novel framework that enhances multi-city mobility prediction. Our model introduces a multi-level encoding module that captures cross-city heterogeneity and within-city spatial relationships. City encodings capture macro-level characteristics unique to each city, while absolute spatial encodings preserve geographic proximity across grid cells. By combining these encodings, HuMob Predictor learns universal mobility representations while retaining city-specific features. To stabilize training across cities, we adopt an incremental training strategy that gradually increases prediction difficulty, significantly improving convergence and cross-city generalization. Experiments on the GISCUP 2025 multi-city datasets demonstrate that HuMob Predictor achieves superior performance in individual mobility prediction.

CCS CONCEPTS

• Information systems → Location based services

KEYWORDS

Multi-City Human Mobility Prediction, HuMob Challenge, Absolute Spatial Encoding, Incremental Training Strategy

* Corresponding author. E-mail address: yang.ls.xu@polyu.edu.hk (Xu. Y).

SIGSPATIAL '25, November 3–6, 2025, Minneapolis, MN, USA

ACM ISBN 979-8-4007-2086-4/2025/11.
https://doi.org/10.1145/3748636.3771309

ACM Reference format:

Guangyue Li, Yuxiao Luo, Shen Liang, Jinghan Wang, Xiaogang Guo, Ling Yin, Luliang Tang, Yang Xu. 2025. HuMob Predictor: Towards a Generalizable Model for Multi-City Individual Mobility Prediction. In *The 33rd ACM International Conference on Advances in Geographic Information Systems (SIGSPATIAL '25), November 3 - 6, 2025, Minneapolis, MN, USA.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3748636.3771309

1 Introduction

Predicting human mobility is essential for understanding and optimizing complex urban systems [1]. Forecasting individual spatiotemporal trajectories uncovers patterns in human activity, providing decision support for traffic modeling [2], urban planning [3], and epidemic simulation [4]. The 14th ACM SIGSPATIAL Cup (GISCUP 2025) focuses on the frontier challenge of "Human Mobility Prediction in Multiple Cities", aiming to shift the research paradigm from traditional single-city analysis to more generalizable and practical multi-city prediction.

However, achieving accurate multi-city prediction faces a core challenge: cross-city spatiotemporal heterogeneity. As illustrated by the GISCUP 2025 dataset [5], shown in Figure 1, each city possesses a unique mobility pattern shaped by its distinct road network topology and distribution of functional zones. This heterogeneity causes the performance of mainstream single-city prediction paradigms [6] to degrade significantly when transferred to new urban environments. While alternative fine-tuning models [7] have shown some progress, they do not fundamentally solve the problem of learning universal, location-agnostic mobility patterns through joint training.

To build a unified and generalizable model, it is imperative to confront and process the mixed data from all cities simultaneously. This presents two severe challenges:

1) Urban structure understanding. A model must comprehend complex urban spaces at both macro and micro scales. At the macro level, it needs to differentiate between cities to identify their unique mobility dynamics. At the micro level, it must explicitly recognize fundamental spatial relationships between geographic units, such as proximity and distance, which form the physical basis of all mobility patterns. Without this multi-level spatial awareness, a model would be confused by the intertwined data and fail to learn effectively.

2) Large-scale training. Even with powerful modal architecture, mixing massive trajectory data from four cities drastically increases the complexity of the learning task. Such a highly heterogeneous dataset can lead to unstable training, with the model struggling to converge or overfitting to the patterns of specific cities. This makes it difficult to learn generalizable long-range spatiotemporal dependencies. Therefore, beyond a powerful model, a more intelligent training paradigm is required.

To address the challenges posed by GISCUP 2025, this paper proposes a novel and generalizable multi-city prediction framework, the Human Mobility Predictor (HuMob Predictor). Its core consists of two mechanisms. First, we design a multi-level encoding module that uses a city encoding to capture macro-level urban characteristics and a novel spatial encoding module to explicitly represent micro-level geographic relationships. Second, to tackle the difficulty posed by large-scale, multi-city data, we introduce an Incremental Training Strategy that guides the model to converge efficiently and stably through tasks of gradually increasing difficulty. The main contributions of this paper are summarized as follows:

1) We design a novel trajectory encoding module. Its city encoding effectively addresses cross-city heterogeneity, while its spatial encoding, combining absolute and learnable components, improves awareness of urban structure.

2) We propose an incremental training strategy to reduce the complexity of learning. By dynamically adjusting task difficulty, this strategy can improve training stability and performance on the large-scale, highly heterogeneous multi-city dataset.

3) Based on these innovations, we build a unified and generalizable multi-city prediction framework. Validated on the GISCUP 2025 dataset, our framework achieves superior performance in multi-city individual mobility prediction.

2 Data

This study uses the GISCUP 2025 dataset, which contains mobility trajectories from four major Japanese metropolitan areas. The data spans 75 days, with a spatial resolution of 500m x 500m grids and a temporal resolution of 30-minute intervals. The competition task is to use the mobility data of all individuals from the first 60 days to predict the complete spatiotemporal trajectories of the last 3000 users in each city for days 61 to 75.

As shown in Figure 1, the mobility patterns exhibit significant cross-city spatial heterogeneity, with each city displaying a distinct geographic layout and activity hotspots. Furthermore, the daily activity volume exhibits high temporal volatility. Figure 2 shows that the daily trajectory point volume in the GISCUP 2025 data fluctuates more dramatically than in previous challenges. We hypothesize that these irregular fluctuations, likely stemming from anomalous real-world events (e.g., lockdown or extreme weather), pose an additional challenge for the model to capture stable, long-term mobility patterns.

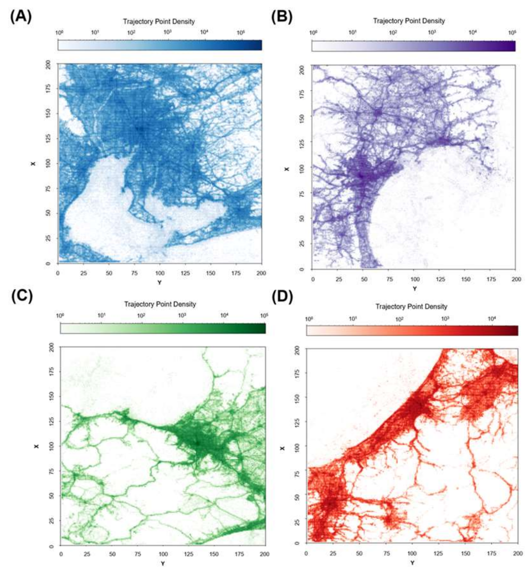

Figure 1. Distribution of trajectory points. (A) Nagoya; (B) Sendai; (C) Sapporo; (D) Ishikawa.

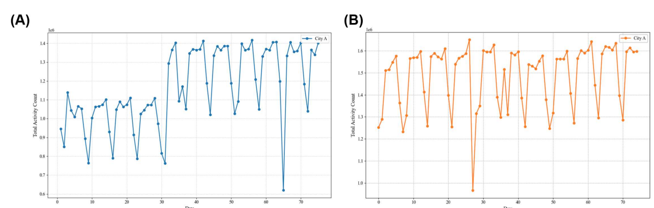

Figure 2. Daily trajectory points of City A. (A) GISCUP 2025; (B) HuMob Challenge 2024.

3 Method

Our proposed framework, the HuMob Predictor, is illustrated in Figure 3(A). It comprises an Encoding Layer, a Spatiotemporal Learning Layer, and a Decoding Layer. To train the model effectively, we designed an Incremental Training Strategy (Figure 3(B)), which dynamically adjusts the task difficulty to enhance training stability and performance on the complex multi-city dataset.

3.1 Human Mobility Predictor

1) Encoding Layer. The encoding layer converts discrete spatiotemporal points into dense vector representations containing rich contextual information from three dimensions: city, space, and time. City Encoding: To distinguish between macro-level mobility patterns, we map the city ID of each user to a vector via an

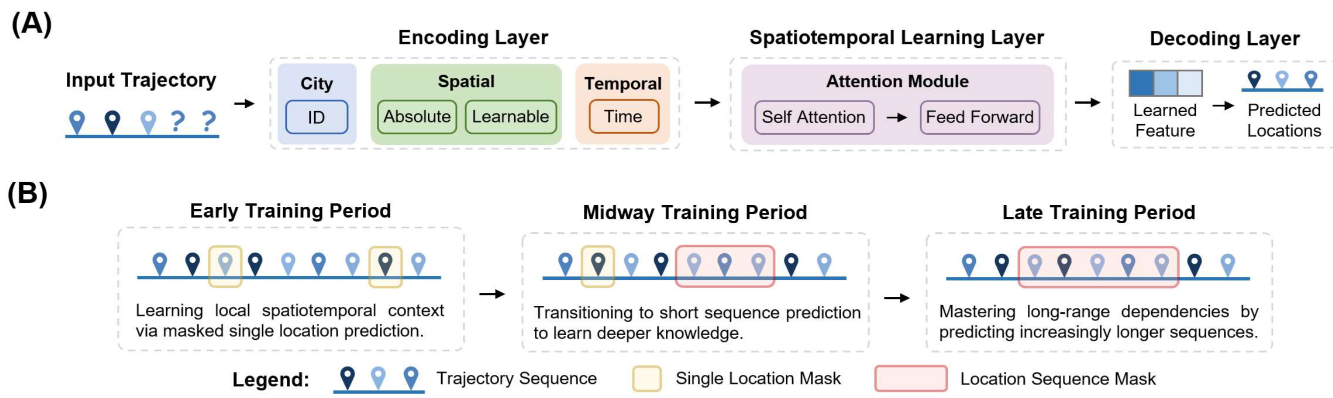

Figure 3. (A) Human Mobility Predictor; (B) Incremental Training Strategy.

Embedding Layer. This explicit representation allows the model to capture city-specific regularities. Spatial Encoding: A key innovation of our method is the combination of Absolute and Learnable spatial encodings. The Absolute Spatial Encoding addresses the limitation of conventional models that lack awareness of geographic proximity. Inspired by positional encoding in Transformers [8], we generate 1D sinusoidal encodings for the X and Y coordinates and concatenate them, assigning a unique, fixed vector to each grid cell, as shown in Figure 4. This ensures that geographically adjacent cells are also close in the feature space. The Learnable Spatial Encoding, implemented via a standard embedding layer, allows the model to learn functional relationships between locations from the data in a data-driven manner. Temporal Encoding: We encode multiple dimensions of time, including the day of the week, the time of day, and the time interval from the previous point. This enables the model to capture daily and weekly rhythms and handle data sparsity.

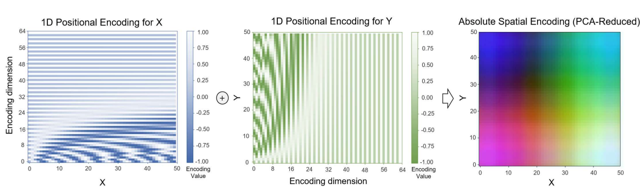

Figure 4. Visualization of Absolute Spatial Encoding.

2) Spatiotemporal Learning Layer. The encoded feature sequence is fed into this layer, which is based on the classic Transformer architecture. It leverages the core multi-head self-attention to efficiently capture complex, long-range spatiotemporal dependencies between any two points in the trajectory.

3) Decoding Layer. The output from the Spatiotemporal Learning Layer is passed to the Decoding Layer. We use two independent MLPs to process the feature for each timestep, predicting the probability distributions over the X and Y coordinates, respectively, to generate the final trajectory.

3.2 Incremental Training Strategy

To tackle the challenges of training on multi-city datasets, we devise an innovative incremental training strategy, as depicted in Figure 3(B). Drawing inspiration from human learning, which progresses from simple to complex, this approach enhances training stability and predictive accuracy by dynamically adjusting task types and complexity. The strategy employs a dynamic masking generator with two modes, Single Location Mask and Sequence Mask. During each epoch, the generator recalibrates its parameters, gradually reducing the use of single location mask while increasing both the frequency and length of sequence mask, thus evolving the model's prediction tasks.

In the early training period, the model focuses on masked single location prediction, where the generator randomly masks isolated time points in the trajectory sequence. This straightforward task enables the model to learn fundamental local spatiotemporal patterns. As training advances to midway, the emphasis shifts, with single-location tasks diminishing and sequence prediction tasks

Table 1. Result of different models.

Model	Hidden dim /heads/layers	Training	Validation	City							
				A		B		C		D	
				GBLEU	DTW	GBLEU	DTW	GBLEU	DTW	GBLEU	DTW
1	128/4/4	2025(A,B,C,D)	Private	0.117	39.79	0.148	36.48	0.153	31.04	0.152	29.63
2	128/4/4	2025(A,B)/(C,D)	Private	0.118	39.78	0.15	36.26	0.155	30.47	0.153	29.47
3	256/8/8	2024	Private	0.121	39.57	0.15	36.75	0.15	33.08	0.15	31.36
4	256/8/8	2025(A,B)/(C,D)	Private	0.127	38.14	0.16	34.56	0.167	29.59	0.163	28.56
5	256/8/8	2025(A,B)/(C,D)	Public	0.152	-	0.149	-	0.184	-	0.177	-

Note: Public validation set (last 3,000 users of 2025 dataset), private set (preceding 1,000 users). GBLEU stand for GEO-BLEU [9].

increasing. The system begins masking short, continuous trajectory segments, encouraging the model to capture short-range spatiotemporal dependencies. In the late training stages, long sequence prediction dominates, with the generator prioritizing longer masked segments, driving the model to master complex, long-range spatiotemporal relationships critical for the ultimate prediction task.

4 Experiment

4.1 Experimental Setting

We train our models using the GISCUP 2025 and HuMob 2024 datasets, abbreviated as 2025 and 2024, respectively. Model validation is performed exclusively on the 2025 dataset. As the official validation results for GISCUP 2025 have not yet been released, we created a private validation set for each city, consisting of 1,000 users (from the last 4,000 to the last 3,000). The official validation set, in contrast, is designated as the public set. We employed the Cross-Entropy Loss function for training. Our evaluation metrics include GEO-BLEU [9] (with Beta=0.5, n=5) and DTW.

4.2 Main Results

Table 1 summarizes the primary results of HuMob Predictor with different settings. We observe a performance gap between models trained on the 2024 dataset (Model 3) and the 2025 dataset (Model 4), indicating the challenge of cross-city generalization. Furthermore, to address a temporal data gap in cities C and D, we test a city-pair training strategy (Model 2) against a unified model (Model 1). As shown Table 1, the performance of Model 1 is slightly lower than that of Model 2, as the unified model's performance is likely influenced by learning from the conflicting patterns of continuous (A, B) and disjointed (C, D) data. By training on pairs, Model 2 avoids this issue, achieving better performance and confirming that tailored training is effective for handling such data irregularities. Finally, comparing the full-scale model (Model 4) with its smaller counterpart (Model 2), the results show that increased model capacity leads to a notable improvement in performance across most metrics. This highlights HuMob Predictor's ability to leverage a larger parameter space for this task.

4.3 Effect of Absolute Spatial Encoding

To validate the effectiveness of our proposed Absolute Spatial Encoding, we conducted an ablation study in our previous research [6]. We compare our model against a baseline variant that omits this component. As shown in Figure 5, the inclusion of Absolute Spatial Encoding yields a substantial improvement in the GEO-BLEU score, confirming its crucial role in enhancing accuracy by providing the model with explicit knowledge of the underlying geography.

5 Conclusion

In this paper, we address the complex problem of multi-city human mobility prediction by proposing a generalizable framework, the HuMob Predictor. Our work introduces two principal contributions to tackle the core challenges of spatiotemporal heterogeneity and large-scale training instability. First, our multi-level encoding module provides the model with essential structural cognition of urban spaces. Second, our incremental training strategy successfully stabilizes the training process on highly complex datasets by adopting a curriculum that progresses from simple to difficult tasks. Together, these innovations create a robust framework that, as validated in the GISCUP 2025 competition, sets a new benchmark for large-scale, multi-city mobility prediction. Future efforts will involve a detailed summarization of the current research and the completion of additional, in-depth experiments.

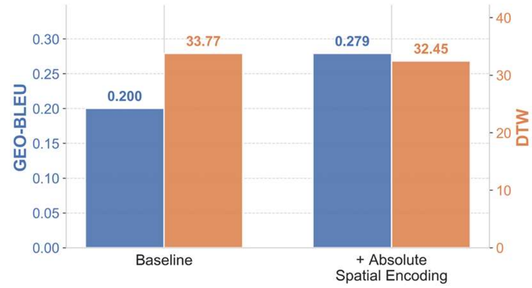

Figure 5. Effect of Absolute Spatial Encoding.

ACKNOWLEDGMENTS

We are grateful to the teammates from the HuMob Challenge 2023 for the invaluable experience they provided. We also thank the organizers of GISCUP 2025 for their efforts and Dr. Takahiro Yabe for providing the dataset and support throughout the competition.

REFERENCES

[1] Barbosa, H., Barthelemy, M., Ghoshal, G., James, C. R., Lenormand, M., Louail, T., Menezes, R., Ramasco, J. J., Simini, F. and Tomasini, M. Human mobility: Models and applications. *Physics Reports*, 734 (2018), 1-74.

[2] Li, G., Zhao, Z., Guo, X., Tang, L., Zhang, H. and Wang, J. Towards integrated and fine-grained traffic forecasting: A spatio-temporal heterogeneous graph transformer approach. *Information Fusion*, 102 (2024), 102063.

[3] Yang, H., Yao, X. A., Whalen, C. C. and Kiwanuka, N. Exploring human mobility: a time-informed approach to pattern mining and sequence similarity. *International Journal of Geographical Information Science* (2024), 1-25.

[4] Perofsky, A. C., Hansen, C. L., Burstein, R., Boyle, S., Prentice, R., Marshall, C., Reinhart, D., Capodanno, B., Truong, M. and Schwabe-Fry, K. Impacts of human mobility on the citywide transmission dynamics of 18 respiratory viruses in pre-and post-COVID-19 pandemic years. *Nature communications*, 15, 1 (2024), 4164.

[5] Yabe, T., Tsubouchi, K., Shimizu, T., Sekimoto, Y., Sezaki, K., Moro, E. and Pentland, A. YJMob100K: City-scale and longitudinal dataset of anonymized human mobility trajectories. *Scientific Data*, 11, 1 (2024), 397.

[6] Li, G., Xu, Y., Gui, Z., Guo, X. and Tang, L. Striking a balance between diversity and regularity: a preference-guided transformer for individual mobility prediction. *International Journal of Geographical Information Science* (2025), 1-32.

[7] Terashima, H., Takagi, S., Tamura, N., Shoji, K., Hossain, T., Katayama, S., Urano, K., Yonezawa, T. and Kawaguchi, N. *Time-series Stay Frequency for Multi-City Next Location Prediction using Multiple BERTs*. City, 2024.

[8] Vaswani, A. Attention is all you need. *Advances in Neural Information Processing Systems* (2017).

[9] Shimizu, T., Tsubouchi, K. and Yabe, T. *GEO-BLEU: similarity measure for geospatial sequences*. City, 2022.