Human Mobility Prediction with Multi-Task Curriculum Training

Tianye Fang
Technical University of Munich
School of Engineering and Design
Munich, Bavaria, Germany
tianye.fang@tum.de

Xuanshu Luo
Technical University of Munich
School of Engineering and Design
Munich, Bavaria, Germany
xuanshu.luo@tum.de

Paul Walther
Technical University of Munich
School of Engineering and Design
Munich, Bavaria, Germany
paul.walther@tum.de

Martin Werner
Technical University of Munich
School of Engineering and Design
Munich, Bavaria, Germany
martin.werner@tum.de

Abstract

Effective human mobility modeling and prediction constitute the core prerequisites for various location-based applications. To encourage research in this direction, the ACM SIGSPATIAL Cup 2025 posed the challenge of predicting human mobility trajectories from a sparse multi-city dataset. This paper presents our solution, MoBERT, a BERT-like model that adapts and leverages mobility semantics with additional direction and distance prediction, providing supplementary supervision signals for robust feature learning. MoBERT models are trained in stages through curriculum learning, where augmented trajectories are ordered by increasing mobility entropy for training with progressively increasing difficulty. The final score of our method is 0.14609, as measured by average GEO-BLEU distances across four cities. Finally, we analyze the results and discuss insights from our approach.

CCS Concepts

• **Information systems → Location based services**; • **Theory of computation → Machine learning theory**; • **Mathematics of computing →** *Information theory*; • **Applied computing**; • **Computing methodologies → Model development and analysis**; **Artificial intelligence**;

Keywords

Human Mobility Prediction, Curriculum Learning, Entropy, Multi-Task Learning, Transformer, GISCUP

ACM Reference Format:

Tianye Fang, Xuanshu Luo, Paul Walther, and Martin Werner. 2025. Human Mobility Prediction with Multi-Task Curriculum Training. 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/10.1145/3748636.3771316

SIGSPATIAL '25, Minneapolis, MN, USA

ACM ISBN 979-8-4007-2086-4/2025/11
https://doi.org/10.1145/3748636.3771316

1 Introduction

Mobility has played a fundamental role in human society throughout history. From ancient hunter-gatherer societies where movement was essential for survival, to contemporary urban environments where people exhibit unprecedented levels of spatial mobility, human movement patterns have remained central to societal organization and individual life experiences. Modern human mobility often follows observable patterns, such as daily work commutes, exhibiting acknowledged predictability [16] that has encouraged substantial research interest across multiple domains, spanning market research, urban planning, transportation optimization, disaster response management, and public health interventions.

Despite certain regularities represented in human mobility, accurate next-location prediction in practice is a non-trivial decision-making process involving intricate combinations of social interactions, personal preferences, and geographic constraints. Thanks to the large-scale proliferation of ubiquitous user devices passively collecting location information, vast amounts of human mobility data have become available, enabling methods based on deep learning models to effectively understand complex human mobility patterns. Specifically, sequential architectures, such as recurrent neural networks and Transformers [8, 18], are preferred choices as they are suitable for modeling chronological mobility data [11, 17, 20]. There are also approaches using convolutional/graph neural networks for their excellence in local spatial pattern extraction [5, 13].

To establish a unified benchmark for promoting research in human mobility prediction, the HuMob Challenge was founded as workshops at ACM SIGSPATIAL conferences in 2023 [1, 2]. In 2025, it serves as the ACM SIGSPATIAL Cup (GISCUP 2025)[1]. This challenge calls for innovative approaches to human mobility prediction based on a large-scale, longitudinal, and discretized dataset collected in four Japanese metropolitan areas, with prescriptive train/test split and evaluation metrics [14, 19].

This paper presents the method utilized in our submission. We design MoBERT, an encoder-only BERT-based model for human mobility prediction [4]. In addition to essential location estimation heads, MoBERT also incorporates two auxiliary predictions for distances and directions to provide complementary supervision signals for multi-task learning (MTL) [3], aiming to obtain comprehensive representations of mobility data. MoBERT is trained in

[1] https://sigspatial2025.sigspatial.org/giscup/index.html (last accessed: 27.09.2025)

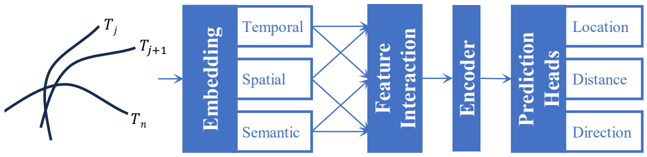

Figure 1: The proposed MoBERT architecture.

a curriculum learning fashion, where training data is ordered by increasing difficulty to learn. Concretely, the predictability of trajectories is quantified by the proposed mobility entropy estimator based on Lempel-Ziv compression [21].

Our submission achieves an average GEO-BLEU score [14] of 0.14609 in the GISCUP 2025. The proposed method was originally designed for the HuMob Challenge 2023 [1] and is capable of outperforming other solutions. We try to analyze potential reasons for the performance degradation for the GISCUP 2025 and conclude that MTL as a regularizer may impede convergence due to the increasing temporal sparsity of data, resulting in model underfitting.

2 The GISCUP 2025

The GISCUP 2025 focuses on human mobility prediction, aiming to estimate future locations given corresponding historical trajectory data, where a trajectory contains a series of points p_i with orders, e.g., by timestamps t_i. A trajectory T of length n can be formulated as $T_{1:n} = \{(p_1, t_1), (p_2, t_2), \dots, (p_n, t_n)\}$. Predicting the next location $\hat{p}_{n+1}$ is to maximize the conditional probability $\mathbb{P}$ given $T_{1:n}$, i.e.,

$$\hat{p}_{n+1} = \arg\max_{p} \mathbb{P}[(p, t) \mid T_{1:n}].$$

Similarly, we can also perform multistep prediction by maximizing $\mathbb{P}[\hat{T}_{n+1:n+k} \mid T_{1:n}]$, where k is the prediction horizon.

The data for this year's GISCUP are extended from a previously published dataset, YJMob100k [19], synthetically constructed from users' smartphone location data in four Japanese cities. Trajectories are discretized using a spatial grid of size 200×200 with a resolution of $500\,m\times500\,m$ per cell. A temporal discretization is also performed at 30 min intervals. Each grid cell also includes spatial semantics and context represented by a list of Points of Interest (POI) categories. Together with the dataset, a standardized evaluation metric based on GEO-BLEU is used for the competition [15].

3 Methodology

This section elaborates on the design details of MoBERT and the training procedure based on curriculum learning.

3.1 Model Architecture

The overall architecture of the proposed MoBERT is depicted in Figure 1. We explain every building block in the following sections.

3.1.1 Input Features. The model input comprises eight features that collectively capture temporal, spatial, and semantic dimensions. Spatially, each trajectory point is represented by its two-dimensional grid coordinates ① x and ② y in the discretized space. Temporal information encompasses four features: ③ the day index, ④ the time slot index within a day, ⑤ the day of the week, and ⑥ the time interval between consecutive trajectory points to model temporal continuity. Additionally, we incorporate ⑦ a binary day/night indicator to distinguish between diurnal and nocturnal movements. The semantic dimension is represented by ⑧ the top-k most frequent POI categories associated with each grid cell as a list. According to our experiments on City A, the optimal value of k is determined to be 3, and this setting is retained in our submission. This multi-faceted feature construction enables the model to learn comprehensive spatiotemporal and semantic patterns in the data.

3.1.2 Feature Interaction. To effectively leverage the eight input features, we design a feature interaction module that progressively fuses information across different modalities. Each raw feature is first independently encoded through a dedicated embedding layer, transforming values into dense vector representations of identical dimensionality. These eight feature embeddings are then processed through two parallel pathways. The first pathway performs basic fusion by directly summing all feature embeddings element-wise, preserving individual semantic contributions. The second pathway employs multi-head self-attention operating across the feature dimension to dynamically learn inter-feature dependencies, computing adaptive weights that emphasize relevant feature combinations based on context. The outputs from both pathways are combined through element-wise summation, yielding a unified representation that maintains individual feature semantics while incorporating learned relational patterns, thereby enhancing the model's capacity to understand complex mobility behaviors.

3.1.3 Backbone. The sequential nature of mobility data motivates the application of recurrent architectures [9], which have emerged as predominant approaches for trajectory modeling, as their recurrent connections naturally accommodate variable-length sequences and adapt to temporal patterns [10, 12]. However, they face inherent limitations, such as susceptibility to vanishing gradients and constrained parallelization capabilities. Alternative paradigms have been explored to address different challenges. Convolutional approaches focus on extracting localized spatiotemporal patterns but struggle with modeling global dependencies and generalizing across diverse spatial contexts [13], while graph-based methods excel at representing spatial relationships through network structures yet often lack temporal modeling mechanisms and comprehensive spatial understanding [5]. In comparison, Transformer-based architectures have gained prominence in human mobility prediction due to their ability to simultaneously capture long-range spatiotemporal dependencies through self-attention mechanisms while enabling parallelism [18]. Transformers typically demand substantial training data volumes. As the GISCUP 2025 utilizes a large-scale dataset, Transformer-based solutions are thereby viable [11, 17].

Given these considerations, the proposed MoBERT adopts a BERT-style encoder-only architecture for its distinctive advantages in mobility prediction tasks [4]. Unlike autoregressive decoder models that generate predictions sequentially and accumulate errors over forecasting horizons, BERT's bidirectional encoding enables parallel prediction of multiple future locations while leveraging contextual information from the entire historical trajectory, thereby preventing error propagation and naturally accommodating the sparse, irregularly-sampled nature of real-world mobility data. To deal with trajectories of different lengths, a padding mask is used.

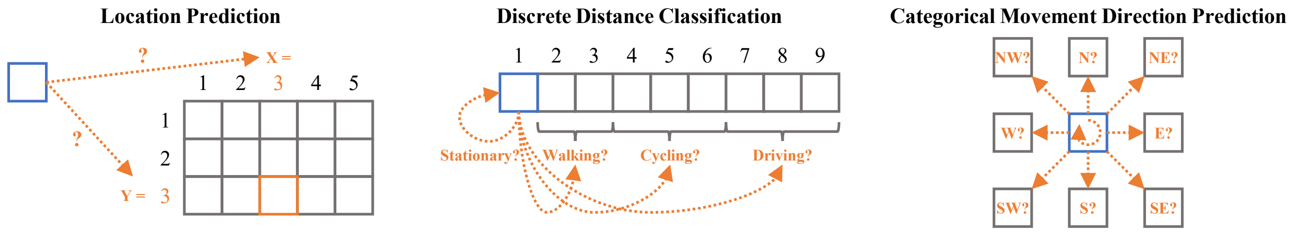

Figure 2: Visualization of the three prediction goals: next location, distance classification, and movement direction

3.1.4 Multi-task prediction. To enhance the primary location prediction task, we incorporate multi-task learning by introducing two auxiliary tasks that provide complementary supervision signals. The motivation stems from the observation that human mobility inherently involves multifaceted decision-making processes where individuals simultaneously consider destinations, movement scales, and directional preferences based on urban structure and personal routines. Therefore, we introduce distance and direction prediction as auxiliary tasks, which are universally derivable from any trajectory dataset without additional annotations, ensuring broad applicability across different mobility datasets. MoBERT produces outputs for different tasks with different prediction heads.

All prediction goals are visualized in Figure 2. The primary location prediction is formulated as a classification task and estimate by direction to reduce the number of classes from $200 \times 200 = 40000$ to $200 + 200 = 400$ in total. This approach leads to a more homogeneous class frequency, allowing for easier learning compared to the traditional grid cell prediction. For the distance prediction task, we discretize the Euclidean distance between consecutive trajectory points into four classes representing typical movement scales:

$$
\begin{aligned}
d_x &\in [0, r_1) && \textit{stationary} \\
&\in [r_1, r_2) && \textit{walking} \\
&\in [r_2, r_3) && \textit{cycling} \\
&\in [r_3, +\infty) && \textit{driving}
\end{aligned}
$$

The parameters r_1, r_2, and r_3 are determined empirically depending on the grid size of the dataset. The direction prediction task discretizes geographic orientation into nine classes comprising four cardinal directions, four ordinal directions, and a stationary class. Both auxiliary tasks are formulated as classification problems consistent with the primary location prediction task.

To ensure the backpropagation still working jointly for end-to-end training, we designed a weighted total loss $\mathcal{L}$ consisting of a linear combination of losses for each prediction head:

$$\mathcal{L} = \mathcal{L}_{Location} + \lambda_1 \mathcal{L}_{Distance} + \lambda_2 \mathcal{L}_{Direction}$$

The weights λ_1 and λ_2 are empirically determined with a grid search in the interval $[0, 1]$. This multi-task formulation acts as an implicit regularizer, encouraging the model to learn more robust and generalizable representations by considering multiple complementary aspects of human mobility simultaneously.

3.2 Training Procedures

We propose an entropy-driven curriculum learning strategy to address the heterogeneous complexity in human mobility data by organizing training progression from simple to complex patterns. Trajectory predictability is quantified using normalized Lempel-Ziv compression entropy, which measures the rate at which new movement patterns emerge in a sequence. This metric is grounded in Fano's inequality [7], establishing that lower-entropy trajectories are fundamentally more learnable due to their higher regularity and deterministic patterns. The trajectory entropy is quantified by the proposed normalized Lempel-Ziv entropy estimator. Concretely, trajectories are first symbolized by flattening coordinates (x, y) through the transformation $x * (y_{max} + 1) + y$. The Lempel-Ziv algorithm [21] then parses the symbolized trajectory, maintaining a dictionary of observed subsequences and iteratively identifying the shortest previously unseen subsequence at each position. The entropy is estimated as $H_{LZ} = (\ln N)/(\bar{Q} \ln 2)$, where N is the total sequence length and $\bar{Q}$ is the average phrase length, then normalized by dividing by $\log_2 N$ to yield values between zero (highly predictable) and one (random patterns).

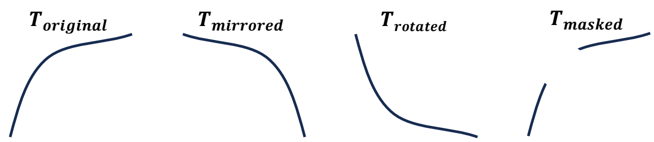

Figure 3: Augmentation of trajectory samples (from left to right: original trajectory, mirrored, rotated, and masked).

The curriculum pipeline begins with trajectory augmentation through horizontal mirroring, vertical mirroring, 180-degree rotation, and masking, as shown in Figure 3. Augmented trajectories are sorted by increasing normalized entropy and assigned progressively longer prediction horizons. Based on the dataset's entropy distribution, curriculum pretraining proceeds through three stages: trajectories with entropy below 0.4 and three-day prediction horizon, entropy below 0.65 with seven-day horizon, and finally all augmented data with the complete fifteen-day horizon. Following curriculum pretraining, the model undergoes finetuning exclusively on original trajectories for optimal adaptation to authentic mobility characteristics. This strategy enables the model to establish robust foundational representations before gradually adapting to increasingly irregular and unpredictable movement patterns.

4 Results

The achieved scores measured by GEO-BLEU in the GISCUP 2025 are presented in Table 1. We originally developed the proposed method under the setting aligned with the HuMob Challenge 2023

City A	City B	City C	City D	Mean
0.13796	0.13767	0.15690	0.15183	0.14609

Table 1: Evaluation results measured by GEO-BLEU.

[1], achieving a GEO-BLEU score of 0.354 and a DTW distance of 26.15 [6], which outperformed other solutions in the challenge according to the public leaderboard[2].

However, the performance of the proposed method on the GISCUP 2025 dataset is unsatisfactory. We compare the temporal sparsity of the GISCUP 2025 dataset and the HuMob Challenge 2023 version, where the sparsity is calculated as the ratio of time slots with data to the total number of time slots within a day. In this way, we find that the GISCUP 2025 dataset is more sparse than the 2023 version, leading to a smaller number of points per trajectory as well as longer time intervals between trajectory points on average. Therefore, this year's dataset is significantly more difficult to learn. Consequently, the MTL strategy as a regularizer impedes model convergence. A possible modification is to remove MTL in the early stages of training and re-introduce it during finetuning, further aligning with the idea of curriculum learning.

5 Conclusion

This paper presents MoBERT, a BERT-based encoder architecture designed for the human mobility prediction challenge posed by the ACM SIGSPATIAL Cup 2025. MoBERT incorporates multi-task learning for complementary supervision signals to enhance the primary location prediction objective. We also propose an entropy-driven curriculum learning strategy that progressively trains the model on trajectories ordered by increasing predictability, as quantified through normalized Lempel-Ziv compression entropy. Our method achieves a mean GEO-BLEU score of 0.14609 for the GISCUP 2025. While the approach demonstrated superior performance on the HuMob Challenge 2023 dataset, we observe performance degradation for the GISCUP 2025, which we attribute to increasingly sparse trajectories. This analysis suggests that the multi-task learning regularization may impede convergence under such sparse data conditions, indicating potential refinements such as adaptive integration of auxiliary tasks during different training phases.

Acknowledgments

This work was supported by the Federal Ministry of Research, Technology, and Space (BMFTR) of Germany under grant number 16DHBKI020 and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under grant number 507196470. For further details of our method, please refer to [6].

References

[1] 2023. *HuMob-Challenge '23: Proceedings of the 1st International Workshop on the Human Mobility Prediction Challenge* (Hamburg, Germany). Association for Computing Machinery, New York, NY, USA.

[2] 2024. *HuMob'24: Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Human Mobility Prediction Challenge* (Atlanta, GA, USA). Association for Computing Machinery, New York, NY, USA.

[3] Rich Caruana. 1997. Multitask Learning. *Machine Learning* 28, 1 (01 Jul 1997), 41–75. doi:10.1023/A:1007379606734

[4] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, Jill Burstein, Christy Doran, and Thamar Solorio (Eds.). Association for Computational Linguistics, Minneapolis, Minnesota, 4171–4186. doi:10.18653/v1/N19-1423

[5] Jiaxin Du and Xinyue Ye. 2023. Batch and negative sampling design for human mobility graph neural network training. In *Proceedings of the 1st International Workshop on the Human Mobility Prediction Challenge (ACM Digital Library)*. Association for Computing Machinery, Erscheinungsort nicht ermittelbar, 47–50. doi:10.1145/3615894.3628504

[6] Tianye Fang, Xuanshu Luo, and Martin Werner. 2025. Entropy-Driven Curriculum for Multi-Task Training in Human Mobility Prediction. arXiv:2509.01613 [cs.LG] https://arxiv.org/abs/2509.01613

[7] Te Sun Han and S. Verdu. 1994. Generalizing the Fano inequality. *IEEE Transactions on Information Theory* 40, 4 (1994), 1247–1251. doi:10.1109/18.335943

[8] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long Short-Term Memory. *Neural Computation* 9, 8 (1997), 1735–1780. doi:10.1162/neco.1997.9.8.1735

[9] S. Hochreiter and J. Schmidhuber. 1997. Long short-term memory. *Neural computation* 9, 8 (1997), 1735–1780. doi:10.1162/neco.1997.9.8.1735

[10] Renhe Jiang, Xuan Song, Zipei Fan, Tianqi Xia, Quanjun Chen, Satoshi Miyazawa, and Ryosuke Shibasaki. 2018. DeepUrbanMomentum: An Online Deep-Learning System for Short-Term Urban Mobility Prediction. *Proceedings of the AAAI Conference on Artificial Intelligence* 32, 1 (Apr. 2018). doi:10.1609/aaai.v32i1.11338

[11] Akihiro Kobayashi, Naoto Takeda, Yudai Yamazaki, and Daisuke Kamisaka. 2023. Modeling and generating human mobility trajectories using transformer with day encoding. In *Proceedings of the 1st International Workshop on the Human Mobility Prediction Challenge (ACM Digital Library)*. Association for Computing Machinery, Erscheinungsort nicht ermittelbar, 7–10. doi:10.1145/3615894.3628506

[12] Qiang Liu, Shu Wu, Liang Wang, and Tieniu Tan. 2016. Predicting the Next Location: A Recurrent Model with Spatial and Temporal Contexts. *Proceedings of the AAAI Conference on Artificial Intelligence* 30, 1 (Feb. 2016). doi:10.1609/aaai.v30i1.9971

[13] Bilong Shen, Xiaodan Liang, Yufeng Ouyang, Miaofeng Liu, Weimin Zheng, and Kathleen M. Carley. 2018. StepDeep: A Novel Spatial-temporal Mobility Event Prediction Framework based on Deep Neural Network. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (London, United Kingdom) *(KDD '18)*. Association for Computing Machinery, New York, NY, USA, 724–733. doi:10.1145/3219819.3219931

[14] Toru Shimizu, Kota Tsubouchi, and Takahiro Yabe. 2022. GEO-BLEU: similarity measure for geospatial sequences. In *Proceedings of the 30th International Conference on Advances in Geographic Information Systems* (Seattle, Washington) *(SIGSPATIAL '22)*. Association for Computing Machinery, New York, NY, USA, Article 17, 4 pages. doi:10.1145/3557915.3560951

[15] Toru Shimizu, Kota Tsubouchi, and Takahiro Yabe. 2022. GEO-BLEU: similarity measure for geospatial sequences. In *Proceedings of the 30th International Conference on Advances in Geographic Information Systems (SIGSPATIAL '22)*. Association for Computing Machinery, New York, NY, USA. doi:10.1145/3557915.3560951

[16] Chaoming Song, Zehui Qu, Nicholas Blumm, and Albert-László Barabási. 2010. Limits of predictability in human mobility. *Science (New York, N.Y.)* 327, 5968 (2010), 1018–1021. doi:10.1126/science.1177170

[17] Haru Terashima, Naoki Tamura, Kazuyuki Shoji, Shin Katayama, Kenta Urano, Takuro Yonezawa, and Nobuo Kawaguchi. 2023. Human Mobility Prediction Challenge: Next Location Prediction using Spatiotemporal BERT. In *Proceedings of the 1st International Workshop on the Human Mobility Prediction Challenge (ACM Digital Library)*. Association for Computing Machinery, Erscheinungsort nicht ermittelbar, 1–6. doi:10.1145/3615894.3628498

[18] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. 2017. Attention is All you Need. In *Advances in Neural Information Processing Systems*, I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.), Vol. 30. Curran Associates, Inc.

[19] Takahiro Yabe, Kota Tsubouchi, Toru Shimizu, Yoshihide Sekimoto, Kaoru Sezaki, Esteban Moro, and Alex Pentland. 2024. YJMob100K: City-scale and longitudinal dataset of anonymized human mobility trajectories. *Scientific Data* 11, 1 (18 Apr 2024), 397. doi:10.1038/s41597-024-03237-9

[20] Dingqi Yang, Benjamin Fankhauser, Paolo Rosso, and Philippe Cudre-Mauroux. 2020. Location Prediction over Sparse User Mobility Traces Using RNNs: Flashback in Hidden States!. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (IJCAI-20)*, Christian Bessiere (Ed.). International Joint Conferences on Artificial Intelligence, Erscheinungsort nicht ermittelbar, 2184–2190. doi:10.24963/ijcai.2020/302

[21] J. Ziv and A. Lempel. 1977. A universal algorithm for sequential data compression. *IEEE Transactions on Information Theory* 23, 3 (1977), 337–343. doi:10.1109/TIT.1977.1055714

[2] https://connection.mit.edu/humob-challenge-2023/ (last accessed: 29.09.2025)