BERT4Traj: Transformer-Based Trajectory Reconstruction for Sparse Mobility Data

Hao Yang
Department of Geography, University of Georgia, Athens, GA, USA

Angela Yao[1]
Department of Geography, University of Georgia, Athens, GA, USA

Christopher C. Whalen
College of Public Health, University of Georgia, Athens, GA, USA

Gengchen Mai
Department of Geography and the Environment, University of Texas at Austin, TX, USA

Abstract

Understanding human mobility is essential for applications in public health, transportation, and urban planning. However, mobility data often suffers from sparsity due to limitations in data collection methods, such as infrequent GPS sampling or call detail record (CDR) data that only capture locations during communication events. To address this challenge, we propose BERT4Traj, a transformer-based model that reconstructs complete mobility trajectories by predicting hidden visits in sparse movement sequences. Inspired by BERT's masked language modeling objective and self-attention mechanisms, BERT4Traj leverages spatial embeddings, temporal embeddings, and contextual background features such as demographics and anchor points. We evaluate BERT4Traj on real-world CDR and GPS datasets collected in Kampala, Uganda, demonstrating that our approach significantly outperforms traditional models such as Markov Chains, KNN, RNNs, and LSTMs. Our results show that BERT4Traj effectively reconstructs detailed and continuous mobility trajectories, enhancing insights into human movement patterns.

2012 ACM Subject Classification Computing methodologies → Machine learning

Keywords and phrases Human Mobility, Trajectory Reconstruction, Deep Learning, CDR, GPS

Digital Object Identifier 10.4230/LIPIcs.GIScience.2025.8

1 Introduction

Understanding human mobility is crucial for various applications, including public health, transportation, and urban planning [12, 2, 15]. With the increasing availability of location data from GPS devices, mobile phones, and other portable technologies, human mobility analysis has gained significant attention. However, despite the abundance of location data, data sparsity remains a persistent challenge. For example, Call Detail Records (CDRs) capture locations only when calls or text messages occur, leaving significant gaps in a user's movement trajectory [4]. Similarly, GPS data may be sparse due to battery-saving modes, signal loss, or intermittent sampling. Consequently, there exist places that individuals have visited but are not recorded in the data, which we refer to as "hidden visits" [1]. The presence of hidden visits impedes the ability to reconstruct a complete view of an individual's daily movement, posing substantial challenges to understanding human mobility. Thus, identifying hidden visits to address data sparsity and reconstructing continuous, detailed, and complete mobility trajectories is a necessary and meaningful research problem.

[1] Corresponding Author

13th International Conference on Geographic Information Science (GIScience 2025).
Editors: Katarzyna Sila-Nowicka, Antoni Moore, David O'Sullivan, Benjamin Adams, and Mark Gahegan; Article No. 8; pp. 8:1–8:9
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Early trajectory reconstruction methods primarily relied on Markov Chains and interpolation-based techniques [6, 16, 8, 7]. Markov Chain models, such as the one proposed by [8], incorporate activity changes to enhance mobility prediction but struggle with long-range dependencies and complex movement behaviors due to the adopted Markov assumption. Interpolation-based approaches leverage spatial-temporal correlations to estimate missing points. For instance, [7] use linear and cubic interpolation to reconstruct human mobility from mobile phone data. However, such assumptions fail to capture real-world non-linear travel patterns effectively.

With advancements in machine learning, researchers have increasingly adopted deep learning models for trajectory reconstruction [3, 9, 10, 14]. Backpropagation (BP) neural networks, as proposed by [10], reconstruct mobility trajectories from sparse Call Detail Records (CDR) to estimate hourly population density. However, this method assumes predictable movement patterns, overlooking detours and irregular trajectories.

More recently, Transformer-based approaches have demonstrated superior performance [13, 5]. TrajBERT, introduced by [13], applies BERT-based trajectory recovery with spatial-temporal refinement to address implicit trajectory sparsity. While effective in predicting missing locations, TrajBERT lacks external context modeling, such as user characteristics, dynamic temporal variations, or real-world events, limiting its adaptability for high-accuracy trajectory prediction.

To overcome these challenges, this paper introduces **BERT4Traj**, a novel Transformer-based model for trajectory reconstruction. By leveraging BERT's bidirectional self-attention mechanism, BERT4Traj effectively captures spatial-temporal dependencies, improving trajectory prediction accuracy. The model is applied to reconstruct complete movement trajectories from both CDR and GPS datasets, demonstrating its robustness in handling data sparsity across different mobility data types. Through context-aware trajectory reconstruction, BERT4Traj enables a more detailed and accurate representation of human mobility patterns, offering valuable insights for applications in public health, urban planning, and transportation analytics.

2 Methodology

To address the challenge of data sparsity and reconstruct continuous and detailed mobility trajectories, we propose a transformer-based architecture, BERT4Traj, inspired by BERT. This model predicts hidden visits in user mobility trajectories by treating each user's daily trajectory as a sequence analogous to a sentence in Natural Language Processing (NLP), where locations correspond to ordered words. The objective is to infer missing locations within this sequence using spatial, temporal, and user-specific demographic information.

The core idea of BERT4Traj is inspired by the masked language modeling in BERT, where predictions are made based on contextual information from surrounding tokens. In the context of human mobility, locations visited on the same day provide contextual clues to infer missing visits. In addition to known locations, background information such as demographic characteristics (e.g., age, gender), key life anchors (e.g., home and workplace), and temporal context (e.g., weekday vs. weekend, holidays) further enrich the representation of an individual's mobility behavior.

As illustrated in Figure 1, BERT4Traj incorporates a BERT-like masking and prediction mechanism. A subset of locations in a trajectory sequence is randomly masked – for example, location P2 in the figure – and the model learns to predict these missing locations using the context provided by the rest of the sequence. This bidirectional prediction process enables BERT4Traj to develop a deep understanding of how visited locations relate to one another in varying contexts.

The input sequence consists not only of the trajectory data but also of unmasked context tokens that provide additional background information, including temporal attributes, user demographics, and travel characteristics. During training, the model learns intricate relationships between visited locations and contextual features, allowing it to accurately predict missing locations at specific times in a day. Ultimately, this approach reconstructs an individual's movement trajectory with finer temporal granularity, effectively addressing data sparsity issues in mobility datasets.

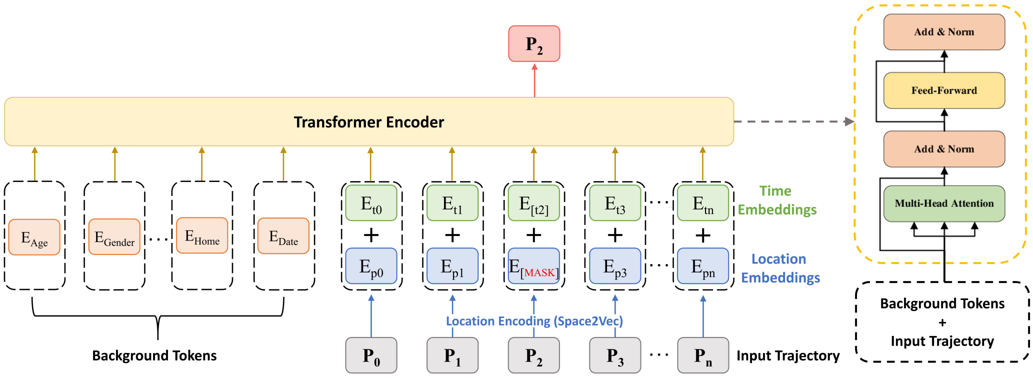

Figure 1 The overall framework of the BERT4Traj model.

2.1 Data Representation and Embeddings

Each user's daily trajectory consists of a sequence of visited locations with corresponding timestamps. To represent this information in the model, we define location embeddings and time embeddings which are analogy word embeddings and position embeddings in the classic BERT model. Each visited location $\mathbf{x_i} \in \mathbb{R}^2$ is embedded as a vector of dimension d, i.e., $\mathbf{l}_i \in \mathbb{R}^d$. Here, i is the index of the location in the trajectory. The location embeddings encode geographical information (latitude, longitude) and may also include semantic attributes, such as Points of Interest (POI) types or travel modes, if available.

To model temporal dependencies, a time embedding $\mathbf{t}_i \in \mathbb{R}^d$ is generated based on the timestamp capturing the time of the visit. The time embedding functions similarly to positional embeddings in NLP models, providing temporal context to the trajectory sequence.

In addition to trajectory embeddings, we incorporate background tokens representing demographic features, anchor points, and temporal attributes. Specifically, $\mathbf{B} = [\mathbf{w}_{\text{age}}; \mathbf{w}_{\text{gender}}; \dots]$ represents the demographic embeddings, where $\mathbf{w}_{\text{age}}$ and $\mathbf{w}_{\text{gender}}$ denote the age and gender embeddings, respectively.

Anchor points, such as home and workplace locations, are encoded as $\mathbf{A} = [\mathbf{w}_{\text{primary}}; \mathbf{w}_{\text{secondary}}; \dots]$, where $\mathbf{w}_{\text{primary}}$ and $\mathbf{w}_{\text{secondary}}$ represent embeddings for primary and secondary anchor points.

Temporal context, including whether the day is a weekday, weekend, or holiday, is represented as $\mathbf{T} = [\mathbf{w}_{\text{weekday}}; \mathbf{w}_{\text{weekend}}; \dots]$, where $\mathbf{w}_{\text{weekday}}$ and $\mathbf{w}_{\text{weekend}}$ denote the corresponding time-related embeddings.

To form the final input to the Transformer encoder, we concatenate these background tokens with the spatiotemporal trajectory tokens. Each trajectory token is constructed by performing element-wise addition of the location embedding $\mathbf{l}_i$ and its corresponding time embedding $\mathbf{t}_i$, effectively combining spatial and temporal context into a single vector.

The complete input sequence is formulated as:

$$\mathbf{X} = [\mathbf{B}; \mathbf{A}; \mathbf{T}; \mathbf{l}_1 + \mathbf{t}_1; \mathbf{l}_2 + \mathbf{t}_2; \ldots; \mathbf{l}_n + \mathbf{t}_n]. \tag{1}$$

2.2 Masking Mechanism

A portion of the location tokens in the trajectory sequence is randomly masked. The objective is to predict these masked locations using unmasked locations and contextual embeddings. Let $\mathbf{M} \in \{0, 1\}^n$ be a binary masking vector, where $\mathbf{M}_i = 1$ if the location $\mathbf{l}_i$ is masked and $\mathbf{M}_i = 0$ otherwise. The masked sequence is represented as:

$$\mathbf{X}_{\text{masked}} = [\mathbf{B}; \mathbf{A}; \mathbf{T}; (\mathbf{M}_1 \cdot \mathbf{l}_1) + \mathbf{t}_1; \ldots; (\mathbf{M}_n \cdot \mathbf{l}_n) + \mathbf{t}_n]. \tag{2}$$

This ensures that spatial and temporal relationships are preserved while training the model to infer missing locations.

2.3 Transformer-Based Sequence Encoder

The masked sequence is processed by a Transformer encoder consisting of multiple self-attention layers. The self-attention mechanism computes dependencies between different locations in the trajectory:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \tag{3}$$

where Q, K, V represent the query, key, and value matrices derived from the input sequence, and d_k is the dimensionality of the key vectors.

Multi-head attention further enhances the model's ability to capture diverse mobility patterns:

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \ldots, \text{head}_h)W^O. \tag{4}$$

The output of the encoder is a sequence of hidden states $H = [h_1, h_2, \ldots, h_n]$, where each h_i encodes contextual information about its corresponding location.

2.4 Masked Location Prediction

In our approach, predicting masked locations is formulated as a classification task rather than a regression problem. This is because the model selects the most likely location from a predefined set of discrete spatial units (e.g., tower IDs for CDR data or grid IDs for GPS data) rather than predicting continuous latitude and longitude values. For each masked location $\mathbf{l}_i$, the model predicts its most probable value using the output hidden states:

$$\hat{\mathbf{l}}_i = \text{softmax}(Wh_i), \tag{5}$$

where $W \in \mathbb{R}^{|P| \times d}$ is a weight matrix, and $\hat{\mathbf{l}}_i$ represents the predicted probability distribution over the possible locations P. Since each masked location must be assigned one discrete label from a finite set of locations, this naturally aligns with a multi-class classification problem.

The model is optimized using a cross-entropy loss function:

$$\mathcal{L} = -\sum_{i \in \mathcal{M}} \log \hat{\mathbf{l}}_i[\mathbf{l}_i]. \tag{6}$$

Where M denotes the set of masked locations. Minimizing this loss encourages the model to correctly predict masked locations, improving its ability to reconstruct missing trajectory data.

3 Application of BERT4Traj to CDR and GPS Data

The BERT4Traj framework is highly flexible and can be extended or modified to reconstruct mobility trajectories across different types of mobility data, such as GPS and CDR. Depending on data availability, additional background information can be incorporated to enhance the model's ability to capture mobility behavior and patterns. In this study, to evaluate its effectiveness, we adapted and applied BERT4Traj to two distinct location datasets: Call Detail Records (CDR) collected from 248 cell phone users and GPS data collected from 586 portable watch users in Kampala, Uganda. All participants – both cell phone and watch users – provided self-reported information, including age, gender, anchor points (such as home, workplace, and school locations), education, and income.

3.1 CDR Data

CDR data captures the tower location and timestamp when a communication event occurs, such as a call or text message. Due to its event-driven nature, CDR data is sparse, In our dataset, each individual has records spanning an average of 34 days, but only around five location records per day resulting in significant gaps in their mobility trajectories. To address this, BERT4Traj predicts hidden visits during unrecorded periods, enhancing trajectory completeness.

Each input trajectory is represented as a sequence of tower locations with timestamps. We generate location embeddings using Space2Vec [11], which provides continuous vector representations based on geographical coordinates:

$$e_{loc}(l_i) = \text{Space2Vec}(\mathbf{x_i}), \tag{7}$$

where $\mathbf{x_i}$ denotes the latitude and longitude of the ith tower location.

For time embeddings, we divide the 17-hour time window (from 6:00 AM to 11:00 PM) into 34 half-hour slots, assigning an index from 1 to 34 to each slot. We apply sinusoidal positional encoding to preserve temporal relationships:

$$t_i = \begin{cases} \sin\left(\frac{s_i}{10000^{\frac{2j}{d}}}\right), & \text{if } j \text{ is even} \\ \cos\left(\frac{s_i}{10000^{\frac{2j}{d}}}\right), & \text{if } j \text{ is odd} \end{cases} \tag{8}$$

where s_i is the time slot index (ranging from 1 to 34), j is the embedding dimension index, d is the total embedding dimension. This encoding ensures that nearby time slots have similar embeddings, allowing the model to recognize the temporal structure of the sequence effectively.

Additional context, including age, gender, primary and secondary anchor points, and temporal indicators (e.g., weekday vs. weekend), is incorporated into the model. These background tokens are combined with trajectory tokens and then fed into the Transformer encoder. BERT4Traj predicts tower IDs for the half-hour time slots where no records exist, reconstructing a temporally detailed trajectory.

3.2 GPS Data

GPS data provides higher temporal resolution than CDR but still contains missing records due to device-related issues such as power-saving modes, signal loss, or shutdowns. In our GPS dataset, each individual has an average of 197 days of records, with a mean time interval of 18 minutes between points. To construct continuous trajectories, we apply BERT4Traj to predict the missing locations during these gaps.

The input GPS trajectory consists of exact coordinate points (latitude, longitude) recorded at each timestamp. However, to facilitate prediction and ensure spatial consistency, these exact coordinates are mapped to a 200m × 200m grid. Each grid cell has a unique Grid ID, which serves as a spatial unit for the model. To ensure spatial coherence, we derive location embeddings using Space2Vec, following the same approach as with the CDR data.

For time encoding, we use normalized time encoding to preserve the full timestamp (hour, minute) in a continuous and periodic manner. First, we normalize the time to a fraction of the day:

$$t_{\text{norm}} = \frac{\text{hour} \times 60 + \text{minute}}{1440} \tag{9}$$

where 1440 is the total number of minutes in a day.

Then, we apply sinusoidal encoding to capture periodicity:

$$t_i = \begin{cases} \sin(2\pi t_{\text{norm}}) \\ \cos(2\pi t_{\text{norm}}) \end{cases} \tag{10}$$

This encoding ensures that time is represented in a continuous way, maintaining smooth transitions between consecutive timestamps while preserving cyclic properties.

Similarly, we incorporate background tokens representing demographic attributes, anchor points, and temporal indicators. BERT4Traj then reconstructs continuous trajectories by predicting the most likely location (Grid ID) at any given time.

3.3 Model Evaluation

The effectiveness of BERT4Traj was evaluated against several baseline models, including Markov Chain, RNN, LSTM, and KNN, using both CDR and GPS datasets. Performance is assessed using the following metrics:

- **Accuracy**: The proportion of correctly predicted locations.
- **Top-3 Accuracy**: Whether the correct location is among the top three predictions.
- **Top-5 Accuracy**: Whether the correct location is among the top five predictions.

Since the goal of this study is to reconstruct mobility trajectories for known users, we adopt a trajectory-level data split, where each user's daily trajectories are partitioned into 80% for training, 10% for validation, and 10% for testing. For both the CDR and GPS datasets, we apply a random masking strategy during training, where 20% of the location tokens in each trajectory are masked. The model is trained to predict these masked locations using the remaining context and background information. During testing, the same masking ratio (20%) is applied to the trajectories in the test set. The trained BERT4Traj model is then used to predict the masked locations in these test sequences. Model performance is evaluated by comparing the predicted locations against the true masked labels using metrics such as accuracy, top-3 accuracy, and top-5 accuracy

Table 1 summarizes the performance comparison.The results clearly demonstrate that BERT4Traj outperforms all baseline models across both datasets. In the CDR dataset, BERT4Traj achieves an accuracy of 87.1%, significantly surpassing LSTM (74.5%) and RNN (70.6%), highlighting its superior ability to handle sparse mobility data compared to recurrent models. Similarly, in the GPS dataset, BERT4Traj attains 71.4% accuracy, outperforming LSTM (62.1%) and RNN (60.3%). The lower accuracy observed in the GPS dataset compared to CDR is due to the fundamental difference in prediction tasks – CDR reconstruction predicts locations within predefined half-hour time slots, whereas GPS trajectory reconstruction requires continuous predictions across time, making the task inherently more challenging. Despite this, BERT4Traj still achieves notable improvements over baseline models, demonstrating its robustness in handling missing data.

Among the baseline models, Markov Chain and KNN exhibit the lowest accuracy, particularly in the GPS dataset, where their accuracy remains below 55%. This indicates that these simpler models struggle to capture sequential dependencies and complex spatial-temporal relationships, reinforcing the advantage of deep learning approaches in trajectory reconstruction.

Table 1 Comparison with baselines in Accuracy, Top-3 Accuracy, and Top-5 Accuracy.

Model	CDR Accuracy (%)	CDR Top-3 (%)	CDR Top-5 (%)	GPS Accuracy (%)	GPS Top-3 (%)	GPS Top-5 (%)
Markov Chain	62.1	65.2	67.8	52.7	53.9	55.3
RNN	70.6	73.4	74.9	60.3	61.2	62.8
LSTM	74.5	77.6	80.1	62.1	64.5	66.4
KNN	67.2	70.4	72.3	54.2	55.4	57.1
BERT4Traj	**87.1**	**89.8**	**91.2**	**71.4**	**73.4**	**74.8**

To examine the contribution of different contextual background features, we conducted an ablation study where demographic information, anchor points, and date information were individually removed from BERT4Traj. The results of this analysis are shown in Table 2 below.

Table 2 Ablation Study: Effect of Removing Contextual Features on Model Accuracy.

Feature Removed	CDR Accuracy (%)	GPS Accuracy (%)
No Demographics	82.7	69.5
No Anchor Points	84.3	71.2
No Date Information	81.5	68.1
Full Model (BERT4Traj)	**87.1**	**71.4**

The findings show that removing any contextual feature leads to a decline in model performance. The most significant drop occurs when removing date information, reducing accuracy to 81.5% in the CDR dataset and 68.1% in the GPS dataset. This suggests that temporal context plays a crucial role in predicting missing locations. The removal of demographic data also results in a notable accuracy drop, indicating that user characteristics contribute valuable information for mobility prediction. Similarly, excluding anchor points reduces accuracy, highlighting their importance in modeling an individual's movement behavior.

Overall, these results demonstrate that incorporating spatial, temporal, and demographic background information enhances the accuracy of BERT4Traj in reconstructing mobility trajectories.

4 Conclusion

This study introduced BERT4Traj, a Transformer-based model for reconstructing complete mobility trajectories from sparse location data. The model effectively captures spatial and temporal dynamic relationships, enabling more accurate trajectory reconstruction. Additionally, BERT4Traj enhances location prediction accuracy by incorporating multi-faceted contextual embeddings, including demographic, anchor point, and temporal features, enriching the representation of human mobility patterns. Moreover, BERT4Traj provides a scalable and adaptable framework for mobility datasets, making it applicable to public health, urban planning, and transportation analytics.

Despite its advantages, BERT4Traj has limitations. Its generalizability across different regions requires further validation, as mobility behaviors vary across geographic and socioeconomic contexts. Privacy concerns also emerge when reconstructing detailed trajectories, necessitating robust safeguards. Future research should explore multi-source mobility data integration, efficiency optimization, and privacy-preserving techniques. Addressing these challenges will enhance BERT4Traj's reliability and applicability in human mobility research and decision-making.

References

1 Ian Barnett and Jukka-Pekka Onnela. Inferring mobility measures from gps traces with missing data. *Biostatistics*, 21(2):e98–e112, 2020.

2 Vitaly Belik, Theo Geisel, and Dirk Brockmann. Natural human mobility patterns and spatial spread of infectious diseases. *Physical Review X*, 1(1):011001, 2011.

3 Cynthia Chen, Ling Bian, and Jingtao Ma. From traces to trajectories: How well can we guess activity locations from mobile phone traces? *Transportation Research Part C: Emerging Technologies*, 46:326–337, 2014.

4 Guangshuo Chen, Aline Carneiro Viana, Marco Fiore, and Carlos Sarraute. Complete trajectory reconstruction from sparse mobile phone data. *EPJ Data Science*, 8(1):1–24, 2019. doi:10.1140/EPJDS/S13688-019-0206-8.

5 Alessandro Crivellari, Bernd Resch, and Yuhui Shi. Tracebert—a feasibility study on reconstructing spatial–temporal gaps from incomplete motion trajectories via bert training process on discrete location sequences. *Sensors*, 22(4):1682, 2022. doi:10.3390/S22041682.

6 Sébastien Gambs, Marc-Olivier Killijian, and Miguel Núñez del Prado Cortez. Next place prediction using mobility markov chains. In *Proceedings of the first workshop on measurement, privacy, and mobility*, pages 1–6, 2012. doi:10.1145/2181196.2181199.

7 Sahar Hoteit, Stefano Secci, Stanislav Sobolevsky, Carlo Ratti, and Guy Pujolle. Estimating human trajectories and hotspots through mobile phone data. *Computer Networks*, 64:296–307, 2014. doi:10.1016/J.COMNET.2014.02.011.

8 Wei Huang, Songnian Li, Xintao Liu, and Yifang Ban. Predicting human mobility with activity changes. *International Journal of Geographical Information Science*, 29(9):1569–1587, 2015. doi:10.1080/13658816.2015.1033421.

9 Mingxiao Li, Song Gao, Feng Lu, and Hengcai Zhang. Reconstruction of human movement trajectories from large-scale low-frequency mobile phone data. *Computers, Environment and Urban Systems*, 77:101346, 2019. `doi:10.1016/J.COMPENVURBSYS.2019.101346`.

10 Zhang Liu, Ting Ma, Yunyan Du, Tao Pei, Jiawei Yi, and Hui Peng. Mapping hourly dynamics of urban population using trajectories reconstructed from mobile phone records. *Transactions in GIS*, 22(2):494–513, 2018. `doi:10.1111/TGIS.12323`.

11 Gengchen Mai, Krzysztof Janowicz, Bo Yan, Rui Zhu, Ling Cai, and Ni Lao. Multi-scale representation learning for spatial feature distributions using grid cells. In *International Conference on Learning Representations*, 2020.

12 Sandro Meloni, Nicola Perra, Alex Arenas, Sergio Gómez, Yamir Moreno, and Alessandro Vespignani. Modeling human mobility responses to the large-scale spreading of infectious diseases. *Scientific reports*, 1(1):62, 2011.

13 Junjun Si, Jin Yang, Yang Xiang, Hanqiu Wang, Li Li, Rongqing Zhang, Bo Tu, and Xiangqun Chen. Trajbert: Bert-based trajectory recovery with spatial-temporal refinement for implicit sparse trajectories. *IEEE Transactions on Mobile Computing*, 2023.

14 Jingyuan Wang, Ning Wu, Xinxi Lu, Wayne Xin Zhao, and Kai Feng. Deep trajectory recovery with fine-grained calibration using kalman filter. *IEEE Transactions on Knowledge and Data Engineering*, 33(3):921–934, 2019.

15 Hao Yang, X Angela Yao, Christopher C Whalen, and Noah Kiwanuka. Exploring human mobility: a time-informed approach to pattern mining and sequence similarity. *International Journal of Geographical Information Science*, pages 1–25, 2024.

16 Haofei Yu, Armistead Russell, James Mulholland, and Zhijiong Huang. Using cell phone location to assess misclassification errors in air pollution exposure estimation. *Environmental pollution*, 233:261–266, 2018.