Human Mobility Challenge: Are Transformers Effective for Human Mobility Prediction?

Ruochen Kong
ruochen.kong@emory.edu
Emory University
Atlanta, GA, USA

Hossein Amiri
hossein.amiri@emory.edu
Emory University
Atlanta, GA, USA

Yueyang Liu
yueyang.liu@emory.edu
Emory University
Atlanta, GA, USA

Lance Kennedy
lance.kennedy@emory.edu
Emory University
Atlanta, GA, USA

Misha Gupta
misha.gupta@emory.edu
Emory University
Atlanta, GA, USA

Joon-Seok Kim
joonseok.kim@emory.edu
Emory University
Atlanta, GA, USA

Andreas Züfle
azufle@emory.edu
Emory University
Atlanta, GA, USA

Abstract

Transformer-based models are popular for time series forecasting and spatiotemporal prediction due to their ability to infer semantic correlations in long sequences. However, for human mobility prediction, temporal correlations, such as location patterns at the same time on previous days or weeks, are essential. While positional encodings help retain order, the self-attention mechanism causes a loss of temporal detail. To validate this claim, we used a simple approach in the 2nd ACM SIGSPATIAL Human Mobility Prediction Challenge, predicting locations based on past patterns weighted by reliability scores for missing data. Our simple approach was among the top 10 competitors and significantly outperformed the Transformer-based model that won the 2023 challenge.

CCS Concepts

• **Information systems** → **Geographic information systems**; **Location based services**.

Keywords

Human Mobility, Patterns of Life, Historical Huristic

ACM Reference Format:

Ruochen Kong, Hossein Amiri, Yueyang Liu, Lance Kennedy, Misha Gupta, Joon-Seok Kim, and Andreas Züfle. 2024. Human Mobility Challenge: Are Transformers Effective for Human Mobility Prediction?. In *2nd ACM SIGSPATIAL International Workshop on the Human Mobility Prediction Challenge (HuMob'24), October 29-November 1, 2024, Atlanta, GA, USA.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3681771.3700130

1 Introduction

Human mobility is a critical component in the emerging field of spatial intelligence and the science of human mobility has a plethora of applications from infectious disease contact tracing to elder care and crime detection [9, 10]. Recent works leverage human mobility data to model and trace infectious diseases [5], analyzing human behavior and detecting anomalies [1, 7, 17, 18], traffic monitoring and prediction [2], and enhancing urban planning [3].

HuMob'24, October 29-November 1, 2024, Atlanta, GA, USA

ACM ISBN 979-8-4007-1150-3/24/10
https://doi.org/10.1145/3681771.3700130

Artificial intelligence techniques have increasingly been applied to recognize patterns in human mobility data. Traditional methods, such as machine learning [16] and deep learning [8], have demonstrated significant potential in analyzing complex mobility patterns. However, these approaches often struggle with the inherent challenges presented by such data, including temporal dependencies, spatial irregularities, and high variability. To address these limitations, generative models have recently emerged as promising alternatives in the field, potentially filling the gaps left by more conventional approaches [6]. By leveraging probabilistic frameworks, generative models aim to improve predictive accuracy and capture nuanced behaviors in human mobility data, advancing the scope of analysis and enabling more sophisticated predictions about movement patterns.

The introduction of Transformers [13] has marked a notable advancement in optimizing deep learning for natural language processing (NLP) tasks, and they have been successfully applied beyond translation, including in areas like medical prediction [4]. Transformers have also been used for human mobility prediction. For example, the winner of the Human Mobility Prediction Challenge 2023 [12] was based on the BERT transformer model and the runner-up [11] was based on the GPT transformer model. Despite their success in NLP, Transformers face challenges when applied to time-series data [15]. Specifically, it has been shown that "the nature of the permutation-invariant self-attention mechanism inevitably results in temporal information loss" [15]. Human Mobility data is a special case of time-series, a time-series of observed locations of individual humans. For short-term predictions of human mobility, such as the prediction of locations over the next minutes or hours, Transformers have shown very good results [11, 12]. But for the prediction of locations of the next days or even weeks, we hypothesize that Transformer-based models may struggle.

To illustrate our intuition, think of the problem of predicting the next words in a text document. The transformer model excels at this problem. However, predicting the next 1000 words is very challenging: Each predicted word has a chance of being incorrect and this chance of errors accumulates until the 1000th word will have a very low chance of being predicted correctly. The same applies for human mobility data: The transformer may do well predicting the next location and even the location afterward. But predicting the chain of next locations for 15 days is likely to incur errors upon which subsequent predictions will be based on.

Table 1: Trajectory data representation

Type	Temporal feature	Interval
Constant time interval (CTI)	sample point	constant
Variable time interval (VTI)	sample point	variable
Event-based interval (EBI)	event	variable
Constant time period (CTP)	period	constant

But can we do better for human mobility prediction? For next-word text prediction, there is no seasonality or periodicity as words don't simply repeat every 100 words. But for human mobility prediction, there is! We know that human mobility follows strong periodic patterns: Rush hours occur every non-weekend non-holidays. Individual humans are likely to be located at their home at 4am every single day, or at work/school during the day on non-weekend non-holidays. Individual humans may also visit the same grocery stores and recreational places on weekends and holidays. If you were to ask yourself: "Where will I be in exactly 14 days?". When answering this question, you're likely not trying to predict your next locations today, then prediction your first location tomorrow followed by all further locations tomorrow, followed by your first location in two days, until you reach 14 days. Instead, you would think of your periodic patterns: Where am I usually on this day of the week during the current time? Where am I right now? Where was I seven and fourteen days ago? You may want to base your prediction on your answer to the former questions.

With these insights, our approach is motivated to predict the location of an individual in 392 hours (the same hour in 14 days), rather than inductively predicting the location at hour i given the locations until hour $i-1$ (and accumulating prediction errors), we look at where is the individual usually at this day of the week. Thus, we exploit the periodicity of human mobility that a transformer is not able to leverage.

2 Methodology

In this section, we describe the methodology, including the dataset representation of trajectories, the preprocessing techniques used to prepare the data for further analysis, the creation of the confidence matrix, and, finally, the prediction model.

2.1 Trajectory Data Representation

There are several representations of trajectory data, as illustrated in Table 1. One advantage of constant time interval (CTI) data is that it enables a model to learn the distribution of all locations at synchronized timestamps. In contrast, variable time interval (VTI) data is often used to compress important segments, facilitating the recovery of data from its compressed form. Event-based interval (EBI) representation is particularly effective for compressing data and identifying semantic sequence patterns, which is advantageous for sequence pattern learning. In this study, we utilize constant time period (CTP) representation to train our model, assuming that the most frequent locations during the period are used. Compared to CTI, CTP requires fewer resources for data representation and allows the model to learn the distribution of all locations within a synchronized time frame. If locations are uncertain, CTP can provide better estimates by averaging the most frequent locations over the defined period, thereby smoothing out noise and capturing underlying patterns more effectively. This averaging effect enhances the model's robustness, especially in scenarios where data might be sparse or exhibit high variability. Additionally, CTP representation can facilitate better generalization by focusing on significant location trends rather than being influenced by outlier data points.

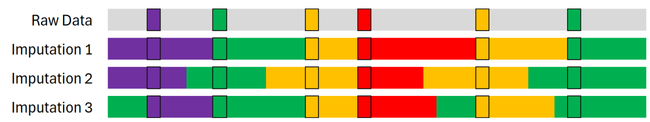

Figure 1: Ideas of data preprocessing. Each color represents a unique location of an individual. Gray is the missing data.

In our analysis of the YJMob100K dataset [14], we identified that the trajectory data can be categorized into VTI and EBI based on different hypotheses. **Hypothesis 1** (VTI) posits that there is a genuine absence of location observations for the individual at a given time, indicating that the data is missing. **Hypothesis 2** (EBI) suggests that the user remained in the same location until the next observation, meaning that the data is not missing but rather omitted for the purpose of data compression.

We also observed that individuals are rarely recorded in the same location, which poses a challenge for the EBI hypothesis. However, given the spatial resolution of $500m \times 500m$ and the temporal resolution of $30min$, it is expected that individuals would primarily remain within the same spatial cell, except during travel. This leads us to strongly favor the second hypothesis as the more plausible explanation. Additionally, we found that many individuals had no recorded data in the initial days, with some users lacking any data for the first 40 out of 75 days. We attribute this phenomenon to the varying starting times of data collection for different individuals.

2.2 Preprocessing Approaches

Based on these observations and assumptions, we employed three distinct data imputation approaches to transform the data into CTP format. These methods are illustrated in Figure 1 and described in detail below:

Imputation 1. We treated all missing locations using Hypothesis 2. For each missing data points, we simply assume that the individual has not moved since their previous observation. Thus, each missing location is replaced with the location of the individuals most recent observation. For missing locations before the individual's first observation, we use the location of the first observation. This approach is shown in the second row of Figure 1 where the imputed locations change when a new location is observed.

Imputation 2. If Hypothesis 2 holds true, then the data points correspond to change-of-location events. In the withheld test data, we would also need to predict these change-of-location events rather than predicting current locations. This makes a difference. For example, assume a very boring individual who goes to work between 8-9am every day, comes back home from work between 5-7pm every day, and never visits any other place. Now, if you were asked to predict the location of this individual on a day at 4:30pm, what location would you predict? If you were tasked to predict their current location, you'd likely respond "at work", because this user is always at work at that time. But if your task is to predict the location of a change-of-location event at 4:30pm, then the answer should be "at home": That's because it is much more likely for the user to arrive at home at 4:30pm (30 minutes earlier than usually) than to arrive at work at 4:30pm (7.5 hours later than usually). Using this intuition of predicting change-of-location events, we may predict that any time before around 1pm is likely a (possibly very late) arrive-at-work event, and any time after 1pm is likely a (possibly very early) arrive-at-home event. Imputation 2 formalizes this intuition. Using Imputation 2, we calculate the length of each missing time block. If such length is larger than a (small) temporal threshold (such as 1 hour), we fill the first half of the missing data with the previous location and the remainder with the next location. Otherwise, if the gap is smaller than the threshold, we still fill the entire block with the previous location as done for Imputation 1. An example of Imputation 2 is shown in the third line of Figure 1.

Imputation 3. Similar to Imputation 2, we assume the locations may have changed during an extended period of missing data. But now, we assume that an user would stay at the previously observed location for no more than a threshold of time. We fill the gap with the previous location until the threshold. After that, we concede that we have no idea where the individual may be, and we simply fill the remainder of the gap with the individual's most common location. This approach is shown in the last line of Figure 1 assuming that the "green" location is the individual's most common location.

2.3 Confidence Matrices

From the preprocessing step, we obtained a dataset with each user having location information at each half-hour time stamp. However, we can no longer discriminate between observed data and imputed data. Intuitively, the further the imputed time is from the nearest (previous or next) observation, the lower the confidence that this data is imputed correctly. For example, for a user missing three days of data, our confidence in the imputation locations on the second day of missing data is very low. But if the imputed value is immediately after an observed location, our confidence in this computation is higher. Thus, we create a confidence matrix to avoid confusing our model by using imputed data. For an individual i and time t, let t^i_{prev} and t^i_{next} denote the time of the previous and next observation of i, respectively. We let Δt^i denote the smaller absolute difference between t and $\{t^i_{prev}, t^i_{next}\}$, that is:

$$\Delta t^i := \min_{t' \in \{t^i_{prev}, t^i_{next}\}} |t - t'|$$

Then we define a confidence matrix that computes a confidence value for each individual i and each time interval t as:

$$\mathrm{C}[i,t] = e^{-\frac{(\Delta t^i)^P}{D}}$$

where D is a normalization parameter that we simply call Denominator and P is a parameter simply called Power that controls the exponential decay of confidence as the temporal distance to an observation increases. This equation maps each individual-time pair to a confidence value in $[0,1]$ having a confidence of $e^0 = 1$ for locations that are not imputed but observed and thus have a distance of Δt^i of zero. For our experiments, we test various combinations of denominators and powers as described in Table 2.

2.4 Prediction Model

For the SIGSPATIAL Human Mobility Prediction Challenge 2024 prediction task, locations are masked for a validation set V of indviduals (the number of which depends on the City to be prediction) for days 61-75 in the dataset which corresponds to time intervals $[2880, 3600]$ as each day has 48 time intervals of 30 minutes each. We denote the set of pair (i,t) where $i \in V$ and $t \in [2880, 3600]$ as the prediction window.

First, we use the imputation method described in Section 2.2 to impute all the missing values outside of the prediction value, that is for all agents $i \notin V$ (at any time), and for all times $t \notin [2880, 3600]$ (for all agents). The result of this imputation is a matrix that is full (no missing values) outside of the prediction window. For any time $t \in [0, 2879]$ and any individual $i \in V$ we let $Impute[i,t]$ denote the imputed location for i at t. Next, we create the confidence matrix Confidence$[i,t]$ as described in Section 2.3.

To predict the location of user $i \in V$ during time $t \in [2880, 3600]$, we iteratively look at the location of i at multiples of seven days (336 30-minute intervals) ago. Thus, we look up location $Impute[i, t-k \cdot 336]$, $1 \le k \le 10$. For each location found this way, we add a weight to that location corresponding to the confidence confidence$[i, t - k \cdot 336]$, $1 \le k \le 10$. We then chose the location having the largest aggregate weight (choosing ties arbitrarily).

As an example, assume we want to predict the location of individual i during time interval $t = 3500$. We first check Impute$[i, t - 1 \cdot 336]$ = Impute$[i, 3164]$. Since time interval 3164 is also part of the prediction window, there is no imputed value, thus Impute$[i, 3164]$ = None. We continue one week further in the past by looking at Impute$[i, t - 2 \cdot 336]$ = Impute$[i, 2828]$. Since time 2828 is not part of the prediction window, we will find an imputed value. Let location $[12, 18]$ be the imputed location. Also assume that this location was directly observed (not imputed), such that confidence$[i, 2828] = 1$. We increment the weight of location $[12, 18]$ by 1.0. Next, we look at $[i, t - 3 \cdot 336]$ = Impute$[i, 2492]$. Assume that this look-up yields a different location $[10, 25]$ having a low confidence confidence$[i, 2492] = 0.1$. We increment the weight of location $[10, 25]$ by 0.1. We repeat this process using the previous weeks Impute$[i, 2156]$, Impute$[i, 1820]$, Impute$[i, 1484]$, Impute$[i, 1148]$, Impute$[i, 812]$, Impute$[i, 476]$, and Impute$[i, 140]$. Then, we use the location having the highest sum of weights as our predicted location. In this example, assume that location $[10, 25]$ has been observed (not imputed) once for an addition weight of 1.0 for a total weight of 1.1, and assume that no other location has a higher total weight, so we would predict location $[10, 25]$ in this example. A formal algorithm of our prediction approach is found in Algorithm 1.

Algorithm 1 Location prediction of user i at day d and time t

Input: Impute Matrix I$[i,t]$, Confidence Matrix C$[i,t]$, User ID i, and Time t
Output: location of the user i at time t

```
procedure PreviousWeeks
    score ← map()            ▷ key: location, value: sum confidence
    for k ← 1 to 10 do       ▷ At most 10 previous weeks
        t_prev ← t − k · 7 · 48
        score[I[i, t_prev]] += C[i, t_prev]
    end for
    location ← argmax_{score.key}(score)
end procedure
```

3 Experimental Results

Conducted Experiments. Our proposed approach has three hyperparameters: The normalization parameter D, the exponential imputation decay P, and the choice among the three proposed imputation methods. To tune these hyperparameters, we used the first 45 days of training and the next 15 days for testing where

Table 2: Accuracy of models with all combinations of preprocessing approaches and confidence matrices

Denominator		24			18	28
Power		8	3	0.7	0.7	0.7
Imputation	1	0.2341	0.2347	0.2348	0.2343	0.2350
Imputation	2	0.2322	0.2330	0.2330	0.2328	0.2331
Imputation	3	0.2343	0.2353	0.2364	0.2357	0.2368

Table 3: Evaluation metrics of the experimental results

City	Model	ACC	GeoBleu	DTW	Time
A	Transformer	0.2186	0.2704	30.45	6 (h)
	Proposed	0.2772	0.3180	34.86	45 (m)
B	kNN	0.1840	0.2085	28.52	4 (m)
	Proposed	0.2529	0.2841	32.02	35 (m)
C	kNN	0.1729	0.1960	22.47	3 (m)
	Proposed	0.2578	0.2750	23.65	33 (m)
D	kNN	0.1417	0.2101	58.76	2 (m)
	Proposed	0.2368	0.3036	52.56	30 (m)

the ground truth is available. Due to the runtimes for the GeoBleu and DTW evaluation metrics, we only use the exact-match Accuracy (ACC) using the smaller City D for tuning. Table 2 shows our hyperparameter tuning results. Based on the results, we chose Imputation 3 with a confidence matrix $e^{(-(dt/28)^{0.7})}$ for our challenge submission.

Comparison to Baselines. We evaluated our approach with 60 days of training and 15 days of validation on City A and with 45 days of training and 15 days of validation on Cities B, C, and D. We also implemented two baseline methods: Transformer (last year's challenge winner, with a similar setup but different datasets) and k-Nearest-Neighbor Imputation (the inspiration for our proposed method). The evaluation results are shown in Table 3. The kNN approach, specifically the 1-NN variant, identifies the most similar training time slot for each testing time slot and uses the corresponding locations of individuals from the selected time slot as the predicted values. We currently only have the result for the Transformer on City A and for kNN on City B, C, and D. Both approaches require specific resources for training, namely a high-end GPU (i.e. NVIDIA H100) for the Transformer, or a large RAM for kNN. The baseline models for the remaining cities are still training at the time of this submission.

Our proposed method outperforms both the Transformer-based approach and the kNN approach using both ACC and GeoBleu as evaluation metrics for all four datasets. However, our result worse using DTW. One reason for this phenomenon is that we chose the parameter by only optimizing the accuracy and hence GeoBleu. Besides, GeoBleu and DTW have competing optimization goals. GeoBleu gives only a few credits to the predicted ones close to the truth, but with DTW, it would be better to predict every location close to the truth.

4 Conclusion

We propose an embarrassingly simple approach for long-term prediction of future human mobility. Our approach starts by imputing missing data and then simply uses the individuals' location at the same time as in previous weeks. Our approach is highly efficient, requiring only milliseconds to predict the location of individuals even for large datasets. In contrast, existing Transformer-based approaches require high-end computing environments and incur excessive compute or memory cost. Despite the simplicity and the efficiency of our proposed algorithm, our results are competitive with the state-of-the-art. Using exact-match accuracy and the GeoBleu metric, our approach outperforms the state-of-the-art while using DTW, our approach is not far behind. Implementation details of our proposed approach can be found in our GitHub repository at https://github.com/RuochenKong/HuMob_Cucumber.git.

5 Acknowledgements

Supported by the Intelligence Advanced Research Projects Activity (IARPA) via Department of Interior/ Interior Business Center (DOI/IBC) contract number 140D0423C0025. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DOI/IBC, or the U.S. Government.

References

[1] Hossein Amiri, Ruochen Kong, and Andreas Zufle. 2024. Urban Anomalies: A Simulated Human Mobility Dataset with Injected Anomalies. arXiv:2410.01844
[2] Wenqiang Chen, Tao Wang, et al. 2022. Lane-based Distance-Velocity model for evaluating pedestrian-vehicle interaction at non-signalized locations. *Accident Analysis & Prevention* 176 (2022), 106810.
[3] Sibren Isaacman, Richard Becker, et al. 2012. Human mobility modeling at metropolitan scales. In *MobiSys'12*. 239–252.
[4] Seibi Kobara, Alireza Rafiei, et al. 2024. Social Media as a Sensor: Analyzing Twitter Data for Breast Cancer Medication Effects Using Natural Language Processing. In *International Conference on Artificial Intelligence in Medicine*. Springer, 345–354.
[5] Will Kohn, Hossein Amiri, and Andreas Züfle. 2023. EPIPOL: An Epidemiological Patterns of Life Simulation (Demonstration Paper). In *SIGSPATIAL SpatialEpi'23 Workshop*. 13–16.
[6] Xishun Liao, Brian Yueshuai He, et al. 2024. Deep Activity Model: A Generative Approach for Human Mobility Pattern Synthesis. *arXiv preprint arXiv:2405.17468* (2024).
[7] Yueyang Liu, Lance Kennedy, Hossein Amiri, and Andreas Züfle. 2024. Neural Collaborative Filtering to Detect Anomalies in Human Semantic Trajectories. *arXiv preprint arXiv:2409.18427* (2024).
[8] Massimiliano Luca, Gianni Barlacchi, Bruno Lepri, and Luca Pappalardo. 2021. A survey on deep learning for human mobility. *ACM Computing Surveys (CSUR)* 55, 1 (2021), 1–44.
[9] Mohamed Mokbel, Mahmoud Sakr, Li Xiong, et al. 2024. Mobility Data Science: Perspectives and Challenges. *ACM Transactions on Spatial Algorithms and Systems* 10 (2024). Issue 2.
[10] Mohamed Mokbel, Mahmoud Sakr, Li Xiong, Andreas Züfle, et al. 2022. Mobility data science (dagstuhl seminar 22021). In *Dagstuhl reports*, Vol. 12. Schloss Dagstuhl-Leibniz-Zentrum für Informatik.
[11] Aivin V Solatorio. 2023. GeoFormer: Predicting Human Mobility using Generative Pre-trained Transformer (GPT). In *HuMob-Challenge '23*. 11–15.
[12] 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 *HuMob-Challenge '23*. 1–6.
[13] A Vaswani. 2017. Attention is all you need. *Advances in Neural Information Processing Systems* (2017).
[14] Takahiro Yabe, Kota Tsubouchi, et al. 2024. YJMob100K: City-scale and longitudinal dataset of anonymized human mobility trajectories. *Scientific Data* 11, 1 (2024), 397.
[15] Ailing Zeng, Muxi Chen, Lei Zhang, and Qiang Xu. 2023. Are transformers effective for time series forecasting?. In *IAAI'24*, Vol. 37. 11121–11128.
[16] Ying Zhang, Zhiwen Yu, et al. 2024. Predictability in human mobility: From individual to collective (vision paper). *ACM TSAS* 10, 2 (2024), 1–17.
[17] Zheng Zhang, Hossein Amiri, et al. 2023. Large Language Models for Spatial Trajectory Patterns Mining. *arXiv e-prints* (2023), arXiv–2310.
[18] Zheng Zhang, Hossein Amiri, et al. 2024. Transferable Unsupervised Outlier Detection Framework for Human Semantic Trajectories. arXiv:2410.00054