Using the Temporal-Trajectory-based K Nearest Neighbor Algorithm to Predict Human Mobility Patterns

An-Syu Li
Department of Computer Science and Engineering, National Chung Hsing University
Taiwan, R.O.C.
yessir0621@gmail.com

Ling-Huan Meng
Department of Computer Science and Engineering, National Chung Hsing University
Taiwan, R.O.C.
lhmeng93@gmail.com

Yu-Ling Zhong
Department of Computer Science and Engineering, National Chung Hsing University
Taiwan, R.O.C.
qazxcvb1425@gmail.com

Yi-Chung Chen
Department of Computer Science and Engineering, National Chung Hsing University
Taiwan, R.O.C.
chenyich@nchu.edu.tw/mitsukoshi901@gmail.com

Tomoya Kawakami
Graduate School of Engineering, University of Fukui
Japan
tomoya-k@u-fukui.ac.jp

Abstract

Researchers are increasingly applying AI to predict the daily routines of individuals and their corresponding trajectories (referred to as mobility prediction), due to its considerable potential for commercial activities and public administration. However, most previous studies focused exclusively on model theory or design without considering the specific characteristics of human flow trajectories. In this study, we analyzed historical data to identify three factors that play key roles in human mobility. Based on these factors, we developed a temporal-trajectory-based K-nearest neighbor algorithm to predict human flow trajectories. Experimental simulations demonstrated the effectiveness of the proposed scheme when applied to the HuMob Challenge 2024 dataset.

CCS Concepts

• **Information systems → Location based services**.

Keywords

KNN, Human mobility, Machine learning

ACM Reference Format:
An-Syu Li, Ling-Huan Meng, Yu-Ling Zhong, Yi-Chung Chen, and Tomoya Kawakami. 2024. Using the Temporal-Trajectory-based K Nearest Neighbor Algorithm to Predict Human Mobility Patterns. 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/10.1145/3681771.3699913

1 Introduction

The ability to predict individual movements at specific points in the future can have significant implications for both commercial ventures and public administration. Figure 1 provides an illustrative example of a mobile app that has compiled a list of locations in which the app was used over the previous week. This location data could serve as input for a human mobility prediction model to determine the route User A regularly takes on workdays. Based on this predicted trajectory, restaurants and coffee shops along the route could be targeted for app-based advertising, under the assumption that User A is more likely to engage with businesses nearby than with those farther away. Additionally, advertising fees could be adjusted based on the effectiveness of the ads.

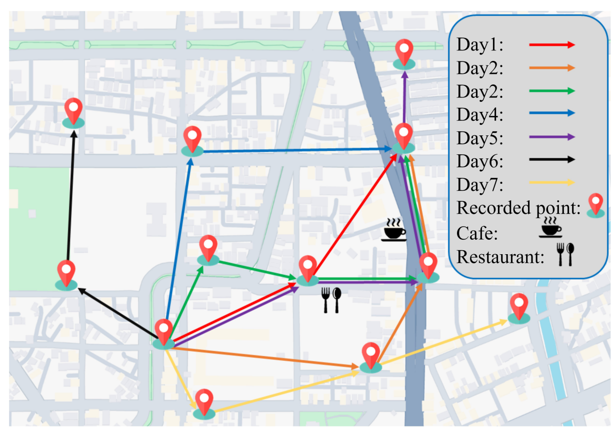

Figure 1: Map marked with user mobility patterns collected over a period of 1 week

Researchers are increasingly applying AI to predict individuals' daily routines and their corresponding trajectories (referred to as mobility prediction). Wang and Deng [10] designed a deep neural network model using spatiotemporal data to model and predict human movement behavior. By treating movement trajectories throughout the day as a sentence and each location as a word, Terashima et al. [9] and Kobayashi et al. [5] were able to use language models to predict an individual's next location. Du and Ye [2] summarized historical human movement records in the form

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

of a graph to derive mobility characteristics for trajectory prediction. He et al. [3] used decision trees to characterize movement trajectories, and then trained a model to predict the next movement location based on these characteristics. Solatorio [7] and Kim et al. [4] adopted a similar approach but implemented a neural network model based on a transformer. The effectiveness of these methods is evident from their high rankings in the HuMob Challenge 2023. However, these works focused primarily on model theory, largely overlooking the importance of model fine-tuning based on the underlying characteristics of human flow patterns. This study is based on the hypothesis that prediction performance can be improved by considering the key factors that influence human mobility.

Generally, crowd flows exhibit three key characteristics: (1) time fixedness, (2) contingency, and (3) uniqueness [8]. First, time fixedness refers to the tendency of human behavior to be influenced by time, resulting in regular, fixed patterns. Figure 2 provides an illustrative example from the HuMob Challenge 2024 dataset, showing location statistics for an app user captured at random times over a 75-day period. The grid in the figure represents 7 days by 24 hours, totaling 168 hours. Darker colors indicate a higher likelihood that the app user remains in a given location during that time, and lighter colors indicate a lower likelihood. For example, this app user tended to remain stationary between 11 p.m. and 3 a.m., and again between 7 a.m. and 8 a.m. Flow contingency refers to the fact that people visit certain places only on rare occasions (e.g., high-end restaurants, historical sites, or specialty shops). Flow uniqueness refers to the fact that the combination of places visited is unique to each individual. For example, students attending the same school may live in different areas and visit different restaurants and entertainment venues.

In the following, we use the three flow characteristics to illustrate how existing human mobility prediction methods could be improved. Time fixedness often causes imbalances in the data distribution, which can hinder AI model training and reduce prediction accuracy. Contingencies often appear as noise in the dataset, further hindering training and leading to erroneous prediction results. The uniqueness of human flow patterns makes it difficult to predict the trajectory of one person based on the trajectories of others. In this study, we sought to fine-tune AI models in accordance with these three characteristics to enhance prediction performance.

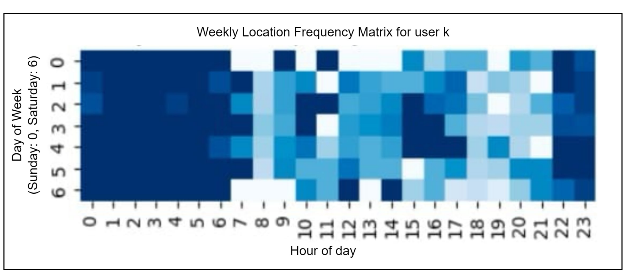

Figure 2: The likelihood that the user remained in a fixed location at timepoints throughout the week (168 hours). In the gradient on the right, darker colors indicate higher probability. The data was collected at random times over a period of 75 days. (HuMob Challenge 2024 dataset)

In this study, we developed a temporal-trajectory-based K nearest neighbors (TT-KNN) model for human mobility prediction. The proposed methodology is based entirely on the three characteristics of human mobility: time fixedness, contingency, and uniqueness. The efficacy of the proposed scheme was assessed using real-world human mobility data provided by the organizers of the HuMob Challenge 2024 [11].

2 Methods

The proposed algorithm comprises offline processing and online processing.

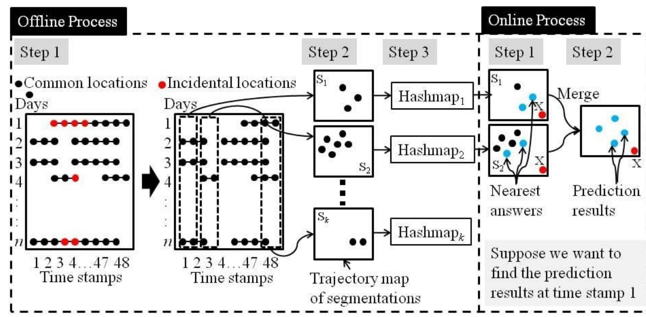

Figure 3: Conceptual illustration showing the concept of the proposed algorithm

2.1 Offline Process

The left half of Figure 3 shows the concept underlying offline processing, which includes three steps: (1) Deletion of incidental locations; (2) Dividing trajectory points based on time segmentation; and (3) Establishing a TT-KNN data structure.

(1) **Deletion of incidental locations.** Processing the trajectory data of user A over the past n days involves performing a statistical analysis of the trajectory locations during this period. Locations with an appearance rate below the user-defined threshold are deleted. The removal of these infrequent locations (first step in Fig. 3) allows us to clean the data by reducing the noise associated with contingencies in human flow patterns.

(2) **Dividing trajectory points based on time segmentation.** As mentioned, many human activities are performed at specific times and in specific locations. Therefore, we divided the user's trajectories and locations throughout the day into multiple time segments. The location associated with each segment is then analyzed with the corresponding locations in future time segments (see the second step in Fig. 3). The size of each time segment is determined by the user, depending on the frequency of data collection or the habits of individuals in that area.

(3) **Establishing the TT-KNN data structure**. When using an algorithm based on the KNN concept, the user typically needs to calculate the distance between the query point and each data point. In this study, we aimed to eliminate this time-consuming task by adopting a hashmap data structure for storing trajectory locations associated with specific time segments. This approach accelerates the online prediction speed of our TT-KNN algorithm.

2.2 Online Process

As shown in the right half of Fig. 3, the Online process involves two steps: (1) Selection of time segments for prediction; and (2) Utilizing TT-KNN for location prediction.

(1) Selection of time segments for prediction Suppose that our aim is to predict the position of user A at time $T+1$, and the predicted reference is the trajectory over m future time points. The algorithm selects the hashmap of TT-KNN time segments from $T+1$ to $T+m$ for use in the next step.

(2) Utilizing TT-KNN for location prediction Suppose that we know the location of the user at time T, referred to as X. Using X as an input, we ask TT-KNN to find the two nearest neighbors in the time segments from $T+1$ to $T+m$. After completing this process, the $2\times m$ results are combined to determine the two nearest neighbors to X. If the nearest neighbor overlaps with the X position of the two neighbors, we take the second nearest neighbor as the final prediction. Otherwise, the nearest neighbor is used as the final prediction. This ensures that the prediction result is close to the current position and likely belongs to the same behavioral trajectory, thereby avoiding significant prediction inaccuracies. Moreover, the fact that KNN-based algorithms can operate with minimal data makes it possible to perform predictions based on the trajectory of a single user (i.e., without relying on the data of other users), thereby ensuring adherence to the uniqueness of flow.

3 Experiment

This section is divided into four parts, including experiment setup, time fixity verification, performance evaluation, and case study.

3.1 Experimental Setup

The dataset used in this experiment was sourced from the HuMob Challenge 2024, containing human mobility data from four cities in Japan (codenamed ABCD). This study focused on data from 3,000 users in city B over a 60-day period. Data from the first 45 days was used as the database for TT-KNN prediction, while data from the following 15 days was used as the prediction target.

When setting the parameters for the TT-KNN algorithm, we employed a threshold value of 5 for the inclusion of infrequent locations. This means that any location not visited at least once a week was considered an infrequent location. The trajectory was recorded at 30-minute intervals, creating time segments in half-hour units. Given that most mobile activities are typically completed within one hour, we selected a time segment that extends from the desired prediction time to one hour beyond this point as the prediction reference.

The performance of TT-KNN was compared with the Generative pre-trained transformer [1] , an award winner in the HuMob Challenge 2023. As recommended by HuMob Challenge 2024, we used Dynamic TimeWarping (DTW) and GEOBLEU[6] as evaluation metrics. Note that GEOBLEU measures similarity (where higher scores indicate better performance), while DTW measures the distance between two time series (where lower scores are better).

3.2 Verifying the time fixity of data

The design of TT-KNN in this paper is based on the time-fixedness of human flow patterns. Thus, we first need to justify the use of TT-KNN by ensuring that the target dataset exhibits this characteristic. Figure 4 presents our statistical results based on the 3,000 users in the target dataset. The X-axis represents the 168 hours in a week, the Y-axis shows the number of users, and the various colored lines indicate different fixed proportions. Consider the only user in which the fixed ratio is 90% (the purple line) and $X = 105$ hours, which means that this user will spend 105 hours in a week remains at some specific location (e.g., in home or in work place) on at least 90% of the previous 60 days. If the fixed rate were set to 50%, then nearly 90% of users would spend more than 75 hours in a specific location each week. If the fixed rate were set to 80%, then many users would still spend at least 50 hours at a specific location every week. This can provide the best verification of the time fixity of the data.

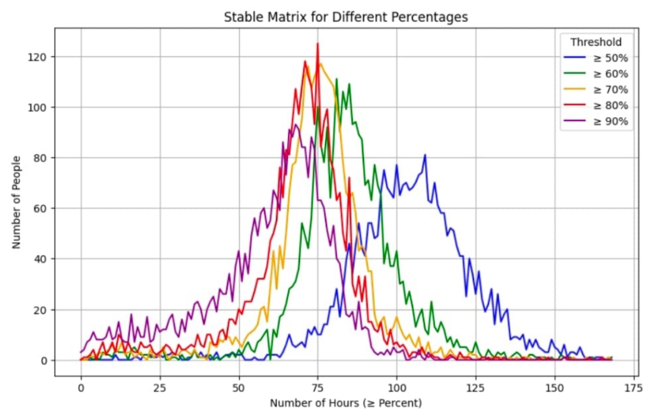

Figure 4: Statistical analysis of human mobility as a function of fixed ratio

3.3 Performance assessment

As shown in Table 1, the proposed TT-KNN model outperformed the GPT in terms of both GEOBLEU and DTW, thereby verifying the effectiveness of the proposed scheme.

3.4 Case Study

In the following, we present a simple case study demonstrating the effectiveness of our TT-KNN model from a qualitative perspective. Due to page limitations, this discussion is limited to a single user. In Fig. 5, the blue markers indicate the actual nodes from the test dataset, while the green and red markers represent the nodes predicted by TT-KNN and GPT, respectively. The degree of overlap between the predicted and actual nodes indicates the accuracy of each model in capturing the ground truth. The overlap of TT-KNN with the actual data points exceeded that of GPT, indicating that TT-KNN is more effective in accurately predicting locations. Furthermore, TT-KNN proved effective in capturing temporal trajectories along the user's path. In contrast, GPT, which relies on previous sequences to predict future positions, was less consistent than TT-KNN in predicting the correct nodes.

Table 1: Comparison of GPT and TT-KNN in terms of average GEOBLEU and DTW

	GPT	TT-KNN
GEOBLEU	0.130049	0.2139882
DTW	94.289506	24.379585

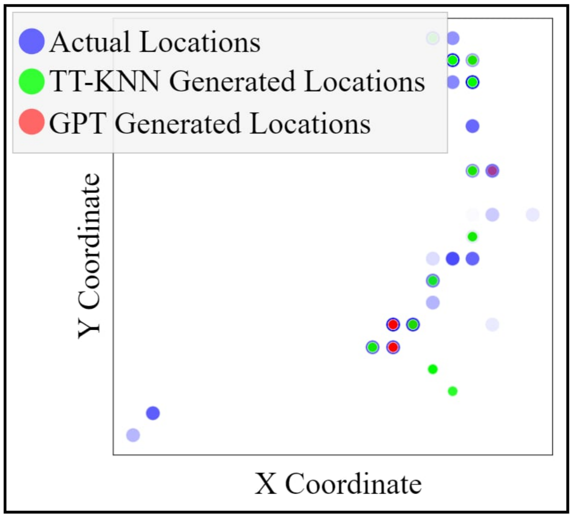

Figure 5: Scatter plot comparing the predicted values of TT-KNN versus those of GPT when applied to the test dataset

4 Conclusion

Researchers are increasingly applying AI to predict individuals' daily routines and their corresponding trajectories (referred to as mobility prediction), due to its considerable potential for commercial activities and public administration. However, most previous studies have focused exclusively on model theory or design without considering the specific characteristics of human flow trajectories. This paper presents a prediction scheme in which three fundamental characteristics of human mobility (time fixedness, contingency, and uniqueness) are used to fine-tune the TT-KNN algorithm, facilitating more accurate flow trajectory prediction. The proposed scheme addresses issues related to data imbalance, noise, and insufficient data volume. Experimental simulations verified the effectiveness of the proposed scheme when applied to the HuMob Challenge 2024 dataset. Note that due to time constraints, we were unable to implement adjustments for holidays or weekends. This will be addressed in future work. Moreover, this paper focuses exclusively on static locations. In the future, we will extend this research to the prediction of movements in dynamic environments (e.g., commuting by tram).

5 Acknowledgments

This work was supported in part by the National Science and Technology Council of Taiwan, R.O.C., under Contracts 112-2121-M-005 -006 and 113-2121-M-005 -003.

References

[1] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language Models are Few-Shot Learners. arXiv:2005.14165 [cs.CL] https://arxiv.org/abs/2005.14165

[2] 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* (Hamburg, Germany) *(HuMob-Challenge '23)*. Association for Computing Machinery, New York, NY, USA, 47–50. https://doi.org/10.1145/3615894.3628504

[3] Haoyu He, Xinhua Wu, and Qi Wang. 2023. Forecasting Urban Mobility using Sparse Data: A Gradient Boosted Fusion Tree Approach. In *Proceedings of the 1st International Workshop on the Human Mobility Prediction Challenge* (Hamburg, Germany) *(HuMob-Challenge '23)*. Association for Computing Machinery, New York, NY, USA, 41–46. https://doi.org/10.1145/3615894.3628507

[4] Taehoon Kim, Kyoung-Sook Kim, and Akiyoshi Matono. 2023. Cell-Level Trajectory Prediction Using Time-embedded Encoder-Decoder Network. In *Proceedings of the 1st International Workshop on the Human Mobility Prediction Challenge* (Hamburg, Germany) *(HuMob-Challenge '23)*. Association for Computing Machinery, New York, NY, USA, 37–40. https://doi.org/10.1145/3615894.3628503

[5] 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* (Hamburg, Germany) *(HuMob-Challenge '23)*. Association for Computing Machinery, New York, NY, USA, 7–10. https://doi.org/10.1145/3615894.3628506

[6] 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. https://doi.org/10.1145/3557915.3560951

[7] Aivin V. Solatorio. 2023. GeoFormer: Predicting Human Mobility using Generative Pre-trained Transformer (GPT). In *Proceedings of the 1st International Workshop on the Human Mobility Prediction Challenge* (Hamburg, Germany) *(HuMob-Challenge '23)*. Association for Computing Machinery, New York, NY, USA, 11–15. https://doi.org/10.1145/3615894.3628499

[8] Masahiro Suzuki, Shomu Furuta, and Yusuke Fukazawa. 2023. Personalized human mobility prediction for HuMob challenge. arXiv:2310.12900 [cs.LG] https://arxiv.org/abs/2310.12900

[9] 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* (Hamburg, Germany) *(HuMob-Challenge '23)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3615894.3628498

[10] Chenglong Wang and Zhicheng Deng. 2023. Multi-perspective Spatiotemporal Context-aware Neural Networks for Human Mobility Prediction. In *Proceedings of the 1st International Workshop on the Human Mobility Prediction Challenge* (Hamburg, Germany) *(HuMob-Challenge '23)*. Association for Computing Machinery, New York, NY, USA, 32–36. https://doi.org/10.1145/3615894.3628502

[11] 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. https://doi.org/10.1038/s41597-024-03237-9