Human Mobility Prediction Challenge: Next Location Prediction using Spatiotemporal BERT

HARU TERASHIMA
haru@ucl.nuee.nagoya-u.ac.jp
School of Engineering,
Nagoya University
Aichi, Japan

NAOKI TAMURA
tam@ucl.nuee.nagoya-u.ac.jp
Graduate School of Engineering,
Nagoya University
Aichi, Japan

KAZUYUKI SHOJI
shoji@ucl.nuee.nagoya-u.ac.jp
Graduate School of Engineering,
Nagoya University
Aichi, Japan

SHIN KATAYAMA
shinsan@ucl.nuee.nagoya-u.ac.jp
Graduate School of Engineering,
Nagoya University
Aichi, Japan

KENTA URANO
vrano@ucl.nuee.nagoya-u.ac.jp
Graduate School of Engineering,
Nagoya University
Aichi, Japan

TAKURO YONEZAWA
takuro@nagoya-u.jp
Graduate School of Engineering,
Nagoya University
Aichi, Japan

NOBUO KAWAGUCHI
kawaguti@nagoya-u.jp
Graduate School of Engineering,
Nagoya University
Aichi, Japan

Abstract

Understanding, modeling, and predicting human mobility patterns in urban areas has become a crucial task from the perspectives of traffic modeling, disaster risk management, urban planning, and more. HuMob Challenge 2023 aims to predict future movement trajectories based on the past movement trajectories of 100,000 users[1]. Our team, "uclab2023", considered that model design significantly impacts training and prediction times in the task of human mobility trajectory prediction. To address this, we proposed a model based on BERT, commonly used in natural language processing, which allows parallel predictions, thus reducing both training and prediction times.

In this challenge, Task 1 involves predicting the 15-day daily mobility trajectories of target users using the movement trajectories of 100,000 users. Task 2 focuses on predicting the 15-day emergency mobility trajectories of target users with data from 25,000 users. Our team achieved accuracy scores of GEOBLEU: 0.3440 and DTW: 29.9633 for Task 1 and GEOBLEU: 0.2239 and DTW: 44.7742 for Task 2[2][3].

HuMob-Challenge '23, November 13, 2023, Hamburg, Germany

ACM ISBN 979-8-4007-0356-0/23/11…$15.00
https://doi.org/10.1145/3615894.3628498

CCS Concepts: • **Human-centered computing** → *Visualization.*

Keywords: Human mobility, Next location prediction, Transformer

ACM Reference Format:
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 *1st International Workshop on the Human Mobility Prediction Challenge (HuMob-Challenge '23), November 13, 2023, Hamburg, Germany.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3615894.3628498

1 INTRODUCTION

Understanding, modeling, and predicting human mobility trajectories in urban areas are crucial tasks with applications spanning traffic modeling, disaster risk management, and urban planning. However, due to privacy concerns, open-source, large-scale mobility datasets are insufficient, making accuracy comparisons with alternative methods challenging. Thus, the HuMob Challenge 2023 was convened to develop and test methods for predicting mobility trajectories using the provided open-source dataset. This challenge specifically aimed to forecast future mobility trajectories based on the past movements of 100,000 users.

Within this challenge, two tasks were defined. Task 1 is to predicting the remaining 15 days of mobility trajectories for 20,000 target individuals using 80,000 individuals' 75-day historical mobility data and the 60-day data of the target individuals. Task 2 is to forecast the remaining 15 days of emergency mobility trajectories for 2,500 target individuals using the 60-day daily activity and 15-day emergency mobility data of 22,500 individuals.

In the context of mobility trajectory prediction tasks, the design of the model significantly affects training and prediction times. Therefore, when creating models, considerations must encompass both accuracy and execution time. With larger datasets, Transformer-based models can reduce execution times compared to RNN or LSTM-based models, which perform sequential predictions. In many time-series prediction tasks, RNN and LSTM-based seq2seq models have been commonly employed[4]. However, these models often fall short in capturing long-term mobility trends and require extensive learning time due to their sequential prediction nature. These issues can be addressed by leveraging Transformers, capable of long-term time-series modeling and parallel processing [5].

Consequently, we developed a model based on BERT[6] extended with Transformer, both of which are natural language processing models, to reduce training time and enable long-term time-series modeling. We refer to this model as the "Location Prediction BERT(LP-BERT)". The LP-BERT achieves reduced training and prediction times by efficiently parallelizing predictions for all location cells based on individual mobility sequences. As a result of employing the LP-BERT, we achieved accuracy scores of GEOBLEU: 0.3440 and DTW: 29.9633 for Task 1, and GEOBLEU: 0.2239 and DTW: 44.7742 for Task 2.

2 RELATED WORK

When predicting human mobility trajectories, temporal and spatial information becomes paramount. According to Yan et al[7], it is presented that embedding vectors that reflect different contexts based on the purpose of visiting functional locations prove advantageous in prediction. Consequently, the Context and Time aware Location Embedding (CTLE) was proposed to adaptively generate location embedding vectors tailored to contexts. Furthermore, according to Dejiang et al[8], it seeks to uncover movement patterns between functional zones and forecast people's subsequent moves within minutes or hours. The paper suggests integrating spatial-temporal influences into the LSTM model organically, presenting the Spatial-Temporal Long-Short Term Memory (ST-LSTM) model. According to Qiang et al[9], recognizing the significance of temporal interval information in mobility trajectory prediction, the proposal extends beyond local temporal context and advocates for modeling periodic time context as well. When creating the models utilized in this challenge, inspiration was drawn from these models presented in the aforementioned papers.

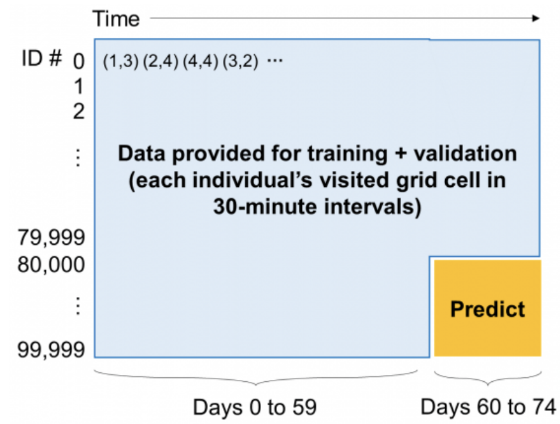

Figure 1. Dataset shape in task1

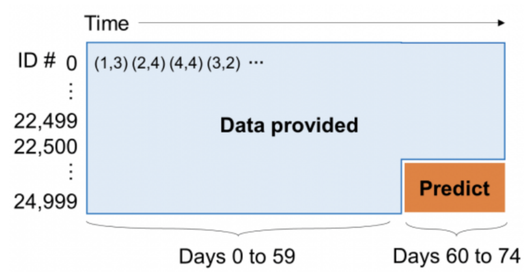

Figure 2. Dataset shape in task2

3 DATASET DESCRIPTION

In this challenge, a dataset of mobility spanning 75 days for 100,000 individuals within a major metropolitan area was provided. The target area was subdivided into $500m \times 500m$ cells, spanning a grid of 200×200. Individual movements were discretized at 30-minute intervals and 500m grid cells. We assigned Location IDs to each cell for prediction purposes.

Task 1 involved predicting the remaining 15 days of mobility trajectories for 20,000 target individuals using 80,000 individuals' 75-day mobility data and the 60-day data of the target individuals. Task 2, on the other hand, required forecasting the remaining 15 days of mobility trajectories in an emergency for 2,500 target individuals using 60-day daily activity mobility data, 15-day mobility data in an emergency, and the 60-day daily activity mobility data of 22,500 individuals. The specifics of the term "emergency" were not disclosed. The dataset columns were structured as follows.

- UserID
- Date($0 \sim 74$)
- Time($1 \sim 48$)
- (x , y) coordinates of stay cell

The dataset's structure is as shown in Figure 1 and Figure 2. In Task 1, when excluding the period of interest, the total

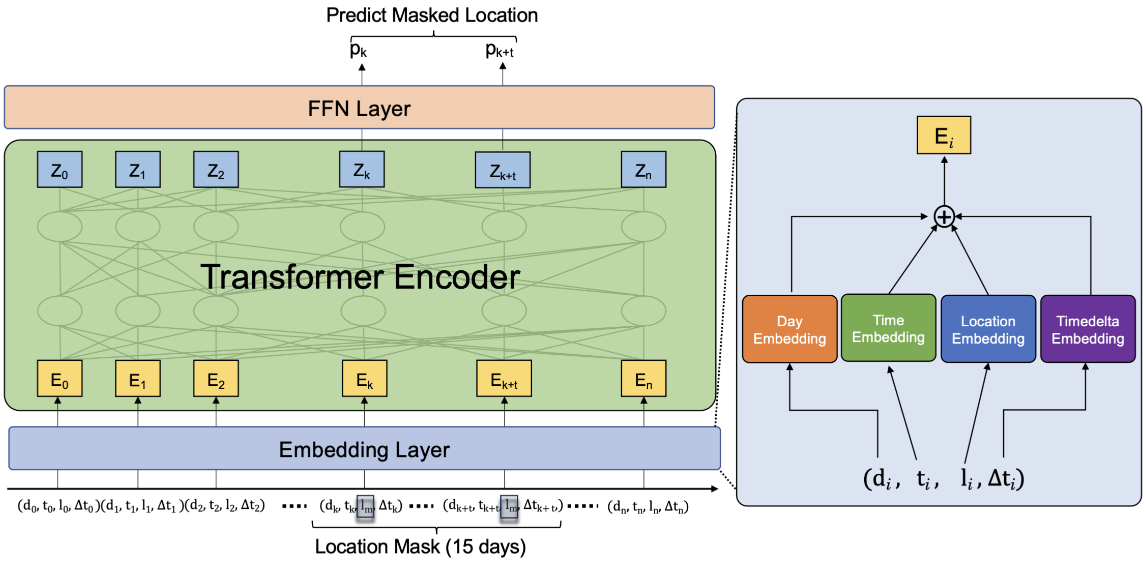

Figure 3. LP-BERT

number of records amounted to 107,204,010. The user with the highest record count possessed 3,497 records, while the user with the fewest had 463 records. Likewise, in Task 2, the total record count was 28,929,120, with the user holding the highest number of records totaling 3,293 and the user with the least amounting to 427 records. It is worth noting, however, that the mobility data is not captured continuously for all time periods, and there are instances of unobserved time intervals.

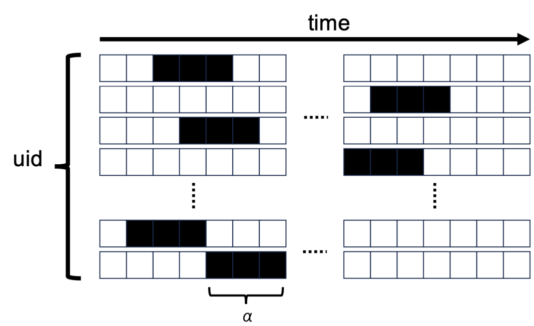

Figure 4. How to mask Location IDs

4 METHODOLOGY

Our team has developed the "LP-BERT", based on BERT, to capture long-term mobility trends. The schematic representation of the LP-BERT is shown in Figure 3. Within the LP-BERT, we define the number of layers in the Transformer encoder as "layersnum" and the number of attention heads as "headsnum."

In the input phase to the encoder, we employ an embedding layer to process date, time, Location ID, and the time difference from the previous movement, summing them together. From here, time differencen will be called timedelta. The inclusion of timedelta in the input is driven by the recognition that mobility data is not uniformly available across all time periods, and we consider the time timedelta from the previous movement to be a crucial piece of information. Additionally, as date and time information is provided for the prediction period, we consider them as valuable inputs for prediction.

The training process, shown in Figure 4, involves randomly masking only Location IDs for a consecutive α -day period from each user's complete mobility sequence. In this process, User ID is not used. Given that date and time information is available for the prediction period, we deem it unnecessary to mask them during the training phase. During prediction, the task is to predict the masked Location IDs.

In the LP-BERT, both in the training and prediction phases, all masked Location IDs are predicted in parallel. This parallel processing reduces both training and prediction times. During prediction, we observed a frequent occurrence of predicting multiple stays in the same Location on the same day. To mitigate this, when predicting the same location on the same day during training, we reduce the probability of staying at that location by a factor of β , reducing the probability of predicting that location during prediction, thereby reducing consecutive stay predictions..

5 VALIDATION RESULT

5.1 Datasets

In preparation for the final submission of our task, our team employed specific training data strategies. For Task 1, we utilized the entire mobility data from day 0 to day 74 for users with User ID ranging from 0 to 79,999, as well as data up to day 59 for users with User ID from 80,000 to 99,999. And for Task 2, we employed the complete mobility data from day 0 to day 74 for users with User ID ranging from 0 to 22,499, as well as data up to day 59 for users with User ID from 22,500 to 24,999.

The parameter settings for this endeavor included an embedsize of 128, 200 epochs, and a batch size of 16. Furthermore, we set the masking duration for days as α =15, and the multiplier for stay probabilities as β =0.9. The model configuration featured $layersnum = 4$, $headsnum = 8$.

5.2 Evaluation experiment

In the process of creating our final model, our team embarked on a thorough comparison of model accuracy employing various parameters. Additionally, we created a model based on BERT to capture long-term mobility trends, and we compared its accuracy with LSTM, a widely utilized model in trajectory prediction. To streamline the evaluation process, we focused on 50,000 users, as an exhaustive comparison across all users would have entailed lengthy execution times. In Task 1, we employed mobility data from day 0 to day 74 for users with User ID in the range of 0 to 39,999 for training, along with mobility data from day 0 to day 59 for users with User ID ranging from 40,000 to 49,999. The prediction period for this group was set as day 60 to day 74. In all cases, the number of epochs was unified at 50. We conducted comparisons for items that were not part of the model under evaluation, using the same conditions as our final submission model.

For evaluation, we adopted the GEOBLEU and Dynamic Time Warping (DTW) metrics, as used in this challenge. GEOBLEU, akin to similarity metrics in natural language processing, emphasizes local features, with higher values indicating greater similarity between two trajectories. In contrast, DTW is a method that compares entire trajectories and gradual adjustments, with smaller values denoting greater similarity between the two trajectories.

5.2.1 Examining models in LP-BERT.

Parameter tuning. Our team deliberated over the dimensionality of date, time, Location ID, and timedelta from the previous movement for embedding. Furthermore, we conducted a comparative analysis of the number of layers in the Transformer model. The outcomes of these considerations are shown in Table 1.

Changing the embedsize did not result in significant alterations in both GEOBLEU and DTW values. Although the differences were marginal, it was observed that the GEOBLEU values were most favorable when $embedsize = 128$, and the DTW values ranked second best. Consequently, we adopted $embedsize = 128$ for this configuration. Additionally, a comparison of the number of layers yielded no substantial distinctions in performance.

Consideration of timedelta. In the LP-BERT used in this study, we incorporated an input feature that embeds the timedelta from the previous movement. Table 2 presents the difference in accuracy when this feature was included and when it was not.

While there was a slight improvement in GEOBLEU accuracy, the DTW accuracy diminished. Therefore, it appears that the timedelta from the previous movement was not a crucial parameter in the prediction process.

Considering how to mask. In this challenge, due to the extended and continuous 15-day prediction period, our team believed that a longer masking period during the training phase would be beneficial. Furthermore, given the continuous nature of the prediction period, we assumed that maintaining consecutive masks during the training phase and providing the task to predict those portions would enhance accuracy. To test this hypothesis, we compared the results between the continuous masking and randomly masking 20% of the input sequence during the training phase, as shown in Table 3.

The results in Table 3 presents that both GEOBLEU and DTW metrics indicate superior accuracy when continuous masking is employed. Given the continuous prediction period in this task, we can conclude that maintaining consecutive masks during the training phase is advantageous.

Considerations for increasing movement rates. Our team discovered a pronounced trend in the LP-BERT, where the same location was consistently predicted during the forecasting stage. Due to the observation that the number of the types of locations appearing in predictions were fewer than those in the true data, we implemented an adjustment during the training phase. Specifically, when the model predicted the same location consecutively during training, we multiplied the probabilities by 0.9. Table 4 presents the comparison between the outcomes with and without this adjustment.

The GEOBLEU metric exhibited an improvement as a result of this adjustment, although DTW accuracy showed a decline. It is important to note that this adjustment was applied only to the probabilities of the locations predicted consecutively, and this could have potentially introduced unnatural patterns to the trajectories as a whole.

5.2.2 Comparison with LSTM. Furthermore, our team created a model using LSTM, a commonly used method in trajectory prediction, and compared its performance with the LP-BERT.

Table 1. Result of parameter tuning

	Embedsize			Layers			
Evaluation index	64	128	256	1	2	3	4
GEOBLEU	0.3111	0.3130	0.3120	0.3076	0.3126	0.3122	0.3130
DTW	34.35707	35.0853	36.0574	34.5410	35.4658	35.1392	35.0853

Table 2. Consideration of timedelta

	Implemented	Unimplemented
GEOBLEU	0.3130	0.3099
DTW	35.0853	34.8306

Table 3. Considering how to mask

	Continuous	Random
GEOBLEU	0.3130	0.2817
DTW	35.0853	45.4109

Table 4. Considerations for increasing movement rates

	Implemented	Unimplemented
GEOBLEU	0.3130	0.2984
DTW	35.0853	33.8231

Like the LP-BERT, we input embeddings of date, time, Location ID, and timedelta with previous movements, each in 128 dimensions. During the training phase, we applied masking to random Location IDs from the end of the time series movement data for each user. The number of Location IDs to mask was specified randomly within the range of 128 to minlen/4, with minlen representing the minimum sequence length within a batch. If the minimum sequence length was less than 128, we used minlen/4 as the number of Location IDs to mask.

In the LSTM model, both the encoder and decoder were utilized, and predictions were made sequentially, unlike the LP-BERT, which performed parallel predictions. Additionally, we used a single layer for the LSTM. The results are shown in Table 5 and Table 6.

From the results, it is evident that the LSTM model outperforms the LP-BERT in terms of both GEOBLEU and DTW. However, it is important to note that the LSTM model required more training and prediction times compared to the LP-BERT, which made parallel predictions. In summary, the LP-BERT demonstrated faster prediction capabilities, while the LSTM, which made sequential predictions using a decoder, displayed superior prediction accuracy. It is worth

Table 5. Comparison with LSTM

	LP-BERT	LSTM
GEOBLEU	0.3130	0.3673
DTW	35.0853	20.4769

Table 6. Average learning time per epoch

	LP-BERT	LSTM
time[seconds]	304	450

mentioning that as the dataset size increased, the Transformer's performance tended to improve. When we utilized data from 100,000 users for training, the accuracy was higher than when using data from 50,000 users. Therefore, with a larger dataset, the LP-BERT might outperform the LSTM in terms of accuracy.

6 CONCLUSION

In this paper, we proposed a method to reduce both training and prediction times by using the LP-BERT based on BERT, capable of parallel predictions, for the HuMobchallenge 2023. While sequential prediction models like LSTM tend to achieve higher accuracy on large datasets, they demand a significant amount of time for both training and prediction. To address this issue, we employed the Transformer architecture to enable parallel predictions. Moreover, by aligning the masking strategy with the prediction pattern, we successfully enhanced accuracy. As a result of these enhancements, our final scores for Task 1 were GEOBLEU: 0.3440 and DTW: 29.9633, while for Task 2, they were GEOBLEU: 0.2239 and DTW: 44.7742. It is worth noting that we used only the Transformer's encoder in this work, but there is potential to further improve accuracy by incorporating the decoder in future research.

Acknowledgements.

This research is supported by JST CREST(JPMJCR21F2), and NICT(222C01).

References

[1] Takahiro Yabe, Kota Tsubouchi, Toru Shimizu, Yoshihide Sekimoto, Kaoru Sezaki, Esteban Moro, and Alex Pentland. Metropolitan scale

and longitudinal dataset of anonymized human mobility trajectories. *arXiv preprint arXiv:2307.03401*, 2023.
[2] Toru Shimizu, Kota Tsubouchi, and Takahiro Yabe. Geo-bleu: Similarity measure for geospatial sequences, 2022.
[3] Pavel Senin. Dynamic time warping algorithm review. *Information and Computer Science Department University of Hawaii at Manoa Honolulu, USA*, 855(1-23):40, 2008.
[4] Pengpeng Zhao, Haifeng Zhu, Yanchi Liu, Zhixu Li, Jiajie Xu, and Victor S Sheng. Where to go next: A spatio-temporal lstm model for next poi recommendation. *arXiv preprint arXiv:1806.06671*, 2018.
[5] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.
[6] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.
[7] Yan Lin, Huaiyu Wan, Shengnan Guo, and Youfang Lin. Pre-training context and time aware location embeddings from spatial-temporal trajectories for user next location prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 4241–4248, 2021.
[8] Dejiang Kong and Fei Wu. Hst-lstm: A hierarchical spatial-temporal long-short term memory network for location prediction. In *IJCAI*, volume 18, pages 2341–2347, 2018.
[9] Qiang Liu, Shu Wu, Liang Wang, and Tieniu Tan. Predicting the next location: A recurrent model with spatial and temporal contexts. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, AAAI'16, page 194–200. AAAI Press, 2016.

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009