Cross-city-aware Spatiotemporal BERT

Meisaku Suzuki*
NTT DOCOMO, INC.
Tokyo, Japan
meisaku.suzuki.fw@nttdocomo.com

Yusuke Fukushima*
NTT DOCOMO, INC.
Tokyo, Japan
yuusuke.fukushima.fw@nttdocomo.com

Ryo Koyama*
NTT DOCOMO, INC.
Tokyo, Japan
ryou.koyama.aw@nttdocomo.com

Hayato Kumagai*
NTT DOCOMO, INC.
Tokyo, Japan
hayato.kumagai.zp@nttdocomo.com

Tomohiro Mimura*
NTT DOCOMO, INC.
Tokyo, Japan
tomohiro.mimura.ez@nttdocomo.com

Keiichi Ochiai*
NTT DOCOMO, INC.
Tokyo, Japan
ochiaike@nttdocomo.com

Abstract

Predicting human mobility has been actively studied for the past decade because of its various possible applications, such as traffic optimization and urban planning. Despite the increasing interest in human mobility prediction, the training and evaluation of prediction methods are often constrained by the use of different datasets (i.e., each study uses their own dataset for an evaluation). In considering these, the Human Mobility Prediction Challenge (HuMob Challenge) 2024 was held aiming at evaluating state-of-the-art models for the prediction of human mobility patterns using large-scale open dataset.

In this paper, we present our solution that ranked in the top 10 among over 100 participating teams. Our method uses LP-BERT as the base model, incorporating a component to account for cross-city and incorporating LSTM, which is suitable for series forecasting. The final prediciton is selected by ensembling these models and a rule-based method that considers the periodicity of human movement based on the probability of each model.

The proposed method is trained using data from all four cities (City A, B, C, and D) and evaluated on the target city data using metrics of GEOBLUE and DTW. As a result, we achieved the accuracy scores of GEOBLEU: 0.3008 and DTW: 23.75 for City B, GEOBLEU: 0.3027 and DTW: 18.71 for City C and GEOBLEU: 0.3305 and DTW: 49.75 for City D.

CCS Concepts

• **Information systems → Spatial-temporal systems**.

Keywords

Human Mobility, Trajectory Prediction

*Equal contribution

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

ACM ISBN 979-8-4007-1150-3/24/10
https://doi.org/10.1145/3681771.3699915

ACM Reference Format:
Meisaku Suzuki, Yusuke Fukushima, Ryo Koyama, Hayato Kumagai, Tomohiro Mimura, and Keiichi Ochiai. 2024. Cross-city-aware Spatiotemporal BERT. 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/10.1145/3681771.3699915

1 Introduction

Understanding and accurately predicting human mobility patterns in urban areas is crucial for optimizing transportation, disaster response, and urban planning. Recently, advanced models based on large-scale data collected from mobile devices and social media have been developed to forecast human movement in cities [3]. However, in most prior studies, the data used for train and test are from the same city. This approach limits the development of models that account for mobility patterns across different cities. Moreover, the training and evaluation of prediction methods are often constrained by the use of different datasets (i.e., each study uses their own dataset for an evaluation). This makes it difficult to fairly compare with other methods. The Human Mobility Prediction Challenge (HuMob Challenge) 2024 aims to address these challenges by providing datasets from four distinct metropolitan areas. Participants are tasked with predicting mobility patterns for each city. The datasets cover the mobility trajectories of 6,000 to 100,000 individuals over a period of 75 days, offering participants the opportunity to leverage this large and diverse dataset to improve the accuracy of their predictions for each city.

In this paper, we introduce a solution that is placed in top 10 in the competition. Our solution consists of three prediction models; (1) the Location Prediction BERT (LP-BERT) model [11], which was the winning solution in HuMob challenge 2023 and incorporates a component to incorporate cross-city features, (2) LP-BERT model with an LSTM layer which is added before the fully connected layers in LP-BERT, and (3) a rule-based method that considers the periodicity of human movement. The final prediction is selected by ensembling these three models based on the probability of each model.

2 Related Work

Human mobility prediction, also known as next location prediction, has been widely studied in the last decade. According to the recent survey paper [3], next location prediction can be categorized into two methods; one is statistical method, and the other is a machine learning-based method. For statistical methods, there are many methods that use Markov chains, such as [1]. However, Markov-based models suffer from the difficulty of representing complex patterns in human mobility [4], recent studies focus on machine learning-based method, especially on deep learning-based method.

Methods based on deep learning models such as Convolutional Neural Networks (CNN) [2], Recurrent Neural Networks (RNN) [7], graph neural networks (GNN) [12], and Transformer [6, 11] have been proposed. In a method called CTLE (Context and Time aware Location Embedding), location embedding is generated using a BERT-based method that considers context, including time and location, to improve the accuracy of movement prediction. Terashima et al. [11] proposed an improved version of CTLE called Location Prediction BERT (LP-BERT), which won first place at last year's HuMob Challenge.

Our method differs from existing works owing to the following reasons:

- To improve prediction accuracy in cross-city setting, we add a component to represent city embedding.
- Because location prediction is sequential prediction, we customize LP-BERT model to incorporate LSTM layer that is suitable for sequential prediction.

3 Preliminaries

3.1 Datasets

This study utilizes trajectory data calculated based on the location history of smartphones provided by LY Corporation [13]. The target areas include four metropolitan regions in Japan (cities A, B, C, and D), somewhere in Japan. Each area is divided into 500-meter by 500-meter cells, forming a 200 x 200 grid. The human mobility datasets contain the movement of individuals across a 75-day period.

Each record is described as (cid, uid, day, time, x, y), where cid represents the city, uid represents the user, day indicates the number of days elapsed from a reference date, time denotes the elapsed time, divided into 30-minute intervals throughout the day, and x and y indicate the coordinates.

3.2 Problem Statement

The goal of this competition is to predict the future trajectories of 3000 individuals each in cities B, C, and D, during days 61 to 75, using movement data of individuals in city A (100,000 individuals' full trajectories from day 1 to 75) and cities B, C, D (total of 25,000, 20,000, and 6,000 individuals each).

4 Methodologies

4.1 Overview

The solution consists of an ensemble of three methods: 1. LP-BERT, 2. A model that incorporate an LSTM into LP-BERT before a MLP layer, and 3. A machine learning approach for selecting the most frequent location.

4.2 Cross-city LP-BERT

This method is based on LP-BERT, which was the winning solution in Human mobility challenge 2023. Because data from four cities is provided in this year's competition, a Cross-city LP-BERT is constructed to improve accuracy by performing data augmentation using the city data for each of these cities. In the original LP-BERT, four features—date, time, Location ID, and time difference from the previous movement—were added to the embedding layer. We enhance the model by adding city embeddings to these four features to incorporate city-specific characteristics. Figure 1 illustrates the overview of adding city embeddings. During training, similar to the original LP-BERT method, a random movement sequence from each user is selected, and Location IDs for consecutive n days are masked, with the model predicting those masked Location IDs.

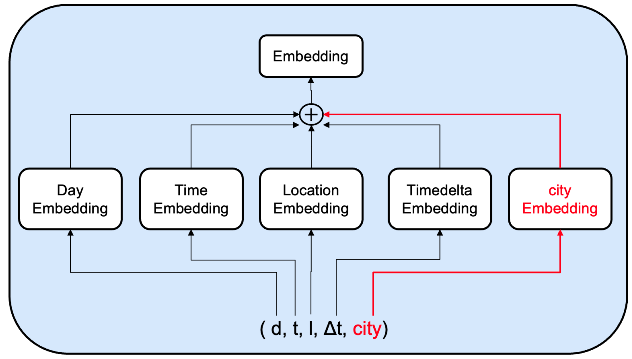

Figure 1: Embedding Layer of Cross-city LP-BERT

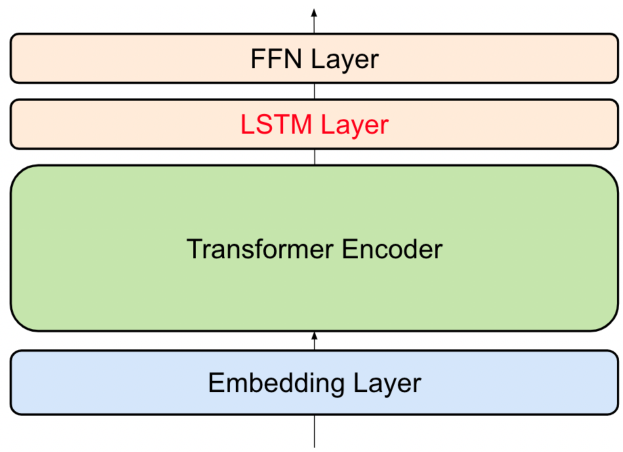

Figure 2: Add LSTM Layer to LP-BERT

4.3 LP-BERT with LSTM Layer

In LP-BERT, since the final prediction is made using fully connected layers after the Transformer layers, we added an LSTM before the fully connected layers. This allows for predictions that take into account the time series of each user. Figure 2 shows an overview of adding an LSTM to the fully connected layers. Additionally, by using the predicted values from the LSTM Layer, we were able to increase diversity in the final ensemble.

Algorithm 1 Calculate Time-Decayed Most Frequent Location [5]

Require: $k_1, k_2, test\_end\_day = 74$

Initialize $locationCount[user][adjustedTimeStep]$ as empty dictionaries
for each $user$ in $users$ **do**
  for $day \leftarrow 0$ to 59 **do**
    for $timeStep \leftarrow 0$ to 47 **do**
      $adjustedTimeStep \leftarrow timeStep$ for Method 1
      $adjustedTimeStep \leftarrow timeStep//3$ for Method 2
      $location \leftarrow$ location of $user$ at day and $timeStep$
      $weight \leftarrow e^{-k \times (test\_end\_day - day)}$ where $k = k_1$ for Method 1 and $k = k_2$ for Method 2
      $locationCount[user][adjustedTimeStep][location] \leftarrow locationCount[user][adjustedTimeStep][location] + weight$
    end for
  end for
end for
for each $user$ in $users$ **do**
  for $adjustedTimeStep \leftarrow 0$ to $adjustedMaxTimeStep$ **do**
    $mostFrequentLocation[user][adjustedTimeStep] \leftarrow$ max key of $locationCount[user][adjustedTimeStep]$ by value
  end for
end for

4.4 Time-Decayed Most Frequent Location

We employ a rule-based method [5] based on the observation that human behaviors tend to repeat approximately every 24 hours. This method achieved third place in the 2023 HuMob Challenge and consists of the following three components:

- **Method 1**: A model that calculates the time-decayed most frequent grid using a 30-minute time window.
- **Method 2**: A model that calculates the time-decayed most frequent grid using a 90-minute time window.
- **Method 3**: GBDT models that determines, for each user and each day, which of Method 1 or Method 2 is optimal.

Method 1 is expected to achieve high accuracy on working days, where behavioral regularity appears in finer time windows. In contrast, Method 2 is expected to be effective on days when behavioral patterns are less pronounced, such as non-working days. The algorithms for Method 1 and Method 2 are presented in Algorithm 1. Time decay refers to a process that computes a weighted most frequent grid, assigning greater importance to data closer to the prediction date.

Method 3 employs ensembled GBDT models using features derived from each user's past behavioral statistics. These models predict, for each user and each day, which trajectory—between Method 1 and Method 2—minimizes the GEO-BLEU score.

4.5 Ensemble

We perform an ensemble of the three methods described in Section 4.2, 4.3 and 4.4. Since LP-BERT had consistently high geo-bleu and DTW scores across all cities in our experiment, we use LP-BERT as the base model. When the probability score of LP-BERT's predictions is low, we adopted the predictions from the LP-BERT with LSTM Layer. If the probability score of the LP-BERT with LSTM Layer is also low, we use the predictions from a rule-based method.

5 Experiments

5.1 Experimental Setting

Metrics. In this competition, Dynamic Time Warping (DTW) [8] and GEO-BLEU [10] were used as metrics. DTW is a measure to evaluate the distance of two given sequences by considering alignment of each sequence. GEO-BLEU is a similarity measure for geospatial sequences inspired by the Bilingual Evaluation Understudy (BLEU)[9] which is often used to evaluate a performance of machine translation.

Hyper-parameters. For LP-BERT model, we configured the batch size to be 128, the number of embedding dimension to be 256 and performed 50 epochs with learning rate of 1e-4 using Adam optimizer. For rule-based method, we set $k_1 = 0.0063$, $k_2 = 0.0178$ in Algorithm 1. These were selected as GEO-BLEU was minimum in offline data.

5.2 Results and discussion

The results for each city for the three previously described methods and the results obtained by the ensemble are shown in Table 1. For all metrics and all cities, our Cross-city LP-BERT outperformed the original LP-BERT. For GEO-BLEU in city B and D, the rule-based method was the best performace except for the result of ensemble. The thresholds used in the ensemble were selected with emphasis on DTW in City B and GEO-BLEU in Cities C and D. Therefore, our submitted results are not the best if we focus only on GEO-BLEU in city B and DTW in Cities C and D. As for stand-alone performance, LP-BERT trained on all cities for GEO-BLEU and the frequency-based method for DTW are better, but we aimed to obtain results that balance both by ensemble.

In Table 1, only Time-dacayed Most Frequent Location is a statistical method; the others use machine learning. Differences can be seen between these two methods. GEO-BLEU tended to be higher for the statistical method, while DTW tended to be better for the machine learning method. The machine learning methods used in this study, such as BERT and LSTM, are learned by considering the immediately preceding time series data. Statistical methods, on the other hand, do not explicitly consider the immediately preceding location information to estimate the next location information. Therefore, the performance of the machine learning method was considered to be better in GEO-BLEU, where it is more important to get the location correct in succession.

5.3 Ablation study

5.3.1 Impact of other city data. From Table 3, we can see that the best performance is obtained when data from all cities are used for training, but at the same time, it is city A, which has the largest amount of data, that contributes the most to the improvement in accuracy. The amount of data is important, but it can also be assumed that the content of the data suggests that City A is a large city and therefore contains all types of movement patterns, which would have had a positive impact on the learning for any of the cities. The other city data also had varying degrees of positive impact, but basically, adding them always had a positive effect.

Table 1: Results for each city by each method

Method	GEO-BLEU			DTW		
	city B	city C	city D	city B	city C	city D
LP-BERT	0.2838	0.2831	0.2817	24.90	20.72	46.38
LP-BERT w/ all city data (Cross-city LP-BERT)	0.3024	0.3018	0.2995	24.08	**18.35**	**39.33**
LP-BERT w/ LSTM head	0.2875	0.2856	0.2810	28.32	20.98	45.93
Time-Decayed Most Frequent Location (Rule-based method)	**0.3041**	0.2940	0.3303	31.11	23.67	50.26
Ensemble	0.3008	**0.3027**	**0.3305**	**23.75**	18.71	49.75

Table 2: Result of adding day-of-week feature

Method	city C	
	GEO-BLEU	DTW
LP-BERT w/ all city data	**0.3018**	18.35
+ day of week feature	0.2923	**17.91**

Table 3: Results when data for each city is added to each

Method	GEO-BLEU			DTW		
	city B	city C	city D	city B	city C	city D
LP-BERT	0.2838	0.2831	0.2817	24.90	20.72	46.38
w/ A data	0.2984	0.2951	0.2995	**24.06**	18.51	41.39
w/ B data	-	0.2840	0.2884	-	19.93	42.51
w/ C data	0.2842	-	0.2796	25.05	-	41.72
w/ D data	0.2858	0.2765	-	25.11	20.04	-
w/ all data	**0.3024**	**0.3018**	**0.2995**	24.08	**18.35**	**39.33**

Given that the nature of each city varies widely, it is conceivable that there are patterns of movement that are independent of the nature of the city.

5.3.2 Addition of day-of-week feature. As an additional validation that could not be performed during the competition, Table 2 shows the results of adding the day-of-week information explicitly as a 1-hot feature, and although the results are limited to city C, GEO-BLEU deteriorated while DTW improved.

6 Conclusion

In this paper, we have proposed a solution to address the user location prediction problem in the HuMob Challenge 2024 and demonstrate its effectiveness. Our solution included three prediction models; (1) the LP-BERT with cross-city consideration, (2) a model in which the head of LP-BERT is changed to LSTM, and (3) a rule-based method that considers the periodicity of human movement. Finally, we proposed combined there three models based on confidence scores, achieving superior GEOBLEU and DTW. From the experient results, we revealed that incorporating comprehensive city information (not only target city information) as a feature and vectorizing it improves user location prediction accuracy. Moreover, incorporating the day of the week as a feature could enhance prediction accuracy, as it explicitly captures the temporal periodicity in the data.

There are several futrue work. Further tuning of the deep learning model's hyperparameters could improve prediction performance. Besides, integrating user persona information into the model could enable movement predictions that account for individual user characteristics. Using the methodology of this study, it is possible to predict both the habitual movements of individuals and typical movements in various urban areas, suggesting potential applications across multiple fields.

References

[1] Daniel Ashbrook and Thad Starner. 2002. Learning significant locations and predicting user movement with GPS. In *Proceedings. Sixth International Symposium on Wearable Computers,*. IEEE, 101–108.

[2] Meng Chen, Yixuan Zuo, Xiaoyi Jia, Yang Liu, Xiaohui Yu, and Kai Zheng. 2020. CEM: A convolutional embedding model for predicting next locations. *IEEE Transactions on Intelligent Transportation Systems* 22, 6 (2020), 3349–3358.

[3] Giovanni Garola, Chiara Siragusa, Arianna Seghezzi, and Riccardo Mangiaracina. 2024. Next Place Prediction Model: A Literature Review. *Emerging Cutting-Edge Developments in Intelligent Traffic and Transportation Systems* (2024), 21–30.

[4] Ye Hong, Henry Martin, and Martin Raubal. 2022. How do you go where? improving next location prediction by learning travel mode information using transformers. In *Proceedings of the 30th International Conference on Advances in Geographic Information Systems*. 1–10.

[5] Ryo Koyama, Meisaku Suzuki, Yusuke Nakamura, Tomohiro Mimura, and Shin Ishiguro. 2023. Estimating future human trajectories from sparse time series data. In *Proceedings of the 1st International Workshop on the Human Mobility Prediction Challenge*. 26–31.

[6] Yan Lin, Huaiyu Wan, Shengnan Guo, and Youfang Lin. 2021. 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*, Vol. 35. 4241–4248.

[7] Qiang Liu, Shu Wu, Liang Wang, and Tieniu Tan. 2016. Predicting the next location: A recurrent model with spatial and temporal contexts. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 30.

[8] Meinard Müller. 2007. Dynamic time warping. *Information retrieval for music and motion* (2007), 69–84.

[9] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. BLEU: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*. 311–318.

[10] 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*. 1–4.

[11] 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*. 1–6.

[12] Dongjing Wang, Xingliang Wang, Zhengzhe Xiang, Dongjin Yu, Shuiguang Deng, and Guandong Xu. 2021. Attentive sequential model based on graph neural network for next poi recommendation. *World Wide Web* 24, 6 (2021), 2161–2184.

[13] 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 (2024), 2052–4463.