

CITY-SCALE HUMAN MOBILITY PREDICTION MODEL BY INTEGRATING GNSS TRAJECTORIES AND SNS DATA USING LONG SHORT-TERM MEMORY

S. Miyazawa^{a*}, X. Song^b, R. Jiang^c, Z. Fan^b, R. Shibasaki^a, and T. Sato^d

^a Center for Spatial Information Sciences, University of Tokyo, Japan - (koitaroh, shiba)@csis.u-tokyo.ac.jp

^b SUSTech-UTokyo Joint Research Center on Super Smart City, Department of Computer Science and Engineering, Southern University of Science and Technology (SUSTech), China and University of Tokyo, Japan - (songxuan@csis.u-tokyo.ac.jp, fanzipei@iis.u-tokyo.ac.jp)

^c Information Technology Center, The University of Tokyo, Japan - jiangrh@csis.u-tokyo.ac.jp

^d ZENRIN DataCom CO., LTD., Japan - t_sato@zenrin-datacom.net

Commission IV, WG IV/5

KEY WORDS: Human mobility, Location-based social network, GNSS trajectory, Human mobility prediction, Deep learning

ABSTRACT:

Human mobility analysis on large-scale mobility data has contributed to multiple applications such as urban and transportation planning, disaster preparation and response, tourism, and public health. However, when some unusual events happen, every individual behaves differently depending on their personal routine and background information. To improve the accuracy of the crowd behavior prediction model, understanding supplemental spatiotemporal topics, such as when, where and what people observe and are interested in, is important. In this research, we develop a model integrating social network service (SNS) data into the human mobility prediction model as background information of the mobility. We employ multi-modal deep learning models using Long short-term memory (LSTM) architecture to incorporate SNS data to a human mobility prediction model based on Global Navigation Satellite System (GNSS) data. We process anonymized interpolated GNSS trajectories from mobile phones into mobility sequence with discretized grid IDs, and apply several topic modeling methods on geo-tagged data to extract spatiotemporal topic features in each spatiotemporal unit similar to the mobility data. Thereafter, we integrate the two datasets in the multi-modal deep learning prediction models to predict city-scale mobility. The experiment proves that the models with SNS topics performed better than baseline models.

1. INTRODUCTION

Today, more than 54 percent of the world's population lives in urban areas (66 percent by 2050)¹. Cities are full of opportunities and services that attract new people. However, cities are also the places with the most serious urban issues such as problems relating to transportation, public safety, and public health. Emerging cities are facing, and will face, unprecedented problems that need to be solved with new ideas and technologies.

Urban dynamics and large-scale human mobility has been major challenges in Urban computing (Zheng et al., 2014) and in recent years, analyzing, mining and visualizing geospatial big data from new sources for decision-support, is also considered to be some of the most important challenges in this era of big data (Li et al., 2016).

The high penetration rate of mobile phones, especially smart phones, enables systematic data collection for longer periods; moreover, users are more willing to connect to multiple services and allow service providers to collect user data (Birenboim and Shoal, 2016). These provide an inexpensive means of collecting data on city-scale mobility.

Traditionally, survey-based human mobility data collect certain background information on mobility. They often include individuals' mobility purpose, home location, household size, and job classification, which complement mobility models and theories.

However, even though digital traces such as GNSS and call detail records (CDR) are accurate and precise, the data lack annotation compared to survey-based data. To fully utilize the data in many estimation, prediction, or classification models, supplemental information with high accuracy and precision are crucial (Gong et al., 2015). One possible solution is to use location-based social network (LBSN) services such as Twitter for the background information (Zhang et al., 2012). LBSN is one kind of SNS that utilizes location information for the services. However, compared to GNSS dataset, trajectory from LBSN is usually sparse and highly biased. To model and understand the city-scale mobility pattern, we need a to extract background information and combine it with the large-scale human mobility dataset.

Our main idea is to develop a city-scale human mobility prediction model by integrating GNSS trajectories and social network service (SNS) data, in order for the human mobility prediction model to achieve more accurate predictions and to have additional application capabilities.

The key idea of this work is summarized in Figure 1 and the work offers the following key contributions, which highlight its uniqueness compared to previous research.

Our key contributions of this paper are as follows:

- Integration of background information for human mobility: the model incorporate spatiotemporal topics from Twitter data as the background information of human mobility,
- Efficient multi-source data integration: multi-source data integration requires novel prediction model structure for efficient learning of inter-source relationship. Mobile phone

*Corresponding author.

¹<http://www.un.org/en/development/desa/news/population/world-urbanization-prospects-2014.html>

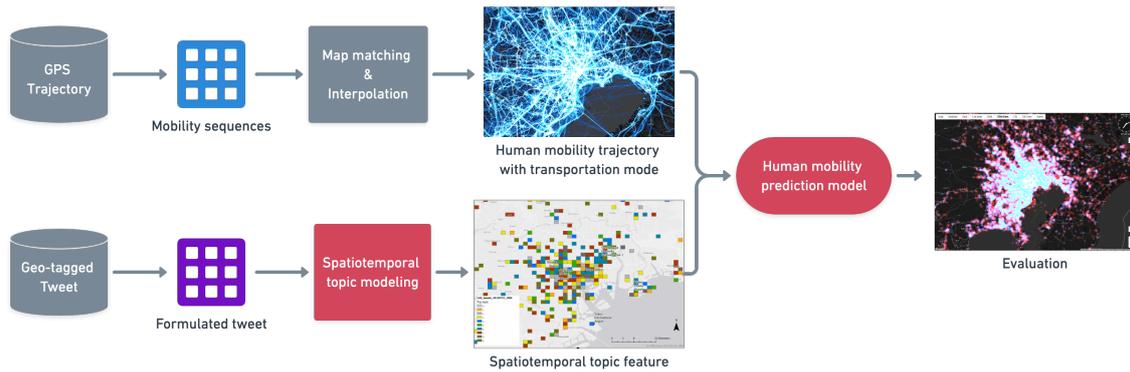


Figure 1: Can we use the cross-domain data for human mobility prediction? Our main contributions are represented in red. Credit: "Konzatsu- Tokei(r)" (c)ZENRIN DataCom CO., LTD.

GNSS trajectory and geo-tagged SNS data are produced differently. Therefore, the resulting difference in the spatiotemporal distributions makes data integration challenging.

The rest of this paper is organized as follows. The related work is discussed in Section 2. We introduce the data and preliminaries for the model in Sections 3 and 4, respectively. Thereafter, in Section 5, we define our prediction scope and experimentally evaluate our methods using real data in Section 6; finally, Section 7 concludes the paper.

2. RELATED WORK

2.1 City scale human mobility analysis and modeling

There have been several studies with location logs and trajectories starting from survey-based (Grinberger and Shoval, 2015, Spaccapetra et al., 2008) to large-scale mobile phone data for city-scale or nation-scale mobility (Calabrese et al., 2013, Kang et al., 2012).

There have been several significant areas of analysis and data mining using mobility data. One typical area is urban analysis, where researchers discover features and evaluate problems in cities. Studies on this topic have been usually motivated by the urban planning perspective such as population estimation (Xu et al., 2016), mobility pattern (Reades et al., 2009), and discovering urban functional zones (Yuan et al., 2012).

2.2 Human mobility prediction

Specifically, mobility prediction focuses on providing information for future decision making. Initially, having large scale human mobility data such as mobile phone GNSS trajectory, the trajectory mining approach enables human mobility prediction models. Song et al. (Song et al., 2010) developed a city-scale human mobility prediction models and showed the high prediction accuracy of human mobility in regular days. More specific prediction scheme, such as disaster evacuation behavior with supplemental disaster information (Song et al., 2014), and taxi travel time estimation (Tang et al., 2016) are also gaining popularity. These specific schemes usually limit the scopes but achieve better accuracy for particular purposes.

2.3 Location based social network (LBSN)

Even though less than 1% of tweets in Twitter are geo-tagged (Morstatter et al., 2013) the mobility data from LBSN shows several biases such as location bias (sparsity), time bias (more data in evening), and attribute bias (more young people and wealthy region) (Mota et al., 2015). The Twitter data has been one of the most widely used LBSN².

In addition to a typical application of using LBSN data as location data (Sakaki et al., 2013), Another application is to use LBSN data as a background information. LBSN data potentially have rich information on people's sentiments, observations, and thoughts. Some notable examples are discovering traffic anomalies (Pan et al., 2013), topic association among cities (Liu et al., 2016), and population density estimation by combining GNSS point density and topics from tweets (Miyazawa et al., 2019).

2.4 Deep Learning on Urban Computing

One of the most significant advancements in Machine Learning is Deep Learning. Urban computing community has indeed adopted the technology for many applications. For mobility prediction, one of the most important concepts is recurrent neural network and especially its extension long short-term memory network (Hochreiter and Schmidhuber, 1997), which enable time series input/output. Significant studies on this application includes traffic flow prediction using taxi GNSS trajectory (Niu et al., 2015), anomaly detection on pedestrian accidents (Zhou et al., 2018), and Region of Interest (ROI)-based short-term human mobility prediction (Jiang et al., 2018a). Another important characteristic of Deep Learning is "multimodality", which enable a model to combine different modal data structures by concatenating input vectors or linking network units. Ngiam et al. introduced multimodal deep learning and demonstrated its superiority over unimodal structure in their experiments (Ngiam et al., 2011). Zheng et al. introduced a multi-source model to detect urban anomalies that are not found by only using one data source (Zheng et al., 2015); they later elaborated the concept with a comprehensive review (Zheng, 2015). There has been an increasing popularity and interest in multimodal deep learning in urban computing communities (Song et al., 2014).

Compared to the aforementioned related work, we intend to apply the methodologies of Feature Extraction and LBSN on the

²Twitter disables "precise location" attached to tweets in 2019. This limits the applicability of this study to future data (<https://twitter.com/TwitterSupport/status/1141039841993355264>).

Deep Learning-based application to improve the accuracy while adding analytical capabilities for the prediction results. Also, we apply the model on a regular day to evaluate the general analytical capability.

3. DATA

3.1 City-scale interpolated and anonymized mobile phone GNSS trajectory

GNSS trajectories used in this study are from anonymized mobile phone users throughout the Greater Tokyo Area from July 1, 2012 to July 31, 2012, which are processed by NTT DOCOMO, INC. Now, NTT DOCOMO INC. collected an anonymous GNSS log dataset, "Konzatsu-Tokei (R)" Data in Japan over a three-year period (Aug 1, 2010 to July 31, 2013). "Konzatsu-Tokei (R)" Data refers to data on the flow of people collected by individual location data sent from mobile phones with the users' consent, through Applications³ provided by NTT DOCOMO, INC. These data are processed collectively and statistically in order to conceal the private information. Original location data are GNSS data (latitude, longitude) sent in a minimum recurring period of approximately five minutes; however, they do not include information on individuals. One important drawback of the data is that the number of points is biased toward daytime (people are awake and moving) and down towns (greater population and activities involving mobility).

Additionally, each trajectory is processed through map matching on transportation network and spatiotemporally interpolated to have a minute interval, and each point is on a transportation network link or node. It also has estimated transportation mode labels (Table 1). This interpolation improved the human mobility prediction result in the previous work (Song et al., 2016).

ID	Description
1	Stay
2	Walk
3	Bicycle
4	Car
5	Train
6	Unknown

Table 1: Estimated mobility mode labels from a previous study (Song et al., 2016).

3.2 Geo-tagged tweets

The geo-tagged tweets from July 1, 2012 to July 31, 2012 were collected using Twitter API. To extract tweets concerning mobility and social activity, only tweets posted from check-in services (e.g. Swarm: a LBSN service by Foursquare Labs Inc.) are used in the following experiment.

3.3 Preprocessing text

The preprocessing is the essential in natural language processing and we follow our previous work (Miyazawa et al., 2019). After the text cleaning to remove noise and irrelevant text, we conduct word segmentation and normalization for part-of-speech (POS) tagging, then finally we remove "stop words" such as Japanese particles, auxiliary verbs, and pronouns. to optimize the learning process.

³Some applications such as "docomo map navi" service(map navi local guide).

4. PRELIMINARIES

4.1 Human mobility trajectory

In this study, similar to the human mobility data, GNSS trajectories from mobile phones are employed. The raw GNSS trajectory is structured as a set of 4-tuple (1):

$$X_{raw} = \{(user, timestamp, lat, lon)\} \quad (1)$$

where *user*, *timestamp*, *lat*, and *lon* are user ID, timestamp, latitude, and longitude, respectively. Typically, GNSS modules in mobile phones creates records in a certain time interval, which varies depending on the scope of the application. For example, navigation applications would measure and record the location in every few seconds, as the applications require frequent location measurement for accurate and precise navigation. In contrast, personal logging applications would measure and record the location in every few minutes, and sometimes it only records the location if some movement is detected in order to save batteries. Additionally, most applications can only record the locations when GNSS or a cell signal is available. Therefore, the timestamp in raw GNSS trajectory usually has no stable interval. If people are in an indoor environment or underground, there would be missing trajectories in the individuals GNSS trajectory.

Additionally, the interpolated GNSS trajectory (Song et al., 2016) we mainly use in this study is structured as a 5-tuple set (2):

$$X_{interpolated} = \{(user, timestamp, lat, lon, mode)\} \quad (2)$$

where *user*, *timestamp*, *lat*, *lon*, and *mode* are user ID, timestamp, latitude, longitude, and travel mode label (Table 1). The location (latitude, longitude) is then encoded as a spatial index, in accordance with the third level Japan Industrial Standards (JIS) X0410 grid square code denoted as *l*. The third level JIS X0410 code is encoded in 10 numeric digits and each grid spans 30 arc-seconds in latitude and 45 arc-seconds in longitude, which approximately match 1 km of the study area.

4.2 Modeling human mobility sequence

Let $\mathbb{U} = \{u_1, u_2, \dots\}$ be a set of users and $\mathbb{T} = \{t_1, t_2, \dots\}$ be a set of temporal indexes with a constant time interval Δt . To model human mobility data for prediction model, we process the dataset to produce human mobility sequence for each user, i.e.,

$$HMS = \{hms_{u_1}, hms_{u_2}, \dots\} \quad (3)$$

In the human mobility sequence of each user $hms_u = r_{t_1}, r_{t_2}, \dots$, *r* is a set of location denotation and supplemental labels for the each temporal index. It is structured as a set of either 3-, 4-, or 5-tuples (4, 5, 6) depending on the input for the prediction model.

$$r_{raw} = \{(l)\} \quad (4)$$

$$r_{mode} = \{(l, mode)\} \quad (5)$$

$$r_{mode,topic} = \{(l, mode, topic)\} \quad (6)$$

where 4 is for the model with only the location code l , 5 is for the model with the location code l and the transportation mode code $mode$, and 6 is for the model with the location code l , transportation mode code $mode$, and the topic feature $topic$ respectively.

4.3 Modeling topic from SNS data

We use Twitter data and discovered topics from the data and integrate it to the model as the background information of the mobility. The tweet dataset is also structured as a table containing the user id, latitude, longitude, timestamp, and raw tweet. Supplemental components of original tweets such as URLs, hashtags, usernames, and location names in “check-in” tweets are removed. For further processing, we use the following equations to define tweets:

$$tw = \{(user, timestamp, lat, lon, \mathbf{w})\} \in \mathbb{TW} \quad (7)$$

and the text in tweets:

$$\mathbf{w} = \{(w_1, w_1, \dots)\} \in \mathbb{V} \quad (8)$$

A tweet tw is a 5-tuple where $user$, $timestamp$, lat , and lon correspond to user ID, timestamp, latitude, and longitude, respectively, of a user with corresponding timestamps, and \mathbf{w} is a bag-of-words containing N_i words. Let the vocabulary $\mathbb{V} = \{1, 2, \dots\}$ be a set of word IDs so that each word appears in the collection of words V at least once.

4.4 Topic Modeling from Twitter data

Several topic modeling models originally developed in (Miyazawa et al., 2019) are applied on the geo-tagged tweets.

The first model is the latent semantic analysis (LSA) based on an “online incremental streamed distributed training algorithm” (Řehůřek, 2011). LSA takes the frequency-inverse term frequency matrix as the input and computes a low-rank approximation of the input matrix using singular value decomposition:

$$\mathbf{X}_{numterm \times numdoc} = \mathbf{U}_{numterm \times k} \mathbf{\Sigma}_{k \times k} \mathbf{V}_{numdoc \times k}^T \quad (9)$$

where \mathbf{X} is the term-document frequency matrix, \mathbf{U} and \mathbf{V} are orthogonal matrices and $\mathbf{\Sigma}$ is a diagonal matrix; $numterm$ is the number of terms, $numdoc$ is the number of documents, and k is the dimension size (the number of topics). The i th column in X represents a vector corresponding to the i th document in relation to each term, while the i th column in $\mathbf{V}(d_i)$ becomes the vector corresponding to the i th document in the low dimensional space, where the number of topics = k ; finally, the $\mathbf{\Sigma}_k d_i$ for each document will be saved and used with the regression models.

Secondly, Latent Dirichlet Allocation (LDA) is performed. LDA is a probabilistic extension of LSA (Blei et al., 2003). LDA assumes that a set of documents are derived from k topics through a generative process where each topic has a multinomial distribution $\beta_k \sim Dirichlet(\eta)$ over the vocabulary. For each document d , the distribution over topics $\theta_d \sim Dirichlet(\alpha)$ is drawn followed by topic index $z_{di} \in \{1, 2, \dots, K\}$ and topic weights

$z_{di} \sim \theta_d$; finally word w_{di} is drawn from the selected topic $w_{di} \sim \beta_{z_{di}}$. In this study, a variation with faster online implementation (Hoffman et al., 2010) is used. Each probability corresponding to the topics for each document will be saved and used with the regression models.

Finally, the topic tensor $T = \{z_{k,l,t}\}$ is defined using the result of the topic models. It contains topic weights z of topic k on a collection of tweets falling under the spatial index l and the temporal index t . We experimentally set the latitude and longitude indices equal to each other as we processed the human mobility trajectory data, and the temporal index for every hour, and the number of topics k as 10.

5. PREDICTION MODEL

Our prediction model $P_\theta(h\hat{m}s_{u,t+1}|hms_{u,t})$ is that given a subset of human mobility sequence $hms_{u,t} = r_{t1}, r_{t2}, \dots, r_t$, it predicts a human mobility sequence of next time step $t + 1$ $h\hat{m}s_{u,t+1} = r_{t+1}$. θ is a set of model parameters that is obtained by minimizing the model loss $L(h\hat{m}s_{u,t+1}, hms_{u,t+1})$ as follows:

$$\theta = \underset{\theta}{\operatorname{argmin}} L(h\hat{m}s_{u,t+1}, hms_{u,t+1}) \quad (10)$$

By applying the prediction model autoregressively, it predicts the human mobility sequence for several steps. As the location is encoded as the spatial index, each spatial index label is converted to a one-hot vector. Thereafter, the model tries to minimize the categorical cross entropy between the predicted distribution \hat{y}_l and the true distribution y_l . Therefore, 10 can be further expanded as:

$$\theta = \underset{\theta}{\operatorname{argmin}} \left(- \sum_l \hat{y}_l \log y_l \right) \quad (11)$$

5.1 Model architecture

The model architecture is described in Fig. 2. It consists of embedding layers, multiple long-short term memory unit layers and one activation layer. For the output activation, we used Softmax in this study.

5.2 Trajectory embedding

The original location denotation in the sequence is the region ID. As the IDs are arbitrarily assigned to each grid cell, the spatial distribution of the grid cells should be learned from trajectory embedding; this idea is similar to Word2Vec in word embedding (Mikolov et al., 2013). This also contributes to reducing the dimension of the input vector, thereby reducing the memory and computational cost.

5.3 Multi-source data integration

Each mobility input, topic input, and travel mode input has an individual embedding layer. The output of the embedding layers are concatenated to one shared hidden layer. This would enable multi-modal network structure to learn multi-source data concurrently.

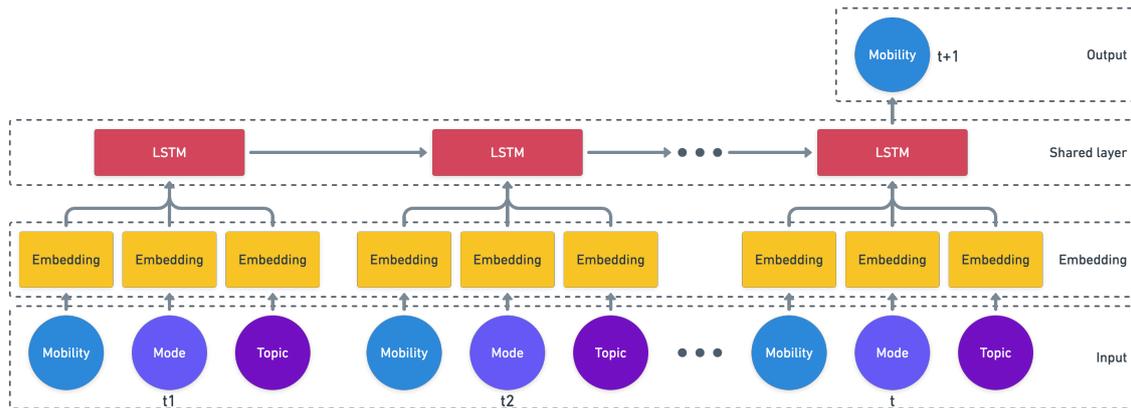


Figure 2: Prediction model architecture

5.4 LSTM network

The shared layer uses the long-short term memory (LSTM) unit (Hochreiter and Schmidhuber, 1997). LSTM network is an extension of recurrent neural network (RNN). It takes a sequence $X = x_1, x_2, \dots, x_T$ as the input and another sequence $H = h_1, h_2, \dots, h_T$ as the output to compute a mapping of each network unit using the following equations:

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \quad (12)$$

$$\tilde{C}_t = \tanh(W_c x_t + U_c h_{t-1} + b_c) \quad (13)$$

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \quad (14)$$

$$C_t = i_t * \tilde{C}_t + f_t * C_{t-1} \quad (15)$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \quad (16)$$

$$h_t = o_t * \tanh(C_t) \quad (17)$$

$$y_t = \phi(W_y h_t + b_y) \quad (18)$$

where i_t , f_t , and o_t are input gate function, forget gate function, and output gate function at time t , respectively. \tilde{C}_t is the candidate values and C_t is the new value for the states of the memory cells at time t . x_t is the input memory cell layer and h_t is the representation layer at time t . $W_i, W_c, W_f, W_o, W_b, U_i, U_c, U_f$, and U_o are weight matrices. b_i, b_c, b_f, b_o , and b_y are bias vectors. ϕ is the network activation function and Softmax was used in this study. Including the forget gate function is an attempt to mitigate the vanishing gradient problem.

5.5 Prediction evaluation scheme

The prediction model is trained to minimize the model loss (categorical cross entropy) defined in (11). The model with different parameter settings and the baseline models are then evaluated based on categorical accuracy. Among the input data, 80% of the data is used for training and remaining 20% is for evaluation.

6. EXPERIMENTS

6.1 Result of Topic modeling

To train the topic models, we selected the area of interest as the Greater Tokyo Area (138.72 to 140.87 in longitude, 34.9 to 36.28 in latitude) and the time period was from July 25, 2013 to July 31, 2013. The model was then applied to the tweets from different

time spans to produce the topic tensor; this is discussed in the following section. We experimentally used the topic tensor from LDA in the prediction model.

Table 2 shows some of the topics that are interpretable as distinct topics. Some topics can be considered strongly related to mobility and transportation infrastructure (topic 2), while some are related to news reporting (topic 7) or seasonal events (topic 10).

Topics	Words
2. Commute/Transportation	到着(arrive) 乗り換え(transfer) 最初(first) 案内(guide) 行か(go)
5. Daily life	仕事(work) 昨日(yesterday) 疲れ(tired) 行っ(go) 大丈夫(good) 帰り(going home) 夜(night) 来る(come)
9. Commute/event	暑い(hot) 開催(event;held) 出勤(commute) 始め(start) オープン(open) か*んばっ(work hard) 早(early)
7. News reporting	km 地震(earthquake) 速報(breaking news) 震源(seismic center) 車線(train/car track) 震度(seismic intensity) 気圧(air pressure) 今月(this month)
10. Event/seasonal	花火(fireworks) オリンピック(Olympic) 休み(holiday) サッカー(soccer/football) 負け(lose) 酒(alcohol)

Table 2: Selected topics and words with high association (LDA). The words with high association to each topic are listed with translation. The authors interpreted and labeled the topic based on the words.

6.2 Prediction scenario and parameter settings

To evaluate the models' performance, we set a prediction scenario and parameter settings (Table 3). The time period from 8 AM to 9 AM on July 26 (Thursday), 2012 was chosen as the target prediction span. Set within a morning commute, we expected the majority of the mobility to be the routine commutes, which have been a challenge for mobility prediction models in previous studies (Jiang et al., 2018b).

Setting	Value
Time span of observation	2012-07-25 08:00:00~07:59:59
Time span of prediction	2012-07-26 08:00:00~08:59:59
Prediction output length	next 1 hour (autoregressive)
Δt	10 minutes
k	10
Prediction LSTM layers	2, 4, 6
Embedding size	1000
LSTM network activation function	tanh, softsign, adam

Table 3: Model architecture and parameter settings

6.3 Baseline models

In addition to Deep Learning models, we added two baseline models: the N-Gram-Like and simple recurrent neural network (RNN) models. The N-Gram-Like model, which is typically used for natural language processing is trained using N and N-1 consecutive sequences of grid IDs; for each N-1 consecutive sequence, it computes the probability distributions for the next step. Here, we adopt Four-gram based model (Jiang et al., 2018b) which uses three steps to predict the next time step. The simple RNN model is comprised of traditional fully-connected RNN structure.

6.4 Performance evaluation

6.4.1 Choosing input data and network structure Table 4 compares the accuracy depending on the deep Learning model and parameter selection. The model based on 2 layers of LSTM and the embedding layer combining grid (location), transportation mode, and topic feature achieved the best accuracy among the other structures. The main model is LSTM with tanh as the output activation layer with dropout function. Other variations (LSTM-without-dropout, LSTM-softsign, LSTM-adam) performed worse than the main model. While adding only the mode feature improved the accuracy, adding only the topic feature did not improve the accuracy. However the combination of the three features improved the accuracy the most.

6.4.2 Computational cost, training data size and learning efficiency To conduct the experiment quickly, we only selected the experimental setting that required less than 24 hours for the training phases. Our primary computation machine consisted of eight CPU cores and one NVIDIA TITAN X GPU and make sure the model is not overfitting to the training data (Figure 3). As Table 5 shows increasing the number of samples indeed improve the prediction accuracy up to 78.5% (Table 5).

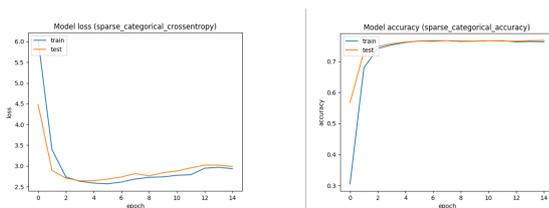


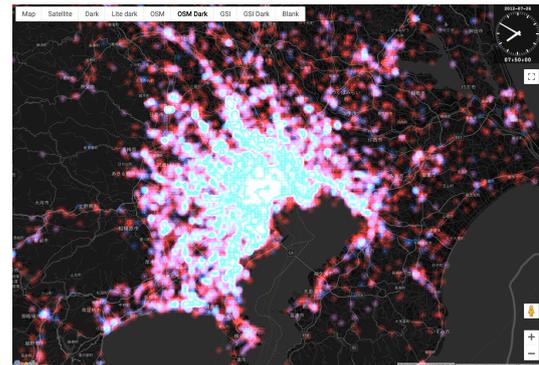
Figure 3: Model loss and model accuracy learning curve

6.5 Spatial evaluation of prediction result

Figure 4 shows the prediction result. Basically, the distribution of points are quite similar between predicted (blue) and ground



(a) 07:00



(b) 07:50

Figure 4: Predicted (blue) and ground truth (red) trajectory at 07:00 (a) and 07:50 (b). Credit: "Konzatsu-Tokei(r)" (c)ZENRIN DataCom CO., LTD.

truth (red). Whereas the individual prediction accuracy represents the accuracy for each individual user, this figure represents the macro scale accuracy to be evaluated if the aggregated results of the model reflects the real-world human mobility. Overall, the predicted result reflects the spatial distribution of human mobility; however, the spatial discrepancy is most significant in rural areas most likely because of the lack of training data.

Figure 5 shows the absolute percentage error of number of users (predicted and ground truth) to evaluate quantitative error in each cell. The spatial distribution of the error is fairly constant that the cells with significant error ($\geq 150\%$) are located not at the center of Tokyo but along transportation infrastructures.

6.6 Prediction result and share of each topic

Figure 6 shows the prediction result of a sample trajectory. While the model inaccurately predicts a middle vertex, it predicts the destination accurately. By extracting the spatiotemporal topic feature, each trajectory can be individually analyzed based on which topics are significant at the moment of prediction.

7. CONCLUSION

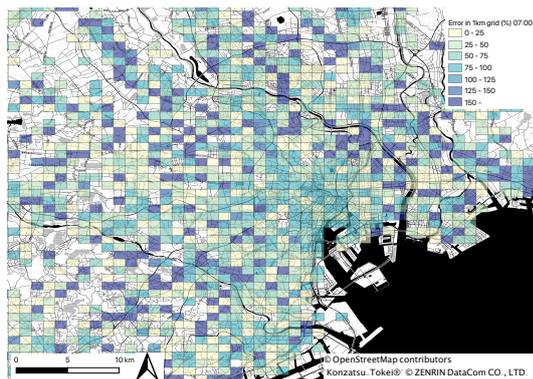
In this study, we developed a multi-modal human mobility prediction model using LSTM combining GNSS trajectory and SNS data. The experiment demonstrated that the model successfully combined the mobility and topic features for the prediction scenario and performed better than typical baseline models. While the access to the input data limits the applicability of the study, the model structure can be introduced to other related applications.

Model	Input features	Layers	Loss	Accuracy
Baseline Ngram	Grid			0.5425
Baseline RNN	Grid	2	8.4946	0.2910
LSTM	Grid	2	5.4619	0.61270
LSTM-without-dropout	Grid	2	5.9633	0.5430
LSTM-softsign	Grid	2	5.5412	0.5932
LSTM-adam	Grid	2	6.7648	0.3581
LSTM-bidirectional	Grid	2	8.0963	0.3289
LSTM-Embedding encoder	Grid, topic	2	5.5070	0.6055
LSTM-Embedding encoder	Grid, mode	2	5.3604	0.6127
LSTM-Embedding encoder	Grid, mode, topic	2	5.3829	0.6153
LSTM-Embedding encoder	Grid, mode, topic	4	7.1844	0.3991
LSTM-Embedding encoder	Grid, mode, topic	6	8.7217	0.1214

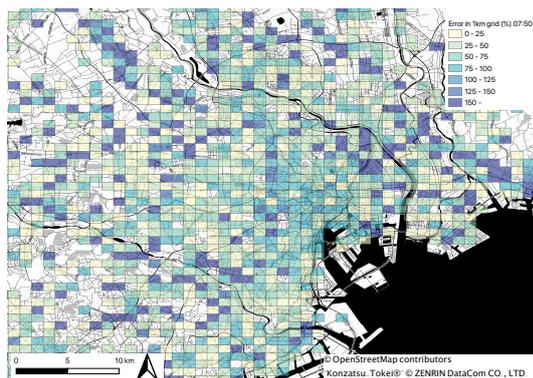
Table 4: Network structure and embedding methods. Credit: "Konzatsu-Tokei(r)" (c)ZENRIN DataCom CO., LTD.

Sample size	Loss	Accuracy
10000	5.3829	0.6153
50000	3.0759	0.7614
100000	2.9287	0.7812
150000	2.8517	0.7822
200000	2.8516	0.7850

Table 5: Effect of number of samples on prediction accuracy



(a) 07:00



(b) 07:50

Figure 5: Absolute percentage error of number of users (predicted and ground truth) in 1km grid at 07:00 (a) and 07:50 (b). Credit: "Konzatsu-Tokei(r)" (c)ZENRIN DataCom CO., LTD.

ACKNOWLEDGEMENTS

We would like to thank Editage (www.editage.com) for English language editing.

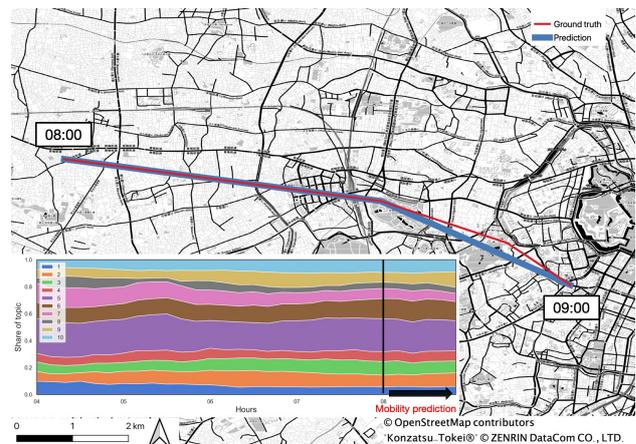


Figure 6: Visualization of individual prediction result. The model produces a small error during the autoregressive process, but predicts correct destination for one hour later. The stacked chart shows the share of each topic with respect to the temporal index (in every 10 minutes). Topic 9 (commute/event) is gaining share in the prediction time span while topic 5 (daily life) is significant throughout the whole time span.

REFERENCES

Birenboim, A. and Shoval, N., 2016. Mobility Research in the Age of the Smartphone. *Annals of the American Association of Geographers* 4452(January), pp. 1–9.

Blei, D. M., Ng, A. Y. and Jordan, M. I., 2003. Latent Dirichlet Allocation. *Journal of Machine Learning Research* 3(4-5), pp. 993–1022.

Calabrese, F., Diao, M., Di Lorenzo, G., Ferreira, J. and Ratti, C., 2013. Understanding individual mobility patterns from urban sensing data: A mobile phone trace example. *Transportation Research Part C: Emerging Technologies* 26, pp. 301–313.

Gong, L., Liu, X., Wu, L. and Liu, Y., 2015. Inferring trip purposes and uncovering travel patterns from taxi trajectory data. *Cartography and Geographic Information Science* 0406(June 2015), pp. 1–12.

Grinberger, A. Y. and Shoval, N., 2015. A temporal-contextual analysis of urban dynamics using location-based data. *International Journal of Geographical Information Science* 29(11), pp. 1969–1987.

Hochreiter, S. and Schmidhuber, J., 1997. Long Short-Term Memory. *Neural Computation* 9(8), pp. 1735–1780.

- Hoffman, M., Blei, D. and Bach, F., 2010. Online learning for latent dirichlet allocation. *Nips* pp. 1–9.
- Jiang, R., Song, X., Fan, Z., Xia, T., Chen, Q., Chen, Q. and Shibasaki, R., 2018a. Deep ROI-Based Modeling for Urban Human Mobility Prediction. In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, Vol. 2 number 1, pp. 1–29.
- Jiang, R., Song, X., Fan, Z., Xia, T., Chen, Q., Miyazawa, S. and Shibasaki, R., 2018b. DeepUrbanMomentum : An Online Deep-Learning System for Short-Term Urban Mobility Prediction. In: *The Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18)*, pp. 784–791.
- Kang, C., Ma, X., Tong, D. and Liu, Y., 2012. Intra-urban human mobility patterns: An urban morphology perspective. *Physica A: Statistical Mechanics and its Applications* 391(4), pp. 1702–1717.
- Li, S., Dragicevic, S., Castro, F. A., Sester, M., Winter, S., Coltekin, A., Pettit, C., Jiang, B., Haworth, J., Stein, A. and Cheng, T., 2016. Geospatial big data handling theory and methods: A review and research challenges. *ISPRS Journal of Photogrammetry and Remote Sensing* 115, pp. 119–133.
- Liu, Z., Huang, Y. and Trampier, J. R., 2016. Spatiotemporal topic association detection on tweets. In: *Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems - GIS '16*, ACM Press, New York, New York, USA, pp. 1–10.
- Mikolov, T., Corrado, G., Chen, K. and Dean, J., 2013. Efficient Estimation of Word Representations in Vector Space. *Proceedings of the International Conference on Learning Representations (ICLR 2013)* pp. 1–12.
- Miyazawa, S., Song, X., Xia, T., Shibasaki, R. and Kaneda, H., 2019. Integrating GPS trajectory and topics from Twitter stream for human mobility estimation. *Frontiers of Computer Science* 13(3), pp. 460–470.
- Morstatter, F., Pfeffer, J., Liu, H. and Carley, K. M., 2013. Is the Sample Good Enough? Comparing Data from Twitter's Streaming API with Twitter's Firehose. *Proceedings of ICWSM* pp. 400–408.
- Mota, T., Munjal, A. and Camp, T., 2015. Large-Scale Human Mobility Analysis Based on Mobile Phone and Social Media Communication: A Case-Study in Africa. *2015 16th IEEE International Conference on Mobile Data Management* pp. 86–91.
- Ngiam, J., Khosla, A., Kim, M., Nam, J., Lee, H. and Ng, A. Y., 2011. Multimodal Deep Learning. *Proceedings of The 28th International Conference on Machine Learning (ICML)* pp. 689–696.
- Niu, X., Zhu, Y. and Zhang, X., 2015. DeepSense: A novel learning mechanism for traffic prediction with taxi GPS traces. In: *2014 IEEE Global Communications Conference, GLOBECOM 2014*, Institute of Electrical and Electronics Engineers Inc., pp. 2745–2750.
- Pan, B., Zheng, Y., Wilkie, D. and Shahabi, C., 2013. Crowd sensing of traffic anomalies based on human mobility and social media. In: *Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems - SIGSPATIAL'13*, ACM Press, New York, New York, USA, pp. 334–343.
- Reades, J., Calabrese, F. and Ratti, C., 2009. Eigenplaces: Analysing cities using the space - Time structure of the mobile phone network. *Environment and Planning B: Planning and Design* 36(5), pp. 824–836.
- Řehůřek, R., 2011. Subspace Tracking for Latent Semantic Analysis. In: *ECIR'11 Proceedings of the 33rd European conference on Advances in information retrieval*, pp. 289–300.
- Sakaki, T., Okazaki, M. and Matsuo, Y., 2013. Tweet Analysis for Real-Time Event Detection and Earthquake Reporting System Development. *IEEE Transactions on Knowledge and Data Engineering* 25(4), pp. 919–931.
- Song, C., Koren, T., Wang, P. and Barabási, A.-L., 2010. Modeling the scaling properties of human mobility. *Nature Physics* 6(10), pp. 818–823.
- Song, X., Kanasugi, H. and Shibasaki, R., 2016. DeepTransport : Prediction and Simulation of Human Mobility and Transportation Mode at a Citywide Level. *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-16)* pp. 2618–2624.
- Song, X., Zhang, Q., Sekimoto, Y. and Shibasaki, R., 2014. Prediction of human emergency behavior and their mobility following large-scale disaster. In: *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '14*, ACM Press, New York, New York, USA, pp. 5–14.
- Spaccapietra, S., Parent, C., Damiani, M. L., de Macedo, J. A., Porto, F. and Vangenot, C., 2008. A conceptual view on trajectories. *Data & Knowledge Engineering* 65(1), pp. 126–146.
- Tang, L., Kan, Z., Zhang, X., Yang, X., Huang, F. and Li, Q., 2016. Travel time estimation at intersections based on low-frequency spatial-temporal GPS trajectory big data. *Cartography and Geographic Information Science* 43(5), pp. 417–426.
- Xu, F., Zhang, P. and Li, Y., 2016. Context-aware real-time population estimation for metropolis. In: *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '16*, ACM Press, New York, New York, USA, pp. 1064–1075.
- Yuan, J., Zheng, Y. and Xie, X., 2012. Discovering regions of different functions in a city using human mobility and POIs. In: *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '12*, ACM Press, New York, New York, USA, p. 186.
- Zhang, D., Wang, Z., Guo, B. and Yu, Z., 2012. Social and Community Intelligence: Technologies and Trends. *IEEE Software* 29(4), pp. 88–92.
- Zheng, Y., 2015. Methodologies for Cross-Domain Data Fusion: An Overview. *IEEE Transactions on Big Data* 1(1), pp. 16–34.
- Zheng, Y., Capra, L., Wolfson, O. and Yang, H., 2014. Urban Computing: Concepts, Methodologies, and Applications. *ACM Transactions on Intelligent Systems and Technology* 5(3), pp. 1–55.
- Zheng, Y., Zhang, H. and Yu, Y., 2015. Detecting collective anomalies from multiple spatio-temporal datasets across different domains. *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems - GIS '15* pp. 1–10.
- Zhou, J., Pei, H. and Wu, H., 2018. Early Warning of Human Crowds Based on Query Data from Baidu Maps: Analysis Based on Shanghai Stampede. *Advances in Geographic Information Science*, Springer International Publishing, Cham, pp. 19–41.